Joint attention and reference construction: The role of pointing and "so"

Angeliki Balantani*, Stefanie Lázaro

Section d'allemand, University of Lausanne, Lausanne, Switzerland

ARTICLE INFO

ABSTRACT

This study examines the multimodal package "so" with a pointing gesture in Swiss German. The analysis shows that by using "so" concurrently with resources generally considered to be in the service of local deixis, such as pointing gestures, participants are directing their addressees' attention to a domain of scrutiny where they should look for additional information to fill the lexico-semantic content of the modal deictic "so". This lexico-semantic content is the outcome of an action, the correct way of accomplishing an action. Based on video-recordings of two different settings, and taking into account the details of embodied conduct, the paper discusses how local deictic resources when combined with modal deixis can be used to highlight the manner.
© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

The study of the organisation of social interaction has demonstrated that talk is a collaborative achievement (Atkinson and Heritage, 1984; Sacks et al., 1974; Schegloff, 2007). In order to achieve meaning in interaction, interlocutors need to monitor each other's talk to coordinate their actions (Sacks et al., 1974). The use of video recordings has made it possible to investigate the role of other resources, such as gaze and gestures, in the production of action, thus allowing us to re-conceptualise action in interaction by taking the multimodal perspective into account (see for example Deppermann, 2013; Goodwin, 2000a; Mondada, 2013). The aim of this paper is to investigate the use of pointing in combination with modal deixis as resources that participants have at their disposal to coordinate their actions.

Pointing is an important resource for joint attention and reference construction in interaction (Bangerter, 2004; Diessel, 2006; Eriksson, 2009; Goodwin, 2000b, 2003; Hanks, 1992; Hindmarsh and Heath, 2000; Mondada, 2014a; Stukenbrock, 2009, 2015, 2018b; Tomasello, 1995). The importance of pointing gestures for local deixis has been widely investigated (Clark, 2003; Fricke, 2007; Haviland, 1993, 2000; Kendon, 2004; Kita, 2003; Levinson, 2004; McNeill, 2005; Stukenbrock, 2015, 2018a), but there is little research on pointing gestures in the service of modal deixis (but see, for example, Harweg, 1990). In this paper, we show that the use of pointing concurrently with the modal deictic "so" (which roughly translates to "like this") forms a multimodal package (Goodwin, 2003; Heath, 1986; Stukenbrock, 2010) or a multimodal Gestalt (Mondada, 2014b) that is implemented by interlocutors in sequences of actions in which establishing a joint focus of attention is important for moving to next-positioned matters.

* Corresponding author.
E-mail addresses: angeliki.balantani@unil.ch (A. Balantani), stefanie.lazaro@unil.ch (S. Lázaro).

The investigated practice mostly occurs in sequences in our data set in which interlocutors are collaboratively engaged in a joint activity (Clark, 1996), for instance when looking together at an object in a museum or building a piece of furniture. In these activities, "so" + pointing is used by participants to point at the result of an action that has already been performed and is either visible as a depiction in an instruction booklet or as an object in the physical surroundings. In this way, participants manage to establish co-orientation and, in turn, coordination (Hausendorf, 2013), necessary prerequisites for the achievement of joint activities (Clark, 1996) such as building a piece of furniture together. By focusing on the interactional function of pointing gestures in combination with the Swiss German modal deictic "so", our analysis contributes to continuing research on establishing joint focus of attention and reference construction in interaction.

2. Theoretical framework

Deictic expressions can serve as a way of achieving intersubjective understanding by bringing about "the transition from individual acts of perception to mutually shared perception" (Hausendorf, 2003, p. 252), and can contribute to the achievement of joint attention (Bangerter, 2004; Bruner, 1995; Diessel, 2006; Mundy and Newell, 2007; Scaife and Bruner, 1975; Stukenbrock, 2018a, 2018b; Tomasello, 1995). Joint attention is an important prerequisite to coordinating attention with a co-participant in talk (Mundy and Newell, 2007) and refers to the fact that the participants not only direct their visual attention to the same object or entity in the visual surroundings, but that they also know that they are sharing the same focus of attention, or in Diessel's (2006, p. 465) words: "(…) joint attention requires that the communicative partners recognize they are attending to the same thing".

The term *deixis* (from the Greek word δεῖξις for *pointing* or *indicating*) refers to grammaticalized forms that can only be interpreted when taking the contextual aspects of an utterance or speech event into account (Levinson, 1983). In this study, we concentrate on deictic expressions that are used gesturally (Fillmore, 1997) and where interlocutors need to monitor the speech event closely to interpret those expressions (Levinson, 1983). In order to function properly, these deictic expressions have to be combined with an embodied action. One of the most studied deictic embodied actions is the pointing gesture (see for example Clark, 2003; Fricke, 2007; Kendon, 2004; Kita, 2003; McNeill, 1992; Mondada, 2014a; Stukenbrock, 2015) with previous work on pointing focusing mainly on the use of this gesture in local deixis (Fillmore, 1997).

Although pointing is a well-established and commonly understood practice to direct the visual attention of others to something in the physical surroundings, addressees of a pointing gesture do not simply have to find the target that is being pointed at by following a "vector" that indicates a certain direction, location or object (Kita, 2003). In fact, addressees can rely on other *meaning-making practices* (Goodwin, 2000a, p. 1516), like the co-occurring talk and the activity in which the pointing gesture occurs, to identify the target and establish a referent (Goodwin, 2000b, 2003; Stukenbrock, 2015). The act of pointing thus only specifies a *domain of scrutiny* (Goodwin, 2000b, 2003) where the addressees of the point have to search for the target, that "is not a simple, physical 'thing' but an entity that is structured through the intersection of a range of different kinds of semiotic practices" (Goodwin, 2000b, p. 74). In such cases, the situated pointing gesture is coordinated with other *concurrent communicative acts* (Streeck, 2013, p. 675) or *semiotic resources* (Goodwin, 2000a) to form a multimodal package (Goodwin, 2003; Heath, 1986; Stukenbrock, 2010) or a multimodal Gestalt (Mondada, 2014b).

Another multimodal package that has been widely investigated is the use of the German particle "so" in conjunction with gestures in the service of modal deixis (Ningelgen and Auer, 2017; Streeck, 2002; Stukenbrock, 2010, 2014, 2015). Streeck (2002) describes the use of the particle "so" in the presence of descriptive gestures (referred to by McNeill (2005) as iconic gestures). According to Streeck, "so" appears in a speaker's turn before the gesture and serves to introduce the nonverbal behaviour that is imminent. He proposes that "so" serves as a *flag* for the interlocutor to look for an additional meaning in the utterance that is being produced. The speaker's gaze direction attracts the addressee's visual attention to the location where this additional meaning can be found. When gaze orientation to the speaker's hands is combined with the flag, the hand gesture is integrated into the structure of the talk and assumes a grammatical status in the utterance. However, Streeck does not distinguish the cases of "so" that do require a gesture from the ones that are accompanied by one but the gesture is not mandatory for the intelligibility of the utterance (Ningelgen and Auer, 2017; Stukenbrock, 2010, 2014).

Stukenbrock (2010, 2014, 2015) concentrates her research on instances of deictic "so" expressions for which the presence of a gesture is obligatory for the intelligibility of an utterance. In such cases, she claims that "so" functions as a *flag* (Streeck, 2002) for an upcoming gesture by the speaker that must be visually perceived by the addressee. The obligatory presence of gaze in deictic "so" references is also identified in instructional sequences (Stukenbrock, 2014). What the author demonstrates with her analysis is that "so" does not merely function as a turn-internal *flag* that directs the addressee's gaze orientation to a gesture, but rather that a deictic "so" projects a bodily behaviour that can range from iconic or depictive gestures to full-fledged bodily performances (Stukenbrock, 2014) by the speaker and an aligning action by the addressee, thus making the addressee's gaze obligatory.

While these studies are concerned with the occurrence of "so" in conjunction with iconic gestures and performances, the use of this particle coupled with pointing gestures has not been dealt with in depth. To our knowledge, only Harweg (1990) argues, albeit theoretically, that pointing gestures, though their use is predominantly associated with local deixis, could nevertheless be used in modal deixis too. Fricke (2007), on the other hand, maintains the hypothesis that, as pointing gestures

only occur with reference to local deictic expressions, such as "here" or "this one", when modal deictic expressions are accompanied by a pointing gesture, they belong to the category of local deixis.

Our aim for this paper is to show that while pointing gestures are used to direct addressees' attention to an object or space in the visual environment, if combined with the modal deictic "so", they can in fact be in the service of modal deixis and point to the *outcome* of an action. This complements the description of multimodal packages with "so" identified so far (Stukenbrock, 2014), where the action in progress has to be monitored to "fill" the lexico-semantic content of "so". To illustrate this, take for instance the following excerpt from Stukenbrock (2014, p. 87):

Example 2: "so abschälen"/ "peel like this" (KK1_1_00:03:05)

```
01    K2    und wenn man die jetzt mit dem MESser schneidet,
            and when you cut it with the knife
02          muss man also SO abschälen,
            you have to peel it like this
```

In this example, the utterance "so abschälen"/ "peel like this" (l.02) has to be accompanied by some action from the speaker that is to be performed more or less concurrently with "so" and will point to its quality. By way of contrast, as we shall highlight in this paper, when "so" is coupled with a pointing gesture, what is being pointed at is the outcome of the action rather than the performance per se.

3. Methodology and data

The methodology used to undertake this study is Conversation Analysis (hereafter CA) (Drew, 2004). As a bottom-up form of analysis, CA puts participants' own displayed orientations at the centre of the analysis. Its objective is to describe how interactants produce their own behaviour and understand the behaviour of others. As social interaction not only consists of verbal but also of embodied conduct, the different modalities people use in interaction have to be taken into consideration in the analysis (see, for example, Deppermann, 2013; Goodwin, 2000a; 2000b; Mondada and Schmitt, 2010; Stivers and Sidnell, 2005; Streeck et al., 2011). Based on the methodological principles of CA, we investigate both the simultaneous combination of individual modalities in the performance of an action, and the simultaneous performance of actions by different interactants (Deppermann, 2013; Stivers and Sidnell, 2005).

Data recordings were conducted following the quality criteria for the collection of naturally occurring interaction (Deppermann, 2008; Mondada, 2013) with the use of mobile eye-tracking glasses (Tobii Pro Glasses 2) worn by the participants. An additional third camera was used during the recordings, allowing for an observer's perspective of the ongoing interaction, in order to account for embodied conduct not visible through the cameras on the eye-tracking devices. The recordings from the eye-tracking glasses and from the third camera were synchronized into one split screen video and imported into ELAN (2019) for manual transcription of verbal (Selting et al., 2009) and embodied conduct (Mondada, 2019) (see Appendix). In the extracts, the first line of the transcription is in Swiss German and the second line is a translation in English. For purposes of accuracy, we will show some embodied actions that appear on the videos in screenshots, choosing the perspective that depicts the action with the most clarity. In some instances, the images will be from one of the eye-tracking glasses and in others from the split screen.

The data analyzed in this paper is part of a larger corpus collected for a research project that focuses on deixis and joint attention. Our collection consists of 11 instances of pointing gesture + "so" utilised to point at the outcome of an action. We will show four instances of the phenomenon in our analysis section that present different variants of the identified practice. Three of the examples pertain to our "relocation" data set, in which participants are building furniture following an instruction booklet of IKEA, and one is from our "museum" data.

4. Empirical analysis: pointing at the outcome of an action

In the first instance, we examine one exemplar of this practice from our "museum" data in order to provide an overview of the main features of the practice: the particle "so" accompanied by a pointing gesture at an object or entity as a visible outcome of an action. We then examine three more comparative instances from our "relocation" data set, where this practice appears in settings with instructional actions either to give instructions or to confirm the result of a just performed action. The examples provided present recognizably the same practice with distinct temporal trajectories.

In excerpt 1, participants are walking around a game-museum and are making comments about the different games they can observe. In this sequence, Anna (see Fig. 1, on the left) comments on a Rubik's cube that is on display.

(Figure 1)

(1) Spielmuseum_02_Rundgang_14:09-14:25 (M: Mike, A: Anna)

* for gaze by A
\+ for gaze by M
\$ embodied action done by A
& embodied action done by M

```
01     M     ou die han &ich immer so GHASST;
             oh I always kind of hated those
       m                 &................-->
02           (0.6)&
       m      -->&
03     M     &du muesch verSUEche s-&
             you have to try to
       m     &PG¹-------------------&
04           &(0.8)&(0.1)
       m     &,,,,,&
05     M     aso eifach ZÄMEmache damits-
             well you simply do them together so that it
06     A     hehe
07     M     wieder die FORM git,
             gives this shape again
08           (0.5)+(0.5)
       m          +Rubik's cubes-->² l.11
09     A     hehe <<lachend> *das hani NIE gmacht gläb;>
             hehe that I have never done I think
       a                    *Rubik's cubes-----------> l.11
10           (0.1)
11     A     $<<lachend> a*ber$ hehehe do *h° *DAS+ chan ich nit*;>
                         but I cannot do this
       a     $...............$PG--------------------------------> l.13
       a             -->*L Rubik's cube-*   *Mike------------*
       m                                    -->+L Rubik's cube--> l.13
12     A     ((lacht))*  ((lacht))    *
             ((laughs))
       a             *L Rubik's cube*
13     A     <<lachend> de RUbiks cube$+;>
                       the Rubik's cube
       a                            -->$
       m                            -->+
14     A     $+((lacht))$[((lacht))                   ]
```

---

¹ PG stands for "Pointing Gesture".
² There are two Rubik's cubes on display (see figures 2 and 3). "L Rubik's cube" refers to the one on the left and "R Rubik's cube" to the one on the right.

```
a      $,,,,,,,,,,$
m      +R Rubik's cube-------------------------> l.16
15  M              &[aso ich würds&SO& #*mache;&]*
                    well I would do it like this
m                  &..............&PG&,,,,,,,,,&
a                                  *R Rubik's cube*
fig                                #fig2/3
```


Figure 2

Figure 3

```
16  M   eifach [useNA+NDnä;]
        simply take apart
m               -->+
17  A           [<<lachend>>geNAU;]
                         exactly
```

The speakers are in the section of the museum with the mind games. Anna points with her little finger at the Rubik's cube and claims that she cannot do this "<<lachend> aber hehe do DAS chan ich nit;>"/ "but I cannot do this" (l.11). Mike affiliates with her at line 15 "aso ich würds SO mache;"/ "well I would do it like this". His turn is formulated with the deictic expression "so" which functions as a *flag* (Streeck, 2002) that attracts the addressee's visual attention. The *flag* is combined with a pointing gesture; at "SO" he points at the cube that is next to the one that his interlocutor pointed at (Fig. 2). While the cube that Anna points at (Fig. 3, left) is the conventional Rubik's cube that one can buy in a store, the cube that Mike points at illustrates a disassembled version of the Rubik's cube (Fig. 3, right). Both interlocutors use a pointing gesture to specify the exhibit they are referring to. Anna uses the demonstrative "das" + pointing gesture to point at the cube that represents the normative way of playing this game, i.e. attempting to solve it by moving the colours around. The demonstrative "das" marks a shift of attention to the Rubik's cube (Diessel, 2006) and introduces the new referent as a "commentable" for the interlocutors. Having introduced the Rubik's cube as a new focus of attention, Mike uses the modal deictic "so" with the pointing gesture to point at an alternative and contrastive version of how this game could be played, namely by cheating. He uses the modal deictic "so" with a gesture to point at the disassembled cube that shows the visible outcome of an action, i.e. playing the game by disassembling the pieces, an action which he also expresses verbally in the next turn in line 16 "eifach useNANDnä"/ "simply take apart".

This is a clear example of the practice we have identified in our data set where the deictic "so" with a pointing gesture is utilised to point at a target in the surroundings that shows the outcome of an action, how Mike would do the Rubik's cube and how the cube would look like after he had played with it.

Having presented our first example which shows the investigated practice, we now turn to our "relocation" data for a more detailed analysis. In all three extracts that follow, two closely related participants are jointly building a cupboard. In extract 2 (Umzug_03), a couple, Arno and Fabienne, are moving into a new flat and are building their furniture. In extracts 3 and 4 (Umzug_01) we have two sisters, Anna and Bettina. Bettina is moving into a new flat and Anna is helping out with building the

furniture. In all of the examples, the participants use instruction manuals and establish a division of labour in an emergent way.

Before we present our empirical analysis of these three extracts, we outline a particularity of our data set, which contrasts to other instructional sequences (De Stefani and Gazin, 2014; Ekström, 2012; Lindwall et al., 2015; Stukenbrock, 2014). The participants in the "relocation" recordings are confronted with a complex task, building a cupboard with the help of an instruction booklet, and although the instructions in the booklet are illustrated with images and follow a step-by-step procedure (see Fig. 4 for an example of the instruction booklet where the steps are numbered, for instance 22, 23 and 24), the process itself is not represented.

(Figure 4)

Example of instruction booklet with numbered steps.

It is not like video tutorials, for instance, where one follows the process of building a piece of furniture step-by-step and the instructor does it virtually with you. The images on the booklet depict the outcome of a process, how the result has to look like once the instructions have been followed correctly (see also Garfinkel, 2002). That is, the instruction booklet only depicts certain aspects of the task at hand (for example, which screw goes where), but for the practical aspects of assembling the furniture, the participants have to rely on their knowledge about the world and their previous experience with similar tasks. In our examples, one assumes the task of reading and understanding the pictures in the instruction booklet and giving "instructions" to the other participant, while the other deals with the more practical task and the material elements in the physical surroundings. There is a certain distribution of labour, and accordingly, roles between the two participants. Consequently, the participants must coordinate two different *spatial frameworks* (Goodwin, 2003, p. 223) and are faced with a coordination problem they need to solve in order to achieve a satisfying outcome. While the verbal component "so" of the multimodal package we are investigating refers to a referent that can be instantiated by two different targets, the pointing gesture establishes a domain of scrutiny either on the instruction booklet or in the physical surroundings, where the participants need to determine the corresponding target. The multimodal package "so" + pointing gesture helps them establish a joint focus of attention necessary to solve the momentary problems before proceeding with the next step of the task at hand.

In sum, the images on the instruction booklet can serve different functions. For one, they are there to "instruct" how to perform the task, but also to, retrospectively, check whether the task has been performed correctly. And the verbal component "so" of the multimodal package emphasizes the way in which the element being pointed at has to be "understood". In our case, the features being pointed at illustrate the outcome of an action. Altogether, the different resources involved complement each other to achieve a joint focus of attention and an intersubjective understanding, prerequisites for achieving coordination and co-operation (Hausendorf, 2013) and a joint course of action (Goodwin, 2003).

We now present the first example of the "relocation" data set. Fabienne (Fig. 5, standing and the one in charge of reading the instruction booklet in this sequence) and Arno (Fig. 5, seated) have put the bottom part of the cupboard on the floor and are about to proceed to the next step when Arno initiates repair (Schegloff, 1992, 2000b).

(Figure 5)

(2) Umzug_03_21:20-21:37 (A: Arno, F: Fabienne)

* for gaze by A
\+ for gaze by F
$ embodied action done by A
& embodied action done by F


```
01   F    ÄM:,
02        (0.8)
03        Aso;
          so
04        (0.4)*(0.6)
     a         *screw--> l.07
05        °h ez müemer das (0.3) BO:deteil-
          now we have to ( ) the bottom part
06        (1.5)
07        e&sO*: *um? (0.3) UF*st*ellä,&
          pu- put up like this
     f      &hand movement up and roll--&
     a      -->*  *instructions*  *screw-->
08        (0.1)*(0.4)
     a       -->*
09        und denn DIE-
          and then these
10        (0.1)
11   A    WIE ufstelle;=
          put up how
12        =gisch mr e RICHtig&sagob.
          will you give me a direction
     f                      &grabs instructions,places them on board-> l.15
13   F    =LUEG emol;=
          look
14        =eSO:;
          like this
```

```
15        (0.8)& &(0.4)&(0.4)*(0.2)#
    f     -->& &.....&PG_LH---------> l.17
    a                       *gz at image->
    fig                            #fig6/7
```
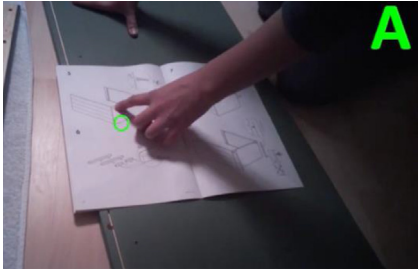

Figure 6

Figure 7

```
16    F   &da[ss DIE&#&o:be& s*ind.&]
            that these are on top
    a                      -->*
    f   &.........&PG_RH---&,,,,,,,&
    fig              #fig8/9
```
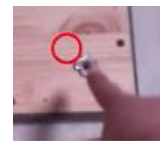

Figure 8

Figure 9

```
17    A       [dass d wolverinechralle] Obe si&nd;=
               that the wolverine claws are on top
    f                                     -->&
18    A   =GUET,
          fine
19        (0.1)
20        das (xxx)WOLverinechralle nach Obe ufstelle;
          this (xxx xxx) wolverine claws up on top
```

In this excerpt, the multimodal format "so" + pointing gesture comes as a repair solution to a problem that the in-terlocutors encounter while putting up the cupboard they are currently building. Fabienne initiates an instruction sequence at line 05 "°h ez müemer das (0.3) BO:deteil-"/ "now we have to ( ) the bottom part", which is left pragmatically, syntactically and prosodically incomplete and projects more to come. Note that the speaker uses the pronoun "wir"/ "we" rather than "du"/ "you", indexing a collaborative involvement of both speakers in the instructing sequence. At line 07 she fulfils the projection by producing an increment to her turn "eso: um (0.3) ufSTEllä,"/ "put- put up like this", followed by another turn constructional unit (Sacks et al., 1974) "und denn DIE-"/ "and then these" (l.09) that is the follow-up step once the cupboard has been "put up". At "eso"/ "like this" (l.07) Fabienne lifts her hands with her palms showing up and

makes a rolling movement indicating how to "put up" the cupboard.[3] During her embodied action, Arno's visual attention is, however, not on Fabienne's embodied action. Instead, he is looking once at the screw, then at the paper instructions and back at the screw.

Before completing the second step suggested at line 09, Arno initiates repair on her turn in line 07 "WIE ufstelle;"/ "put up how" (l.11) by partially repeating Fabienne's declarative statement and adding the interrogative word "how" (Schegloff et al., 1977), locating the trouble source in the "manner". Arno is the one holding the piece of furniture they have to put up (Fig. 7). While Fabienne, who is holding the instructions, is concerned with explaining the overall layout, being concerned with the larger activity their actions are embedded in, Arno deals with the practical problem of the piece's orientation.

Latched to his question is a request for more specific instructions "gisch mr e RICHtigsagob."/ "will you give me a direction" (l.12). Fabienne complies both verbally and in an embodied way with the request; in the middle of "RICH-tigsagob", she reaches for the instruction booklet, produces "LUEG emol"/ "look" (l.13) and places the manual in front of Arno so that he has easy access to the depictions, thus inviting him also in an embodied way "to look" at the instructions. These preparatory actions are helping Fabienne provide the repair solution both verbally, with the modal deictic "eSO:"/ "like this", and in an embodied way, by an accompanying pointing gesture at an image depicted on the instruction booklet (l.15, Fig. 7). Fabienne's pointing gesture attracts Arno's visual attention to the image she is pointing at (Fig. 6). While pointing at the image, Fabienne introduces a subordinate clause "dass DIE o:be sind."/ "that these are on top" (l.16) that verbalises the result of his action if he follows her instructions. If he puts up the board in the way she is indicating (with "eso" + a pointing gesture on the image of the booklet), then the screws she is referring to with "die"/ "these" would be on top ("obe").

The deictic "die"/ "these" is accompanied by a pointing gesture with her right hand on the screw of the cupboard while she simultaneously keeps her left hand pointed at the image (Fig. 8). In this way, Fabienne gives meaning to the visible graphic structure on the booklet by embedding the representation within its contextual field (Goodwin, 2000a). Arno holds his gaze on the image Fabienne is pointing at with her left hand and does not look at the screw she points at. But he does not need to look at Fabienne's second pointing gesture to identify "how" the cupboard has to go "up", because he has already identified the correct direction and displays this understanding in line 17, where he overlaps with Fabienne in formulating a subordinate clause that completes Fabienne's "eSO:" from line 14. However, the formats differ slightly; while Fabienne indicates the screw by pointing at it, Arno denominates the screw by mentioning the nickname they invented for the screw prior to this excerpt.[4] As already mentioned, the participants have to coordinate different *spatial frameworks* (Goodwin, 2003, p. 223) when building a cupboard with the help of an instruction booklet. This example shows how Fabienne uses the multimodal package "so" + pointing gesture to help solve their coordination problem by establishing a joint focus of attention on relevant targets.

To sum up, in this extract, the modal deictic expression "so" is combined with a pointing gesture directed at an image in the instruction booklet in order to guide the co-participant's attention to the location where relevant information to his question can be found. By combining a pointing gesture with the modal deictic expression "so"/ "like this", Fabienne can specify that the referent of her target is not the object itself, but rather how something has to look like if the action in question (putting up the cupboard) is performed correctly. As we have already specified, the instructional sequences in our data set present the particularity that the assembling process, the "how", is not presented in the instruction booklet; the images only depict the results of the actions that need to be performed in order to build the cupboard correctly. So when Fabienne points at a picture in the booklet, she does not show Arno the process of putting the cupboard up, but is pointing at the result, "how" it looks like when the cupboard is put up correctly. This instance shows that Fabienne can point at the result of an already performed action to instruct Arno to perform a prospective action.

In the next excerpt, the sisters, Anna and Bettina, are also building a cupboard together. Bettina is the one in charge of the instruction booklet (Fig. 10, on the right) and Anna (Fig. 10, on the left) is handling the physical objects. Prior to this sequence, they had to figure out which side of the board has to go "up" when laying it down on the floor. Anna has performed the manual task of laying the board on the floor, while Bettina was giving her instructions. Now, Bettina wants Anna to look at the instruction booklet and confirm that the board is in fact laying correctly on the floor. The multimodal package "so" + pointing gesture is used here as a request for confirmation.

---

[3] Her turn at line 07 is another multimodal package, where "so" is combined with an iconic gesture (see Stukenbrock, 2015), but this is not the focus of this paper.

[4] The screw has quite a peculiar form and looks like a "wolverine claw" to Arno and Fabienne (see Fig. 9).
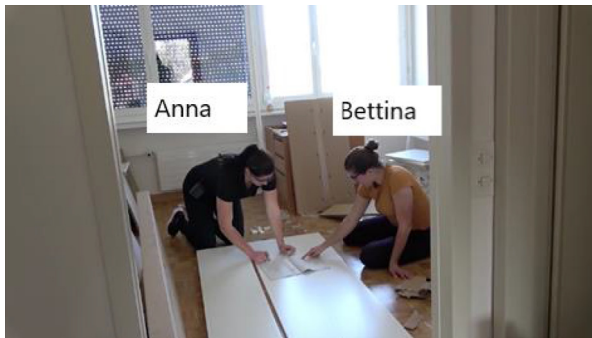
(Figure 10)

(3) Umzug_01_03:17-03:36 (A: Anna, B: Bettina)

+ for gaze by A
* for gaze by B
$ embodied action done by A
& embodied action done by B


```
01   B    &<<p> ALso;>
                So
     b    &puts instructions on board-->
02        $(0.2)&(0.2)
     b       -->&
     a    $leans forward--> 1.04
03        lueg &schnäll ob das &STIMMT;#&
          look quickly if it's right
     b         &..............&PG------&
     fig                           #fig11
```


Figure 11

```
04        =+&dass dA$s#& &SO&:is+ch;&#
          that this is like this
     b       &PG--------& &..&PG-----&
     a             -->$
     a     +below  point--------+to point-->
     fig              #fig12          #fig13
```
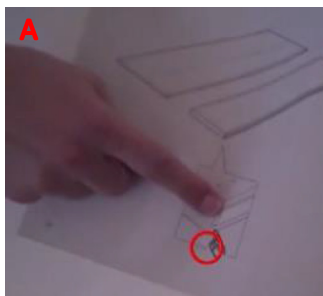

Figure 12

Figure 13

```
05   B    =&und &[O&be ]& &sO.#&+
          and on top like this
     b       &....&PG&,,,,& &PG--&
     a                       -->+
     fig                     #fig14
```
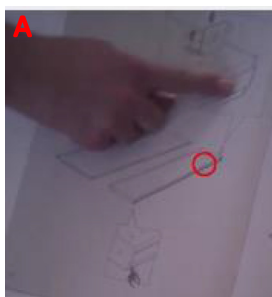

Figure 14

```
06   A           [m_H,]
                  mh
07        (0.5)

08   A  das dunkt mi GUET;
        this seems right to me
```

Bettina initiates the sequence in line 1 with "also"/ "so"[5] while placing the instructions on the board so that Anna has a good visual access to them (Fig. 11). Her embodied action along with the verbal request "lueg schnäll ob das stimmt"/ "look quickly if it's right" (l.03) invites her sister to "have a look" at the instructions, where she foregrounds for Anna, with multimodal packages containing deictic expressions and pointing gestures, the domains of scrutiny where the relevant targets can be found (Figs. 12–14).[6] The domains of scrutiny Bettina is pointing at are themselves already highlighted in the instructions,[7] which indicates that these areas are in fact important for the assembling process. In a prior sequence, not included in the transcript, Bettina had turned to the depicted elements (a hole and a smooth edge) visible in the highlighted areas for discerning which side of the board goes "up" when instructing her sister how to lay the board on the floor. In this prior sequence, it was Bettina who was reading the instruction booklet and Anna was the one handling the board. Now, Anna has to confirm, at Bettina's request, the correctness of the action she completed before, which is placing the proper side of the board upwards. Anna has had a good visual access to the "real" object in question in the prior sequence and now has to check if the relevant targets (the hole and a smooth edge) of the board match indeed the images depicted in the instructions. This is another example of the coordination

---

[5] For the use of "also" as a reformulation marker see Deppermann and Helmer (2013).
[6] We only show Anna's perspective because her sister's pointing gesture and the domain of scrutiny are better visible in this recording.
[7] Certain elements are depicted in a bigger way than others and appear in a framed box.

problems participants have to solve, as they need to coordinate different *spatial frameworks* (Goodwin, 2003, p. 223) when using an instruction booklet for their furniture assembling process. In the prior example, Fabienne used the investigated multimodal package to instruct Arno by showing him the correct outcome of the next relevant action. This example now shows how the participant who was giving the instructions during the assembling process uses the same practice, "so" + pointing gesture, but this time to request confirmation of an already performed action. This confirmation-seeking sequence happens at a point where the sisters are dealing with the practical problem of verifying the correct outcome of Anna's previously performed action. The multimodal package "so" + pointing gesture is used to establish a joint focus of visual attention, necessary for Anna to comply in lines 06 and 08 with her sister's request for confirmation. The confirmation sequence contains three turns with multimodal packages of deictic expressions coupled with pointing gestures in lines 3 ("das"/ "this"), 4 ("das"/ "this" and "so"/ "like this") and 5 ("so"/ "like this"). Whilst the pointing gesture foregrounds the domain of scrutiny where the relevant targets in question can be found, Anna specifies that the referent is either an object ("das"/ "this") or the visible result of an action performed with an object ("so"/ "like this"). By checking now, the two sisters can prevent a wrong placement of the board leading to subsequent errors and it is thus important that both check if, so far, they had followed the instructions correctly.

By now, we have seen two instances from the "relocation" data set where the interlocutors use "so" + pointing gesture. In excerpt 2, Fabienne shows how the outcome of the action of "putting the cupboard up" has to look like if the action is performed correctly by pointing at the corresponding image on the instructions booklet. Thus, she is instructing Arno how to perform the action by showing him a picture of the correct outcome of the action. In the third excerpt, Bettina invites her sister Anna to look at the instruction booklet in order to confirm if the action of laying the board down on the floor has indeed been performed correctly. Thus, the multimodal package is used both retrospectively to confirm the correctness of an already performed action and prospectively to instruct the co-interactant to perform a next action correctly.

In the last excerpt (4), Anna and Bettina need to establish how to screw a hinge onto the cupboard. In contrast to the previous excerpt, where Bettina (Fig. 15, right) was giving instructions and asking her sister for confirmation, it is now Anna (Fig. 15, left) who, after having performed the task of placing the hinge onto the board, asks her interlocutor, who is holding the instruction booklet, for confirmation before moving on with the next step of screwing the hinge onto the cupboard.
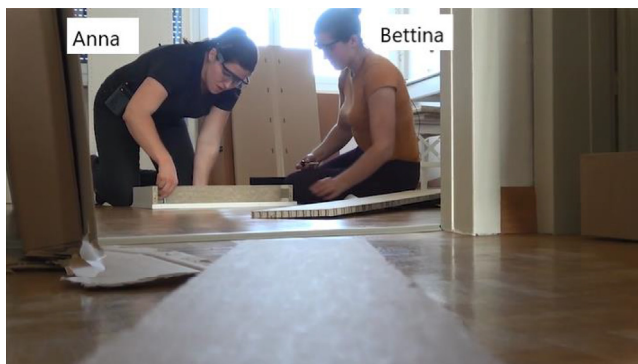
(Figure 15)

(4) Umzug 01_04:20-04:50 (A: Anna, B: Bettina)
* for gaze by A
\+ for gaze by B
\$ embodied action done by A
& embodied action done by B

```
01    A    woisch jetz de SCHRUbezieher,
           where is now the screwdriver
02         (0.5)
03    A    ah HESCH mr de schrUbezieher;=
           ah can you give me the screwdriver
04    B    =Jä; (0.1) ÄH;
           yes eh
```

```
05        (0.1)
06   A    är isch HINter dir;
          it is behind you
07        (2.3)
08   B    ((lacht))
          ((laughs))
09        (2.7)
10   A    ÄHM;
11        (2.5)
12   A    MITe::*m-_wi wie *CHUNNTS jetz ane,
          with the ho how does it have to be placed
     a         *..........*paper instructions-->
13        (0.5)$(0.4)*(0.2)*(0.1)+(0.9)#*
     a              -->*      *Bettina-----*
     a         $PG at hinge-------------->  l.16
     b                            +Anna---> l.15
     fig                                    #fig16
```
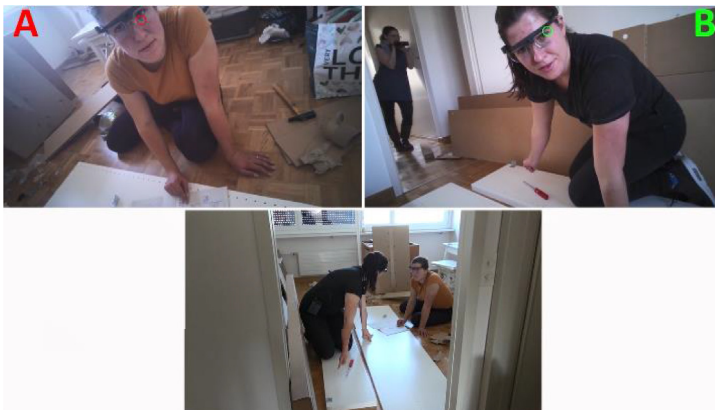

Figure 16

```
14   B    *hm,
     a    *...-->
15        (0.1)+(0.1)*(0.1)+(0.1)
     a           -->*hinge--> l.23
     b      -->+         +A's arm direction-->
16   A    eSO; #(0.2)+O$der-
          like that or
     a                -->$
     b            -->+hinge--> l.22
     fig      #fig17
```
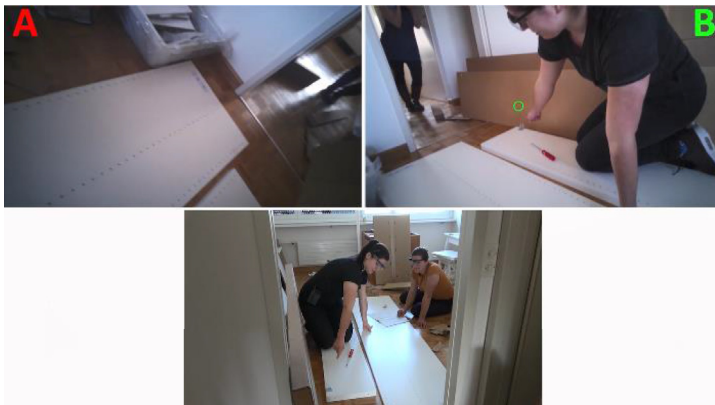

Figure 17

```
17        (0.6)
18   B    $JA;
          yes
     a    $OM⁸--> l.22
19        (0.3)
20   A    e[So,
          like that
21   B    #[SO;
           like that
```

```
     fig #fig18
```

```
     Figure 18
22        (0.1)+(0.3)+$
     b       -->+
     a               -->$
23   A    +SO,*
           like that
     a       -->*
     b    +paper instructions-->>
24        *(0.6)
     a    *...-->
25   B    äh?* (.) WART; ((lacht))
          eh wait ((laughs))
     a    -->*paper instructions--> l.29
26        (0.3)
27   A    JA; (0.3) SO.=
          yes like that
28   B    =JA; (0.4) JA; SO;
          yes yes like that
29        (1.1)*(0.1)
     a       -->*
30   A    <<p> oKE::;>
          okay
```
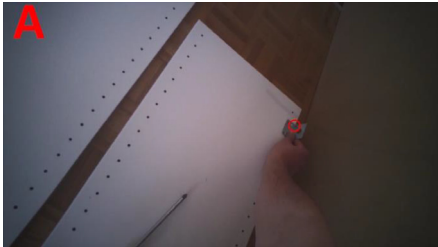
In this excerpt, Bettina has once again assumed the role of reading the instructions on the paper while Anna is dealing with the physical objects. At the start of this sequence, Anna tries to summon her sister's attention with a request for a screwdriver "woisch jetz de SCHRUbezieher,"/ "where is now the screwdriver" (l.01). Once the request is granted, Anna establishes a new interactional focus by asking her sister, who is holding the paper instructions, for instructions on how to screw the hinge onto the cupboard "MITe::m-_wi wie CHUNNTS jetz ane,"/ "with the ho how does it have to be placed" (l.12)[9]. At the end of "MITe::m", she shifts her gaze from the hinge to the paper instructions that are laid out in front of Bettina. While keeping her gaze at the instructions, she points with her right-hand index finger at the hinge and shifts her gaze to her sister. After 0.1 seconds Bettina meets Anna's gaze and the interlocutors are mutually monitoring each other (Fig. 16). However, instead of providing the instructions as requested by Anna, she initiates repair "hm," (l.14) (Schegloff, 1992, 2000b). Anna then shifts her gaze to the hinge, while maintaining the pointing gesture at it, and produces "eSO; (0.2) Oder[10]-"/ "like that or" (l.16) (Fig. 17). By pointing at the hinge, she invites Bettina

to identify the target in the domain of scrutiny and confirm the correctness of her prior embodied action, i.e. the positioning of the hinge on the cupboard. Bettina understands it to be a confirmation seeking device (Stivers and Rossano, 2012); she directs her gaze to the hinge following Anna's pointing gesture and provides a type-conforming response "JA;"/ "yes" (l.18) (Raymond, 2003) followed by a confirmation "SO"/ "like that" (l.21). In overlap with Bettina's "JA"/ "yes" at line 18, Anna moves the hinge slightly, so that the holes of the hinge fit the holes on the board more accurately, without changing the initial position of the hinge (Fig. 18). The second "eSO" in line 20 is seeking confirmation to her current embodied action, her manipulation of the hinge. Since Anna's second "eSO" (l.20) comes in overlap with Bettina's confirming "SO" (l.21), Anna repeats her prior turn "SO" (l.23)[11] and turns her gaze to the paper instructions. However, Bettina does not provide the confirmation that Anna has requested. Instead, she initiates self-repair (Schegloff et al., 1977) "äh (.) WART;"/ "eh wait" (l.25) on her prior turn "SO;" (l.21), having already turned her gaze back to the instructions, and turns the page to the correct one. Anna then self-confirms her prior embodied action at line 27 "JA; (0.3) SO."/ "yes like that", after monitoring the image that portrays the correct position of the hinge in the paper instructions that Bettina is holding, and Bettina agrees with her "JA; (0.4) JA; SO;"/ "yes yes like that" (l.28).

With the use of "SO" and a pointing gesture at the hinge, Anna attempts at mobilising her sister's co-orientation to the result of an already performed action, her positioning of the hinge on the board and so lays the ground for their joint attention towards its correct positioning. The excerpt presents a complex sequence where the modal deictic expression "so" is either used for confirmation seeking (lines 16, 20 and 23) or for (self-)confirmation (lines 21, 27 and 28). The confirmation seeking turns are produced by Anna, the sister who is performing the manual task of positioning the hinge correctly onto the board, whereas the confirming turns are produced either by Anna or by Bettina. For the confirmation seeking instances, we see that the first time Anna produces her request for confirmation containing the modal deictic "so", she couples it with a pointing gesture. This instance shows that the multimodal package "so" + pointing gesture can be used to direct the addressee's visual attention to a domain of scrutiny where the outcome of an already performed action has to be taken into account in order to perform the relevant next action (in this case, confirming the correct positioning of the hinge). Anna's second request for confirmation containing the modal deictic "so" (lines 20 and 23) presents another multimodal package, where "so" is coupled with an object manipulation in which Anna is performing the action (positioning the hinge correctly onto the cupboard) for which she wants a confirmation from her sister. This instance is a case of the multimodal package "so" + performance, presented at the end of Section 2 in this paper, where the addressee needs to monitor the action in progress in order to "fill" the lexico-semantic content of the modal deictic expression (Stukenbrock, 2014). All in all, the confirmation seeking turns performed by Anna contain two different multimodal packages, both produced with the modal deictic expression "so", which underlines that the target in question is either an already performed action ("so" + pointing gesture) or a performance in progress ("so" + object manipulation). In both multimodal packages, the deictic expression is used to highlight the qualitative aspect of the action in question, "how" something was done (Stukenbrock, 2014, p. 87).

When the modal deictic "so" is not accompanied by a pointing gesture or an object manipulation, it is used as a confirmation token, either produced by Anna when self-confirming or by Bettina when she complies with Anna's requests for confirmation. Requests for confirmation are "polar questions" that make a confirming or disconfirming response conditionally relevant (Rossano, 2010; Stivers and Enfield, 2010) and can be answered by interjections like "yes" (lines 18 and 28) or by repetitions (lines 21, 27 and 28) (Enfield et al., 2010, 2019; Stivers and Enfield, 2010).

With the confirmation sequence presented above, Anna is ensuring that the result of her performed action is correct before initiating the next step, that is screwing the hinge onto the board. Thus, "so" + pointing gesture is used to request a confirmation of an already performed action and for establishing co-orientation and coordination with regard to the overarching project of building the cupboard. Since each step in the process is dependent on the correct implementation of the prior one, the confirmation of the correct outcome of the prior action is important in order to ensure that the following step can also be performed successfully.

5. Discussion

The cases we have studied in this article present a systematic use of the particle "so" accompanied by a pointing gesture. What transverses the examples in the empirical analysis is that the point establishes a domain of scrutiny where the visible outcome of an action has to be taken into account, and the verbal component "so" emphasizes that the target has to be perceived in a certain way, foregrounding the qualitative aspect. While in some instances the practice is used prospectively, in others it is used retrospectively. For instance, in excerpt 2, Fabienne employs the multimodal package "so" + pointing gesture to a domain of scrutiny within the instruction booklet to instruct Arno, showing him how the outcome of the action she wants him to carry out has to look like if the action is performed correctly. The multimodal package is employed to illustrate prospectively the correct outcome of an action that the participants still have to achieve and is therefore forward-oriented. Excerpt 3, on the other hand, illustrates an instance where "so" + pointing gesture is employed to ensure the correct outcome of an action that has just been performed. Bettina is in charge of reading the instruction booklet and invites her sister Anna, who has assumed the practical role of dealing with the furniture, to look at the instruction booklet too, in order to confirm that the action she has just performed has been indeed carried out correctly. And in the last excerpt, Anna wants a confirmation of the correct outcome of an action she had just performed, placing the hinge onto the board in the right way. All of the cases show that the multimodal package "so" + pointing gesture serves either to retrospectively confirm the correctness of a prior action or prospectively to

---

[11] Partial or full repeats are a common characteristic of overlapping talk (Schegloff, 2000a).

ensure that the next action can be performed successfully. By confirming the correctness of a prior action or checking the outcome of a prospective action, interlocutors are in a position to move from prior to next-positioned matters. This is especially important in settings where there is a common practical goal that has to be achieved collaboratively by the participants by following step-by-step procedures. In our data, participants use this practice frequently when building furniture together with the help of an instruction booklet, where they need to solve complex coordinational problems of different *spatial frameworks* (Goodwin, 2003, p. 223). By using a pointing gesture, the participants can direct the addressee's visual attention to a domain of scrutiny where additional information necessary for the task at hand can be found. They either point to the "real" objects they need to manipulate when building furniture or to the images depicted on the instruction booklet. The deictic expression "so" in these instances functions as a *summons for gaze* (Stukenbrock, 2018a) laying the grounds for a successful establishment of joint attention, a basic requirement for successfully performing a joint activity (Clark, 1996).

Accordingly, the practice is also found recurrently in settings where instructional actions occur in environments in which the participants pursue a collaborative activity, for instance when participants have to cook something together or figure out the rules of a game together. The reason for the occurrence of this practice in such settings is that there is an overarching project involved and the "end result" depends on the correct implementation of smaller steps. The practice "so" + pointing at the outcome of an action serves to coordinate the individual actions of the participants involved so that they, in turn, can contribute to the progressivity of their overall "project". By confirming the correctness of their prior or next actions, they can ensure the successful achievement of their collaborative project.

With regard to the temporality of the components in the multimodal package, the examples presented above show that there is a variation in the emergence of the gesture and the correspondent modal deictic. In excerpt 3, Bettina has already placed the instructions in front of Anna before producing the deictic expressions, which are relatively precisely timed with the pointing gestures; the deictics either slightly precede the apex ("das"/ "this" in line 3 and "so"/ "like this" in line 4) or are produced during the apex ("das"/ "this" in line 4 and "so"/ "like this" in line 5). Similarly, in excerpt 4, the deictics (lines 16 and 20) are produced during the apex of the pointing gesture or whilst manipulating the object in question. In excerpt 2, on the contrary, Fabienne produces the deictic expression whilst placing the instruction booklet in front of her interlocutor, allowing him to have good visual access to the domain of scrutiny, before performing her pointing gesture. The participants do not express the verbal component of the multimodal package during the apex of the pointing gesture but locally adapt the timing, taking the addressee's visual attention into account. This is in line with Hindmarsh and Heath's (2000, p. 1863) observation, that "the relationship between the utterance – in particular the deictic term or expression – and the gesture […] is subject to a range of local contingencies".

Taking into account the results presented above, by using "so" concurrently with resources generally considered to be in the service of local deixis (Fillmore, 1997), such as pointing gestures (Goodwin, 2000b, 2003; Kendon, 2004; Kita, 2003; Fricke, 2007; Stukenbrock, 2015), participants are directing their addressee's attention to a domain of scrutiny, the *place* where they should look for additional information to fill the lexico-semantic content of the modal deictic expression "so". This observation is on par with Harweg's (1990) theoretical positioning on the use of pointing gestures in the service of modal deixis and complements the description of multimodal packages with "so" identified so far (Stukenbrock, 2014), where the action in progress has to be monitored to "fill" the lexico-semantic content of "so". However, while in Stukenbrock's (2014) study "so" functions as a con-textualisation cue or a flag (Streeck, 2002) that attracts the addressee's attention to the gesture or bodily performance, what our examination of the multimodal package further illustrates is that "so" combined with a pointing gesture directs the addressee's visual attention to the "outcome" of this performance. So, unlike local deixis where the focus of the pointing gesture is a feature in the environment, such as an object, with this practice the focus remains on the performance of an action.

On the whole, the practice examined has been shown to be an efficient mechanism for establishing joint attention. Joint attention is a basic requirement for the successful accomplishment of a collaborative activity (Clark, 1996). While the pointing gesture directs the addressee's visual attention on a domain of scrutiny, the deictic expression "so" foregrounds that the target of the point has to be perceived in a certain way. In other words, "so" indicates how the pointed-at elements have to be "understood", which in the excerpts examined here illustrate the outcome of an action. In combination then, the two elements of the multimodal package, the deictic expression "so" and the gesture, complement each other to achieve a joint focus of attention and an intersubjective understanding, prerequisites for achieving coordination and co-operation (Hausendorf, 2013) and a joint course of action (Goodwin, 2003).

What our examination of the multimodal package "so" + pointing gesture further demonstrates is that co-orientation may be established with respect to actions and not objects within everyday activities. Co-participants display their orientation to actions that need to be performed with respect to the activity at hand and direct their interlocutor's attention to the outcome of an action that is relevant for their collaborative work. By pointing at the outcome of an action, participants can, for one, check whether something has been done correctly before moving to the next task or instruct co-participants to perform a next action adequately by showing them how the correct result of the next action looks like. Those actions highlighted by "so" + pointing are necessary prerequisites for the successful accomplishment of their task and need to be performed in a certain way.

To conclude with, we would like to highlight an important matter to resolve for future studies on this practice. In some of the extracts we have examined in this paper, there are follow-up actions by the speaker who produces "so" + pointing gesture. These follow-up actions are of different types, one is a verbalization of the action referred to by "so" (excerpt 1), and another is a verbalisation of the result of his action if the addressee follows the speaker's instructions (extract 2). However, such follow-up actions do not arise recurrently in the data. In some examples from our data set the verbalisation even comes before the investigated practice and other times it is produced by the addressee and not the speaker. The instances in our data set do not show a clear pattern, therefore more research is needed in order to address the interactional function of these turns

satisfactorily. Future research could build on the results presented here and investigate the role of these formulations by speaker and addressee in combination with the multimodal package "so" + pointing gesture.

Declaration of competing interest

The authors declare no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Appendix A

GAT 2 transcription conventions (GAT, Selting et al., 2009; for the English translation see Couper-Kuhlen and Barth-Weingarten, 2011)

Sequential structure

[ ]	Overlap and simultaneous talk
=	Fast, immediate continuation with a new turn or segment (latching)

In- and outbreaths

h̊ / h̊	In-/outbreaths of approx 0.2--0.5 s duration
hh̊ / hh̊	In-/outbreaths of appr. 0.5--0.8 s duration
hhh̊ / hhh̊	In-/outbreaths of appr. 0.8--1.0 s duration

Pauses

(.)	Micro pause, estimated, up to 0.2 s duration appr.
(0.5)/(2.0)	Measured pause of appr. 0.5 / 2.0 sec. duration (to tenth of a second)

Other segmental conventions

and_uh	Cliticizations within units
uh, uhm, etc.	Hesitation markers, so-called "filled pauses"
:	Lengthening, by about 0.2-0.5 sec.
::	Lengthening, by about 0.5-0.8 sec.
:::	Lengthening, by about 0.8-1.0 sec.
ʔ	Cut-off by glottal closure

Laughter and crying

haha hehe hihi	Syllabic laughter
((laughs)) ((cries))	Description of laughter and crying
<<laughing> >	Laughter particles accompanying speech with indication of scope
<<:-)> so>	Smile voice

Continuers

hm, yes, no, yeah	Monosyllabic tokens
hm_hm, ye_es, no_o	Bi-syllabic tokens
?hm?hm	With glottal closure, often negating

Accentuation

SYLlable	Focus accent
sYllable	Secondary accent
!SYL!lable	Extra strong accent

Final pitch movements of intonation phrases

?	Rising to high
,	Rising to mid
--	Level
;	Falling to mid
.	Falling to low
<<surprised> >	Interpretive comment with indication of scope
<<f> >	Forte, loud
<<ff> >	Fortissimo, very loud
<<p> >	Piano, soft
<<pp> >	Pianissimo, very soft
<<all> >	Allegro, fast
<<len> >	Lento, slow
<cresc> >	Crescendo, increasingly louder
<<dim> >	Diminuendo, increasingly softer
<<acc> >	Accelerando, increasingly faster

Other conventions

(xxx), (xxx xxx)	One or two unintelligible syllables
(may i)	Assumed wording
((. . .))	Omission in transcript
●	

Appendix B

Embodied actions are transcribed according to the following conventions developed by Lorenza Mondada (see Mondada (2018) for a conceptual discussion).

* *	Descriptions of embodied actions are delimited between
+ +	two identical symbols (one symbol per participant and per type of action)
ΔΔ	that are synchronized with correspondent stretches of talk or time indications.
*--->	The action described continues across subsequent lines

---->*	until the same symbol is reached.
>>	The action described begins before the excerpt's beginning.
--->>	The action described continues after the excerpt's end.
.....	Action's preparation.
----	Action's apex is reached and maintained.
,,,,,	Action's retraction.
ric	Participant doing the embodied action is identified in small caps in the margin.
fig	The exact moment at which a screen shot has been taken
#	is indicated with a sign (#) showing its position within the turn/a time measure.

References

Atkinson, J.M., Heritage, J. (Eds.), 1984. Structures of Social Action: Studies in Conversation Analysis. Cambridge University Press, Cambridge.

Bangerter, A., 2004. Using pointing and describing to achieve joint focus of attention in dialogue. Psychol. Sci. 15 (6), 415–419.

Bruner, J.S., 1995. From joint attention to the meeting of minds: an introduction. In: Moore, C., Dunham, P. (Eds.), Joint Attention: its Origins and Role in Development. Psychology Press, New York and London, pp. 1–14.

Clark, H.H., 1996. Using Language. Cambridge University Press, Cambridge.

Clark, H.H., 2003. Pointing and placing. In: Kita, S. (Ed.), Pointing. Where Language, Culture, and Cognition Meet. Lawrence Erlbaum, Mahwah, NJ, pp. 243–268.

Clausen, Y., Scheffler, T., 2020. A corpus-based analysis of meaning variations in German tag questions: evidence from spoken and written conversational corpora. Corpus Linguist. Linguistic Theory. https://doi.org/10.1515/cllt-2019-0060. Published online ahead of print 2020.

Couper-Kuhlen, E., Barth-Weingarten, D., 2011. A system for transcribing talk-in-interaction, translated and adapted for English by Elizabeth Couper-Kuhlen and Dagmar Barth-Weingarten. In: Gesprächsforschung - Online-Zeitschrift zur verbalen Interaktion, vol. 12, pp. 1–51.

Deppermann, A., 2008. Gespräche Analysieren: Eine Einführung. Verlag für Sozialwissenschaften, Wiesbaden.

Deppermann, A., 2013. Multimodal interaction from a conversation analytic perspective. J. Pragmat. 46 (1), 1–7.

Deppermann, A., Helmer, H., 2013. Zur Grammatik des Verstehens im Gespräch: inferenzen anzeigen und Handlungskonsequenzen ziehen mit also und dann. Z. Sprachwiss. 32 (1), 1–39.

De Stefani, E., Gazin, A.D., 2014. Instructional sequences in driving lessons: mobile participants and the temporal and sequential organization of actions. J. Pragmat. 65, 63–79.

Diessel, H., 2006. Demonstratives, joint attention, and the emergence of grammar. Cognit. Ling. 17 (4), 463–489.

Drake, V., 2016. German questions and turn-final *oder*. In: Gesprächsforschung: Online-Zeitschrift zur verbalen Interaktion, vol. 17, pp. 168–195.

Drew, P., 2004. Conversation analysis. In: Fitch, K., Sanders, R. (Eds.), Handbook of Language and Social Interaction. Lawrence Erlbaum, New Jersey, pp. 71–102.

Eisenberg, P., 1994. Grundriss der deutschen Grammatik. Metzler, Stuttgart [u.a.].

Ekström, A., 2012. Instructional Work in Textile Craft: Studies of Interaction, Embodiment and the Making of Objects. PhD thesis. Stockholm University, Stockholm.

ELAN (Version 5.6), 2019. [Computer Software]. Nijmegen: Max Planck Institute for Psycholinguistics. Available at: https://tla.mpi.nl/tools/tla-tools/elan/. (Accessed 30 November 2020).

Enfield, N.J., Stivers, T., Levinson, S.C., 2010. Question-response sequences in conversation across ten languages: an introduction. J. Pragmat. 42 (10), 2615–2619.

Enfield, N.J., et al., 2019. Polar answers. J. Linguist. 55 (2), 277–304.

Eriksson, M., 2009. Referring as interaction: on the interplay between linguistic and bodily practices. J. Pragmat. 41 (2), 240–262.

Fillmore, C., 1997. Lectures on Deixis. CSLI Publications, Stanford.

Fricke, E., 2007. Origo, Geste und Raum: Lokaldeixis im Deutschen. de Gruyter, Berlin, New York.

Garfinkel, H., 2002. Instructions and instructed actions. In: Rawls, A.W. (Ed.), Ethnomethodology's Program: Working Out Durkheim's Aphorism. Rowman and Littlefield Publishers Inc., Lanham, pp. 197–218.

Goodwin, C., 2000a. Action and embodiment within situated human interaction. J. Pragmat. 32 (10), 1489–1522.

Goodwin, C., 2000b. Pointing and the collaborative construction of meaning in Aphasia. Tex. Ling. Forum 43, 67–76.

Goodwin, C., 2003. Pointing as situated practice. In: Kita, S. (Ed.), Pointing. Where Language, Culture, and Cognition Meet. Lawrence Erlbaum, Mahwah, NJ, pp. 217–241.

Hanks, W.F., 1992. The indexical ground of deictic reference. In: Duranti, A., Goodwin, C. (Eds.), Rethinking Context: Language as an Interactive Phenomenon. Cambridge University Press, Cambridge, pp. 43–76.

Harweg, R., 1990. Studien zur Deixis (Bochumer Beiträge zur Semiotik 25). Brockmeyer, Bochum.

Hausendorf, H., 2003. Deixis and speech situation revisited: the mechanism of perceived perception. In: Lenz, F. (Ed.), Deictic Conceptualisation of Space, Time and Person. John Benjamins, Amsterdam, pp. 249–269.

Hausendorf, H., 2013. On the interactive achievement of space – and its possible meanings. In: Auer, P., et al. (Eds.), Space in Language and Linguistics: Geographical, Interactional and Cognitive Perspectives. De Gruyter, Berlin, pp. 276–303.

Haviland, J., 1993. Anchoring, iconicity, and orientation in Guugu Yimithirr pointing gestures. J. Ling. Anthropol. 3 (1), 3–45.

Haviland, J., 2000. Pointing, gesture spaces, and mental maps. In: McNeill, D. (Ed.), Language and Gesture. Cambridge University Press, Cambridge, pp. 13–46.

Heath, C., 1986. Body Movement and Speech in Medical Interaction. Cambridge University Press, Cambridge.

Hindmarsh, J., Heath, C., 2000. Embodied reference: a study of deixis in workplace interaction. J. Pragmat. 32 (12), 1855–1878.

Kendon, A., 2004. Gesture: Visible Action as Utterance. Cambridge University Press, Cambridge.

Kita, S., 2003. Pointing: a foundational building block in human communication. In: Kita, S. (Ed.), Pointing. Where Language, Culture, and Cognition Meet. Lawrence Erlbaum, Mahwah, NJ, pp. 1–8.

Levinson, S.C., 1983. Pragmatics. Cambridge University Press, Cambridge.

Levinson, S.C., 2004. Deixis. In: Horn, L.R., Ward, G. (Eds.), The Handbook of Pragmatics. Blackwell Publishing, Malden/Oxford, pp. 97–121.

Lindwall, O., Lymer, G., Greiffenhagen, C., 2015. The sequential analysis of instruction. In: Markee, N. (Ed.), The Handbook of Classroom Discourse and Interaction. John Wiley and Sons, Inc., Hoboken, NJ, pp. 142–157.

McNeill, D., 1992. Hand and Mind: what Gestures Reveal about Thought. The University of Chicago Press, Chicago/London.

McNeill, D., 2005. Gesture & Thought. University of Chicago Press, Chicago.

Mondada, L., 2013. Video as a tool in the social sciences. In: Müller, C., et al. (Eds.), Body-Language-Communication: an International Handbook on Multimodality in Human Interaction. De Gruyter Mouton, Berlin/Boston, pp. 982–992.

Mondada, L., 2014a. Pointing, talk, and the bodies: reference and joint attention as embodied interactional achievements. In: Seyfeddinipur, M., Gullberg, M. (Eds.), From Gesture in Conversation to Visible Action as Utterance: Essays in Honor of Adam Kendon. John Benjamins, Amsterdam/Philadelphia, pp. 95–124.

Mondada, L., 2014b. The local constitution of multimodal resources for social interaction. J. Pragmat. 65, 137–156.

Mondada, L., 2018. Multiple temporalities of language and body in interaction: challenges for transcribing multimodality. Res. Lang. Soc. Interact. 51 (1), 85–106.

Mondada, L., 2019. Transcription Conventions. Available at: https://344cc026-c96f-49aa-b4bc-071b454d3061.filesusr.com/ugd/ba0dbb_986ddd4993a04a57acf20ea06e2b9a34.pdf. (Accessed 3 December 2020).

Mondada, L., Schmitt, R., 2010. Zur multimodalität von Situationseröffnungen. In: Mondada, L., Schmitt, R. (Eds.), Situationseröffnungen. Zur Multimodalen Herstellung Fokussierter Interaktion. Tübingen, Narr, pp. 7–52.

Mundy, P., Newell, L., 2007. Attention, joint attention, and social cognition. Curr. Dir. Psychol. Sci. 16 (5), 269–274.

Ningelgen, J., Auer, P., 2017. Is there a multimodal construction based on non-deictic *so* in German? Linguistics Vanguard 3 (s1), 1–14.

Raymond, G., 2003. Grammar and social organization: yes/no interrogatives and the structure of responding. Am. Socio. Rev. 68 (6), 939–967.

Rossano, F., 2010. Questioning and responding in Italian. J. Pragmat. 42 (10), 2756–2771.

Sacks, H., Schegloff, E.A., Jefferson, G., 1974. A simplest systematics for the organization of turn-taking for conversation. Language 50 (4), 696–735.

Scaife, M., Bruner, J.S., 1975. The capacity for joint visual attention in the infant. Nature 253, 265–266.

Schegloff, E.A., 1992. Repair after next turn: the last structurally provided defence of intersubjectivity in conversation. Am. J. Sociol. 97 (5), 1295–1345.

Schegloff, E.A., 2000a. Overlapping talk and the organization of turn-taking for conversation. Lang. Soc. 29 (1), 1–63.

Schegloff, E.A., 2000b. When "others" initiate repair. Appl. Linguist. 21 (2), 205–243.

Schegloff, E.A., 2007. Sequence Organization in Interaction: A Primer in Conversation Analysis. Cambridge University Press, Cambridge.

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.

Selting, M., et al., 2009. Gesprächsanalytisches transkriptionssystem 2 (GAT 2). In: Gesprächsforschung - Online-Zeitschrift zur Verbalen Interaktion, vol. 10, pp. 353–402. www.gespraechsforschung-ozs.de.

Stivers, T., Enfield, N.J., 2010. A coding scheme for question-response sequences in conversation. J. Pragmat. 42 (10), 2620–2626.

Stivers, T., Rossano, F., 2012. Mobilizing response in interaction: a compositional view of questions. In: De Ruiter, J.P. (Ed.), Questions: Formal, Functional and Interactional Perspectives. Cambridge University Press, Cambridge, pp. 58–80.

Stivers, T., Sidnell, J., 2005. Introduction: multimodal interaction. Semiotica 156 (1/4), 1–20.

Streeck, J., 2002. Grammars, words, and embodied meanings: on the uses and evolution of *so* and *like*. J. Commun. 52 (3), 581–596.

Streeck, J., 2013. Praxeology of gesture. In: Müller, C., et al. (Eds.), Body-Language-Communication: an International Handbook on Multimodality in Human Interaction. De Gruyter Mouton, Berlin/Boston, pp. 674–688.

Streeck, J., Goodwin, C., LeBaron, C. (Eds.), 2011. Embodied Interaction: Language and Body in the Material World. Cambridge University Press, Cambridge.

Stukenbrock, A., 2009. Referenz durch Zeigen: zur Theorie der Deixis. Dtsch. Sprache 37 (4), 289–316.

Stukenbrock, A., 2010. Überlegungen zu einem multimodalen Verständnis der gesprochenen Sprache am Beispiel deiktischer Verwendungsweisen des Ausdrucks *so*. InLiSt-Interact. Linguist. Struct. 47, 165–193.

Stukenbrock, A., 2014. Take the words out of my mouth: verbal instructions as embodied practices. J. Pragmat. 65, 80–102.

Stukenbrock, A., 2015. Deixis in der face-to-face-Interaktion. De Gruyter, Berlin/München-Boston.

Stukenbrock, A., 2018a. Blickpraktiken von SprecherInnen und AdressatInnen bei der Lokaldeixis: mobile Eye Tracking-Analysen zur Herstellung von joint attention. In: Gesprächsforschung-Online-Zeitschrift zur verbalen Interaktion, vol. 19, pp. 132–168.

Stukenbrock, A., 2018b. Forward-looking: where do we go with multimodal projections? In: Deppermann, A., Streeck, J. (Eds.), Modalities and Temporalities: Convergences and Divergences of Bodily Resources in Interaction. John Benjamins, Amsterdam, pp. 31–68.

Tomasello, M., 1995. Joint attention as social cognition. In: Moore, C., Dunham, P.J. (Eds.), Joint Attention: its Origins and Role in Development. Lawrence Erlbaum Associates, Inc., New York/ London, pp. 103–130.

Zifonun, G., Hoffman, L., Strecker, B., 1997. Grammatik der deutschen Sprache. De Gruyter, Berlin/ New York.

Angeliki Balantani is a postdoctoral research fellow on the project "Deixis and Joint Attention: Vision in Interaction (DeJA-VI)" funded by the Swiss National Science Foundation (SNSF) where she investigates the role of gesture and gaze in the use of the deictic expression "SO" with the help of mobile eye-tracking glasses. She recently completed her PhD at the University of Essex under the supervision of Dr. Rebecca Clift where she investigated responses to informings in Greek talk-in-interaction.

Stefanie Lázaro is conducting her PhD studies at the University of Lausanne (Switzerland) in a project on "Deixis and Joint Attention: Vision in Interaction (DeJA-VI)" funded by the Swiss National Science Foundation (SNSF). Her research interests include sociolinguistics and conversation analysis and her PhD thesis focuses on pointing practices and gaze patterns when participants are establishing a joint focus of attention.