

Cornelia Ebert, Giovanna Pirillo & Sebastian Walter

2022

The Role of Gesture-Speech Alignment
for Gesture Interpretation



in

Robin Hörnig, Sophie von Wietersheim, Andreas Konietzko & Sam Featherston (eds.),
Proceedings of Linguistic Evidence 2020: Linguistic Theory Enriched by Experimental Data
pp. 65–77

Tübingen: University of Tübingen

<https://publikationen.uni-tuebingen.de/xmlui/handle/10900/119301>

The Role of Gesture-Speech Alignment for Gesture Interpretation

Cornelia Ebert, Giovanna Pirillo & Sebastian Walter

University of Frankfurt/Main

ebert@lingua.uni-frankfurt.de, pirillo@em.uni-frankfurt.de, s.walter@em.uni-frankfurt.de

1 Introduction

Ebert & Ebert (2014) argue that the semantic contribution of a co-speech gesture is partly determined by the temporal alignment of gesture and speech. They claim that an iconic gesture (e.g., a manual gesture using both index fingers to draw a rectangular shape in the gesture space in front of the speaker's torso) that accompanies an indefinite (e.g., *a window*) makes a different contribution than the same gesture that accompanies a definite (e.g., *the window*) or one that is temporally aligned only with the NP complement (e.g., *window*). Crucially, it is argued that gesture alignment with a full DP (e.g., *a window* or *the window*) gives rise to what we call an interpretation of *comparison*, whereas alignment with only the NP complement (*window*) triggers an interpretation of *exemplification*. Gestures are interpreted as rigid designators to an intended gesture referent. In the comparison interpretation, the gestural individual concept (which is a rigid designator) is compared to the accompanying speech concept in predefined ways that are dependent on the nature of the accompanying speech expression. In case of an indefinite (*a window*), it is required that the gesture concept be similar (in certain contextually given ways) to the speech concept. In the exemplification interpretation, the speaker simply exemplifies the NP concept, here: *window*, and illustrates a prototypical type of window.

In this paper, we report the findings of an online rating study¹ that we conducted in order to find experimental support for the semantic alignment rules postulated by Ebert & Ebert (2014), namely that an alignment of gesture with NP alone triggers exemplification readings and an alignment of gesture with DP triggers comparison readings (here: in the case of indefinites). In this experiment, we compared potentially referring DPs, here: indefinites (*a window*), with non-referring DPs such as *no window*, which were accompanied by iconic gestures, and created items where we artificially synchronized gesture and speech with the help of a video tool such that a given gesture was aligned with an NP in one condition and with a DP in another. The results of this study corroborate Ebert & Ebert's (2014) hypothesis that NP alignment gives rise to exemplification readings and DP alignment yields comparison readings. We interpret this finding as a general experimental confirmation of the often-made claim that the semantic contribution of a gesture is determined by its accompanying speech expression.

This paper is structured as follows: In Section 2, we introduce some background on gesture semantics that we draw upon in the described experiment. Subsection 2.1 discusses seminal work by Fricke (2012), where she introduces the distinction between what she calls *object-* and *interpretant-related* gesture interpretations (German: "*objekt-*" and "*interpretantenbezogene Interpretationen*", Fricke, 2012: 230ff). This distinction comes very close to the distinction we make between comparison- and exemplification-based interpretations of gestures. We then briefly introduce Ebert & Ebert's (2014) account of gesture semantics and the semantic

¹ The study was conducted as an online study due to the Covid-19 pandemic, which did not allow for testing in the lab.

temporal alignment effects they propose in Subsection 2.2. In Subsection 2.3, we relate Fricke’s (2012) distinction to ours and discuss some consequences of Ebert & Ebert’s (2014) alignment rules for the alignment of gestures with quantifiers other than indefinites or definites, in particular negative quantifiers such as *no*. Section 3 provides an overview of our experimental study. Section 4 discusses our findings.

2 Gesture Semantics

In this paper, we are concerned with iconic gestures only. An iconic gesture is a hand movement that “seem[s] to bear a formal similarity to some aspect of the situation described by the accompanying speech” (McNeill & Levy, 1982: 273). As the following example shows, such iconic gestures are able to modify a verbal utterance (cf. Fricke, 2012: 220).

- (1) a. I want a pullover.
 b. I want a pullover. (+ iconic gesture that imitates a V-neck)

If a speaker utters the sentence in (1a), she does not specify the kind of pullover she wants. However, if the same sentence is uttered and accompanied by an iconic gesture as in (1b), the speaker expresses that she wants a V-necked pullover.

This shows that gestures do contribute to the overall meaning of a message (McNeill, 1992; Kendon, 2004) and the combined meaning of such gestures and speech has been a topic of interest in different fields such as psychology, semiotics, or robotics for decades. However, gesture semantics has only very recently been considered within formal semantic theory. Different formal semantic analyses for the treatment of co-speech gestures have been suggested in the literature (Lascarides & Stone, 2009; Ebert & Ebert, 2014; Schlenker, 2018; Esipova, 2019; a.o.) with various proposals as to how the meaning of a gesture contributes to the meaning of a multimodal utterance.

Note that (1a) and (1b) are both ambiguous between a specific and an unspecific reading of the indefinite. This means that the iconic gesture in (1b) does not force the indefinite into one of its readings. It can still be read non-specifically: the speaker wants a pullover with a V-neck, any exemplar will do. Additionally, and not surprisingly, it can also mean that the speaker has a certain pullover with a V-neck in mind; and she wants this exact pullover. This corresponds to the specific reading.

2.1 Two Interpretation Strategies of Co-Speech Gestures

Fricke (2012) points out that gestures accompanying (potentially referential) nominal phrases can receive different interpretations: they can either enrich or further illustrate the meaning of the referential object that the nominal phrase refers to or they can make reference to the underlying NP concept. Consider the following example for illustration.

- (2) My living room has only [one window].
 + *iconic rectangular gesture*

In (2), the speaker makes a gesture drawing a rectangular shape in front of her torso while uttering *one window*. Here, a referential reading of the indefinite, where the speaker aims to illustrate the shape of her actual window in the living room, is very prominent. Fricke refers to this reading as one that is triggered by an *object-related* (German: *objektbezogene*, cf. Fricke, 2012: 230ff) interpretation of the accompanying gesture.

There is another reading a gesture can give rise to, namely one that, according to Fricke (2012), is triggered by an *interpretant-related* (German: *interpretantenbezogene*) interpretation

of the gesture.² We believe this is the reading when the gesture illustrates the underlying NP concept, as in the following example.

- (3) [Windows] are usually made of glass.
+ *iconic rectangular gesture*

Example (3) makes a statement about windows in general and the gesture seems to illustrate the window concept. There is no specific object to which the speaker makes reference.

The distinction between *object-* and *interpretant-related* gestures of Fricke (2012) is mainly based on an observation she made in a study where she asked people to walk a certain route through Berlin and describe her route to some other person afterwards and this person had to retell the route to somebody else again. She observed that in these descriptions, people used different kinds of gestures when talking about the same object. For example, on the route, participants encountered a rectangular opening in a building, which they often referred to as *gate* (German: *Tor*). Some participants had used a bow-formed gesture when making reference to the rectangular opening, while others iconically mapped the shape of the actual opening onto a rectangular-shaped gesture. Instead of assuming that participants who used the bow-formed gesture and not a rectangular one produced a gesture-speech mismatch, Fricke concluded that a speaker may use gestures in different manners or with different aims: 1., to depict a specific property of a related reference object in the real world (i.e., the real opening that people encountered), and 2., to illustrate the meaning of a corresponding concept (i.e., the concept of an opening in a building or a gate). To explain these two uses of co-speech gestures, Fricke (2012) builds on the gesture-speech model of Kita & Özyürek (2003), who showed that people's gestures are not only influenced by (non-linguistic) visual images, but also by the grammatical and lexical structures of the underlying native language of the person performing the gestures. Interpretant-related gestures refer to an abstract mental prototype that is associated with the speech concept in a specific language and not to an object a person has actually encountered. According to Fricke (2012), these prototypes arise from the shared expectations within a linguistic community which this entity needs to fulfill in order to be labeled in such a way. As opposed to this, object-related gestures refer to a specific object, which is denoted by the associated referentially interpreted DP and which actually exists. Thus, those participants in Fricke's study who referred to the rectangular gate with a bow-formed gesture used an interpretant-related gesture; they referred to a prototypical form of a gate (cf. Fricke, 2012: 244ff for a study determining the prototypical shape properties of a *gate*). The participants who used a rectangular gesture for the gate, however, used an object-related gesture; they illustrated the actual shape property of the gate seen on the route they had taken.

Note that this model predicts that a gesture that gives rise to an interpretant-related reading must visualize a prototypical aspect of the concept under discussion. This means that in (3), where the gesture illustrates the shape of a window, it has to be a gesture depicting a rectangular shape and not a round one because windows are usually rectangular in our cultural communities. In (2), on the other hand, the accompanying gesture could also depict a circular shape, illustrating a round object, because the gesture makes reference to a property of a certain referential object and this object, a window, might happen to be a round window.

Ebert & Ebert's (2014) distinction between comparison and exemplification meaning contributions of gestures is broadly related to Fricke's (2012) distinction of object- vs. interpretant-related interpretations but differs in certain important details. Ebert & Ebert (2014) also argue that it is not a genuine ambiguity of gesture use that we are dealing with in the cases at hand, but that the two readings actually come about via different time-alignments of gesture and speech. Alignment narrowed to the NP triggers the exemplification reading and alignment

² The terminology goes back to Peirce's distinction between *sign*, *object*, and *interpretant* (which is roughly what can be called the sign's meaning), cf. Peirce (1931-58).

with the whole DP results in a comparison reading.³ Ebert & Ebert's approach is discussed in the next subsection.

2.2 Semantic Effects of the Temporal Alignment of Gesture and Speech

Empirical studies have shown that a gesture and its corresponding speech segment are temporally aligned in systematic ways. Usually, the core part of the gesture, its *stroke*, coincides with an intonational peak, the main accent, of a phrase (see e.g., Pittenger et al., 1960; Kendon, 1980; McNeill, 1992; Loehr, 2004; a.o.). It has also been suggested that gesture phrases (usually consisting of preparation phase, stroke, and retraction phase⁴) align with *tone groups* (i.e., “the smallest grouping of syllables over which a completed intonation tune occurs”, Kendon, 1972: 184), *intermediate phrases* (Loehr, 2004), or *focus phrases* (Ebert et al., 2011). Despite the different suggestions for the actual alignment categories, there is a general agreement among gesture researchers that the content of a gesture is intertwined with the content of the accompanying speech signal. And it is sometimes claimed that gestures are actually only interpretable in the context of the speech signal (e.g., Kopp et al., 2004). Gesture information can be co-expressive and carry the same information as the speech signal, but it can also be complementary, i.e., convey information that is additional to what is said in the speech signal (cf. Kendon, 2004).

In this paper, we are concerned with gestural contributions that add information to the multimodal utterance, i.e., gestures that bring additional information not present in speech. Ebert & Ebert (2014) aim at a formalization of systematic gestural semantic contributions based on the temporal alignment of gesture and speech. Before we present these alignment rules, we will briefly introduce some background on Ebert & Ebert's (2014) formal system.

In Ebert & Ebert (2014), iconic and deictic co-speech gestures receive a uniform semantics. The authors argue that these co-speech gestures function like supplements and contribute *non-at-issue* meaning. To appreciate the difference between *at-issue* and *non-at-issue* information, consider the following example from Potts (2005).

- (4) a. Ed's claim, which is based on extensive research, is highly controversial.
b. I have to mow the damn lawn.

(Potts, 2005: 7)

Sentences (4a) and (4b) each make two types of contributions, a main assertion (which is at-issue) and a more peripheral contribution that is not the main point of the conversation (which is not at-issue). In (4a), the contribution made by the main clause, i.e., that Ed's claim is highly controversial, is at-issue and the contribution of the appositive clause, that the claim is based on extensive research, is supplemental or non-at-issue, but makes an additional point, for example that the speaker wants to convince the audience that the controversy should not be taken as a criterion to eventually dismiss Ed's claim. What is at-issue in (4b) is the speaker's contribution that she has to mow the lawn. The expressive *damn* adds the non-at-issue contribution that the speaker has a negative attitude toward this obligation. Potts (2005) argues that non-at-issue content is independent of the at-issue content of an utterance and that there are hence two different dimensions of meaning. Furthermore, non-at-issue content is non-negotiable. At-issue content is directly deniable, while non-at-issue content can only be denied

³ Fricke (2012: 250) surmises that it is impossible for the hearer to decide whether a gesture is object- or interpretant-related, as there seems to be no formal criterion that helps to differentiate the two uses. But interestingly, she adds in footnote 29 on page 250 that an examination of exactly this question – whether there are formal means to differentiate the two gesture interpretations – would be worthwhile. Our enterprise in this paper and in the experimental study we present is to do exactly this.

⁴ Gestures can be divided up into different phases: 1., the preparation phase, where arms and hands are brought into position, 2., the actual stroke, which constitutes the core part of the gesture and is the part with the most kinetic energy, and 3., the retraction phase, where arms are brought back into a resting position. Pre- or post-stroke holds, with the hands just holding a certain position for a while, can also be found (cf. e.g., McNeill, 1992).

via other, more discourse-interrupting, means such as with the help of a *hey, wait a minute* construction (cf. von Stechow, 2004, based on Shanon, 1976, or Potts, 2015; see Syrett & Koev, 2015 for critical discussion).

- (5) a. A: Ed's claim, which is based on extensive research, is highly controversial.
 B: # That's not true! The claim is not based on extensive research.
 B': Hey, wait a minute. Actually, I don't think the claim is based on extensive research.
- b. A: I have to mow the damn lawn.
 B: # That's not true! You like mowing the lawn!
 B': Hey, wait a minute. Actually, I thought you liked mowing the lawn.

The responses that B gives in (5) attempt to directly deny the non-at-issue content of A's utterances. However, this is not possible since A does not offer B the option of discussing or questioning the non-at-issue content in (5). A direct denial of the non-at-issue content is only possible with the discourse-interrupting *hey, wait a minute* protest by B'.

Ebert & Ebert (2014) propose a model which makes use of the uni-dimensional dynamic system of AnderBois et al. (2015), originally proposed for the treatment of appositives, and adapt it to speech-accompanying gestures. For illustration, consider the following example:

- (6) Ludger Beerbaum, an outstanding show jumper, was accused of doping.
 (adapted from Ebert & Ebert, 2014)

AnderBois et al. (2015) point out that updating the context with an utterance as in (6) results in integrating two different propositions into the context set. By uttering (6), the speaker makes the at-issue *proposal* to update the common ground with the propositional content that Ludger Beerbaum was accused of doping. Additionally, there is also the non-at-issue *imposition* that Ludger Beerbaum is an outstanding show jumper. This imposition enters the common ground without the speaker's acknowledgment. Ebert & Ebert (2014) adopt the distinction between at-issue and non-at-issue contributions of sentences with appositives, as proposed by AnderBois et al. (2015), and transfer it to the treatment of speech-accompanying gestures. Two types of propositional variables, p (at-issue) and p^* (non-at-issue), keep track of the distinction between at-issue and non-at-issue contributions. A formalization of (6) in AnderBois et al.'s (2015) framework is given in (7)⁵:

- (7) $[x] \wedge x = \text{LUDGER_BEERBAUM} \wedge \text{OUTSTANDING_SHOWJUMPER}_{p^*}(x) \wedge \text{ACCUSED_OF_DOPING}_p(x)$

AnderBois et al. (2015) introduce variables over individual concepts, which are of type $\langle s, e \rangle$ (for example the variable x in (7)). Ebert & Ebert (2014) extend this system and propose that pointing gestures as well as iconic gestures refer to an intended referent g . We write \mathbf{g} for the corresponding formal language expression for that referent g . It is interpreted as the individual concept with the same value g for all possible worlds, in other words: a rigid designator. The performance of a gesture introduces a gesture discourse referent (cf. Umbach & Gust, 2014). Crucially, the intended referent g relates to the co-occurring speech signal in different ways. Temporal alignment of gesture and speech is meaningful and the gestural contribution to the overall semantic representation is dependent on the aligned speech expression. Alignment with a DP (or more precisely, with the determiner within this DP) results in an interpretation where the gesture concept is related in certain predefined ways to the speech concept. In other words: a comparison interpretation. Co-occurrence with an indefinite determiner expresses similarity between the gesture referent and the discourse referent introduced by the determiner with respect to certain contextually salient features (cf. Umbach & Gust, 2014, for such an analysis of the German similarity demonstrative *so*). If the aligned speech expression is a definite, it is strict identity between the gesture and the speech concept. As mentioned above, co-occurrence

⁵ $[x]$ stands for a reset of assignments at position x , which comes down to the introduction of a dynamic existential quantifier with bound variable x . Details of the formal system that are not essential to the cases at hand are omitted. See Ebert & Ebert (2014) and Ebert et al. (2020) for further details.

with an NP results in an exemplification relation. The gesture has to exemplify the NP concept (cf. Fricke, 2012; Lücking, 2013). Note that all these gestural meaning contributions are non-at-issue.

For illustration purposes, consider the example in (8). As pointed out above, if a gesture is aligned to an NP, the intended gesture referent has to exemplify the NP concept. If it is aligned to an indefinite determiner, this results in expressing a similarity relation between the verbal and the gesture concept. (8a) and (8c) are simplified variants of example (2), with an iconic gesture depicting the rectangular shape of a window aligned to only the NP in (8a) and the whole DP in (8c). The formal representation of (8a) is shown in (8b), (8c)'s formalization is given in (8d).

- (8) a. My living room has a [window].
 + *iconic rectangular gesture*
- b. $[x] \wedge x = \text{MY_LIVING_ROOM} \wedge [y] \wedge \text{WINDOW}_p(y) \wedge [z] \wedge z = \mathbf{g} \wedge$
 $\text{WINDOW}_{p^*}(z) \wedge \text{HAS}_p(x, y)$
- c. My living room has [a window].
 + *iconic rectangular gesture*
- d. $[x] \wedge x = \text{MY_LIVING_ROOM} \wedge [y] \wedge \text{WINDOW}_p(y) \wedge [z] \wedge z = \mathbf{g} \wedge \text{SIM}_{p^*}(y, z) \wedge$
 $\text{WINDOW}_{p^*}(z) \wedge \text{HAS}_p(x, y)$

In the formalization, z is the gestural discourse referent that stands for the individual concept \mathbf{g} of the rectangular gesture, which refers to the intended referent \mathbf{g} (i.e., a window). The variable y stands for the verbal discourse referent introduced by the indefinite DP *a window*. Due to the temporal alignment of the gesture with the NP, which gives rise to the exemplification requirement that the gesturally depicted referent has to exemplify the NP predicate (in other words: has to illustrate (the shape of) a window), the non-at-issue requirement $\text{WINDOW}_{p^*}(z)$ is added in both formulas, (8b) and (8d). Crucially, in (8d), there is an additional requirement, $\text{SIM}_{p^*}(y, z)$, which is added due to the temporal alignment of the gesture and the indefinite determiner. $\text{SIM}(y, z)$ is true iff y is similar to z in certain contextually relevant aspects (see Umbach & Gust, 2014, for a three-valued formally spelled out implementation of the SIM predicate), here: with respect to shape.

As can easily be verified, in (8d), the gesture concept z , which is rigid, is compared to the verbal concept y and it is required that the two are similar in certain contextually relevant respects. It is required that in all p^* -worlds, the window that the speaker's living room has is similar to a rectangular window. This corresponds to the comparison interpretation of (8c), which is triggered in case of a DP-aligned gesture. In (8a), only the exemplification requirement is triggered and for the formula in (8b) to be true, it is sufficient that the gesture the speaker performs exemplifies the window concept, which it does. This corresponds to the exemplification interpretation. Crucially, (8b) is still true even if the window in the speaker's living room is round and does not resemble a rectangular shape.

2.3 Semantic Interpretation Effects of Gesture Alignment with Different DP Types

Both discussed approaches, Fricke (2012) and Ebert & Ebert (2014), focus on the interpretation of gestures in the context of accompanying (potentially) referring expressions such as definites and indefinites. While Fricke (2012) argues for an object-related reading on the one hand and an interpretant-related reading on the other hand, Ebert & Ebert (2014) suggest differentiating between comparison readings and exemplification readings. We believe that Ebert & Ebert's (2014) exemplification readings, which come about when the gesture is aligned to the NP only, can be understood as constituting essentially the same kind of readings as those that Fricke (2012) describes as interpretant-related interpretations of co-speech gestures. The gesture exemplifies or illustrates a prominent feature of the underlying concept that is associated with the NP. In contrast, what Fricke (2012) claims to be the other possible interpretation, the object-

related interpretation, is a referentially based reading. This reading is hence predicted to arise only for referentially interpreted DPs. Here, the gesture restricts the possible extension of a referential DP. Ebert & Ebert's (2014) comparison interpretation, however, leads to a reading that constitutes a certain (non-at-issue) relation (of similarity or identity) between two individual concepts, the gesture concept and the speech concept. This does not necessarily require the DP under discussion to be referential.

Empirical support for this second view comes from examples such as (1), presented at the beginning of this article (*I want a pullover*). As we have seen, the indefinite can be interpreted non-specifically, i.e., non-referentially, even in cases when it is accompanied by an iconic gesture. This reading is clearly not the exemplification or concept-related reading since the V-gesture does not characterize a prototypical concept of a pullover. It restricts the type of pullover that the speaker wants to have to only V-necked ones. This non-specific gesture-enriched reading is as follows: the speaker wants a V-necked pullover but has no specific exemplar in mind. This is clearly a comparison-based interpretation of the gesture, where the gestural concept (a V-necked pullover) has to be similar to the verbal concept corresponding to the DP *a pullover*.

We find related empirical arguments when we consider the readings of the following example.

- (9) When she moves to London, Maria wants to have a living room with [a huge window].
+ *iconic circular gesture*

The sentence has a prominent *de dicto* reading, where Maria wishes to have a living room with a huge round window at some point. Clearly, this is a non-referential reading – neither the living room nor the window have to exist – and yet this cannot be an exemplification reading either since the circular gesture does not exemplify the typical window concept. This reading is, however, derivable via the similarity requirement that Ebert & Ebert (2014) suggest for the interpretation of indefinites plus accompanying gestures.

While we do find non-specific readings of indefinites with co-speech gestures that seem to act restrictively, as exemplified in examples (1) or (9), gestures accompanying other quantifiers such as, for example, negative ones, do not have the same restrictive meaning that we observe for the examples above.

- (10) ?Maria's living room has [no window].
+ *iconic circular gesture*

The utterance in (10) with the iconic co-speech gesture cannot be interpreted to mean that Maria's living room has no round window (but possibly some rectangular ones). That there is no such reading can be explained in Ebert & Ebert's (2014) approach under the assumption that the two concepts that are compared have to be externally dynamic since the corresponding variables (z and y in the SIM relation in (8d), for example) have to be available across dimensions in order to be able to set them into relation to each other.⁶ This then makes the prediction that all quantifiers which are externally dynamic can receive comparison readings and those which are not, cannot. We predict that non-referential quantificational DPs such as *no N* or *every N*, and all other externally static quantifiers such as *at least one N* or *at most two N*, when accompanied by a gesture, do not exhibit a comparison-based reading (see e.g., Groenendijk & Stokhof, 1991; Kamp & Reyle, 1993; Szabolcsi, 1997 for discussion on the dynamic behavior of quantificational DPs). Externally dynamic quantifiers are quantifiers that can bind their arguments not only locally, but across boundaries, e.g., across sentences. For

⁶ Note that Ebert & Ebert (2014) show that indefinites plus accompanying gesture which are embedded under negations cannot receive a reading where the gesture acts restrictively. If we understand *no* in (10) as equivalent to negated *a*, i.e., the sentence in (10) as equivalent to *Maria's living room does not have a window* (plus iconic gesture), (10) makes the same point.

illustration purposes, consider the externally dynamic indefinite quantifier in (11a) and the static negative quantifier in (11b).

- (11) a. Peter owns a horse_i. It_i is very pretty.
 b. Peter owns no horse_i. *It_i is very pretty.

The indefinite *a horse* in (11a) is introduced in the first sentence and can bind the pronoun, i.e., a variable, in the follow-up sentence. The negative quantifier in (11b), in contrast, cannot bind a variable in a follow-up sentence although it can bind pronouns in its domain (e.g., *No horse_i likes to have its_i tail pulled.*).

To derive comparison readings, dynamic binding across dimensions is necessary because gesture concepts are set into relation with speech concepts. To be able to achieve this, variables have to be bound across dimensions, which means that their binding quantifiers have to be externally dynamic. For further illustration, consider the following example in (12a) and its semantic representation in (12b)⁷.

- (12) a. ?In Maria's house, [no window] is sealed.
 + *iconic circular gesture*
 b. $\text{NOT}_{p'}([x] \wedge \text{WINDOW_in-Maria's-house}_p(x) \wedge \text{SEALED}_p(x)) \wedge [z] \wedge z = \mathbf{g} \wedge$
 $\text{WINDOW}_{p^*}(z) \wedge \text{SIM}_{p^*}(x, z)$

As in (10), we predict that accompanying the full DP with the iconic gesture is illicit since the *no*-quantifier does not allow for a comparison reading. That there is no comparison reading for (12a) is predicted by our formalism, which can be verified by consulting the representation in (12b). Glossing over many details, it can be seen that the variable *x* in its last occurrence, where it appears as an argument of the SIM predicate, $\text{SIM}_{p^*}(x, z)$, is free and not bound by the existential quantifier since the existential quantifier is embedded in the NOT-predicate, which does not allow for external binding (in general and also in the system of AnderBois et al., 2015). Hence, the formula does not express a similarity relation between gesture and speech concept and contains a free variable *x*, which we consider responsible for the fact that a DP alignment as assumed in (12a) is not possible.

In the following, we use this insight that externally static quantifiers and negative ones in particular do not allow for comparison readings to experimentally test for Ebert & Ebert's (2014) treatment of co-speech gestures, or more precisely: their proposal concerning the semantic effects of the temporal alignment of gestures with NPs and DPs.

3 Experimental Study

The study used a rating task, which was set out to show that DP-aligned gestures result in what we have defined as comparison-based readings and NP-aligned gestures give rise to exemplification readings. Our two-factorial design crossed the two within-factors QUANTIFIER (indefinite *ein* (English: *a*) vs. negative quantifier *kein* (English: *no*)) and ALIGNMENT (DP-aligned gestures vs. NP-aligned gestures). The core hypothesis predicts an interaction of the two factors such that a *kein*-quantifier is rated worse than an *ein*-indefinite if aligned with the DP, but not if aligned with the NP. We expected different ratings for DP-aligned quantifiers because *kein*-DPs are thought to disallow comparison readings, which are predicted to arise with DP-aligned gestures, but which are unavailable for negative quantifiers. We did not expect such rating differences with NP-aligned gestures because they should give rise to exemplification readings, which are possible with all kinds of quantifiers.

⁷ For the details of the semantic framework, we refer the reader to AnderBois et al. (2015).

3.1 Method

Participants. We tested 27 participants (15 female, 12 male), of which 7 participants were excluded from the analysis due to not meeting the predefined criterium of rating the filler items with mismatching gestures significantly worse than the matching ones ($n = 5$), for not being a native German speaker ($n = 1$), or due to technical problems during the study ($n = 1$). The participants' ages ranged from 20-65 ($mean = 27.9$ years, $sd = 9.33$). All participants were naïve with respect to the research question.

Materials. The study featured 24 experimental items, each one a video with two sentences; a context sentence, which remained constant across conditions, and a target sentence with a co-articulated iconic gesture. A male professional speaker verbalized all sentence stimuli while being videotaped. The videos were recorded in a professional video studio of the *studiumdigitale* at the University of Frankfurt in one session. The speaker used a teleprompter to read the sentences while being recorded.

Items were recorded in each of the four conditions: *ein*-indefinite plus NP-alignment, *kein*-indefinite plus NP-alignment, *ein*-indefinite plus DP-alignment, *kein*-indefinite plus DP-alignment. Consider the following sample item, shown in each of the four conditions.

- (13) a. *Jonas hat gestern ein seit Jahren unbewohntes Haus besichtigt.*
 Jonas has yesterday a since years unoccupied house viewed
Überraschenderweise sah er dort [ein großes, sauberes Fenster.]
 Surprisingly saw he there a big clean window
 ‘Yesterday, Jonas visited a house that has been unoccupied for years. Surprisingly, he saw a big, clean window there.’
- b. *Jonas hat gestern ein seit Jahren unbewohntes Haus besichtigt.*
Überraschenderweise sah er dort ein großes, sauberes [Fenster.]
- c. *Jonas hat gestern ein seit Jahren unbewohntes Haus besichtigt.*
 Jonas has yesterday a since years unoccupied house viewed
Es war also nicht überraschend, dass er dort [kein großes, sauberes Fenster] sah.
 It was hence not surprising that he there no big clean window saw
 ‘Yesterday, Jonas visited a house that has been unoccupied for years. It was hence no surprise that he saw no big, clean window there.’
- d. *Jonas hat gestern ein seit Jahren unbewohntes Haus besichtigt. Es war also nicht überraschend, dass er dort kein großes, sauberes [Fenster] sah.*

Squared brackets indicate the gesture alignment, which was either on the whole DP ((13a) and (13c)) or solely on the NP ((13b) and (13d)). For the gesture, the speaker traced a rectangle with his index fingers in front of him while uttering the DP or NP. The gesture that we chose was meant to be a prototypical gesture that was suited to represent the respective NP concept, here: that of a window.

We initially intended to set apart the different alignment conditions (NP vs. DP) only by instructing our speaker to start the accompanying gesture late (with the NP-complement) in the NP-condition and early (with the determiner) in the DP-condition. In order to ensure that the different gesture alignment positions could be clearly told apart, we designed the items in such a way that each target noun in a target sentence was separated from its determiner by two adjectives. However, when taking a closer look at the experimental items after the recording session, we noticed that the alignment of the gesture with the NP in the one condition and the DP in the other condition was not always as accurate as we wanted it to be.

We hence decided to adapt a technical procedure to control for the alignment positions. For each item, we selected only one of the video recordings of the four conditions, namely the one

where the gesture, according to our intuitions, was optimally pronounced by the speaker. The audio track was silenced. We then used the audio tracks from the four conditions and fitted the chosen gesture to these audio tracks according to the respective alignment condition (i.e., to the beginning of the DP or the beginning of the NP). The stroke of the gesture then started either with the DP or with the NP. If necessary, the beginning of the video track (where no movement was performed) was repeated to match the length of the audio track in order to ensure proper alignment of the gesture to the target phrase. To avoid disturbances resulting from unaligned lip movements or face gestures, we pixelated the speaker's face in the recordings of our items. For the cutting of the videos, we used *Windows Movie Maker*, for the pixelation of the speaker's face, we used *Sensarea*.

The four variants of the 24 experimental items were assigned to four lists according to a Latin square design. Moreover, 26 filler items were added to all four lists. Thirteen of the filler items contained gestures that did not match the context of the verbal utterance and thirteen contained gestures that were perfectly suitable in the given context. These fillers served to control whether participants actually paid attention to the video track. Participants were expected to rate matching fillers significantly better than non-matching ones and would be excluded from the final analysis if this was not the case. In the following, we list an example item of a matching (14a) and non-matching (14b) filler item.

- (14) a. *Anna hat versehentlich ein Glas Wasser über ihr Laptop geschüttet. [Der Maus] geht's aber noch gut.* [+ gesture imitating handling a computer mouse on a mouse pad]
 'Anna accidentally spilled water over her laptop. Luckily, the mouse is still ok.'
- b. *Am Abend gibt es einen großen Boxkampf. Um 20h soll [der Ring] freigegeben werden.* [+ gesture imitating slipping over a finger ring]
 'In the evening, there will be a big boxing event. At 8 pm the ring will be opened.'

In the item recording session, we recorded 36 potential experimental items in the four conditions and chose the best 24 among them for the final study. The main criterion for inclusion of an item in the final study was how well the gesture illustrated the underlying NP concept. To decide on this, we conducted a pre-study on *Prolific*⁸ with ten participants. All participants were native speakers of German and naïve with respect to the research question. In the pre-study, participants saw (muted) videos containing the gesture from our potential experimental items. Below the video, participants saw a question of the type *How typical do you consider the gesture for an NN?* NN was replaced with the corresponding target noun concept of each item. The participants' task was to rate on a 7-point Likert scale how typical they considered the gesture seen in the video for the given NP concept (1 = 'totally untypical'; 7 = 'totally typical'). For the experimental study, we chose those 24 items that were rated most typical as illustrating the NP concept.

Procedure. Participants were recruited via the online platform *Prolific* and used their own computers for the experiment. On average, it took the participants about 24 minutes to accomplish the task. The session started with two practice trials followed by the experimental items intermixed with the fillers in a randomized order for each participant. For each item, a video appeared and could be started by pressing a play button on the screen. Participants were free to watch the video several times, if they wanted to. Below the video screen, there was a scale extending from 1 = 'entirely unacceptable' to 7 = 'entirely acceptable'. Participants had to rate how acceptable they thought the target sentence (i.e., the second sentence in the video they saw) was in the given context (i.e., the first sentence in the video). They gave their judgment by means of a mouse click on the number of the scale. In an introductory text, they were asked to pay attention not only to the audio track but also to the video track. With the help of the trial items, they were made familiar with what should be understood as the target sentence

⁸ Prolific is an online platform on which one can recruit participants for experimental studies of any kind. The participants are paid for their participation.

and what counted as context. The questionnaire was built with *SoSci Survey* (Leiner, 2019), an online platform where one can create web-based questionnaires free of charge for academic purposes.

3.2 Results and Discussion

Data were aggregated per condition per participant (*F1* analysis) or item (*F2* analysis) and subjected to a repeated measures ANOVA.

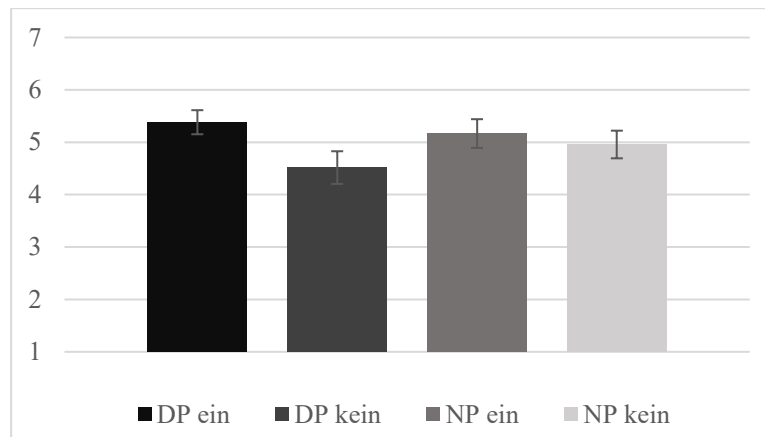


Figure 1. Mean ratings of the items in each condition. The error bars represent ± 1 *se* as obtained in the *F1* analysis

Figure 1 shows the mean ratings of the experimental items in each condition: for items with *ein* and DP-alignment: 5.38 (*se*: .23), for items with *ein* and NP-alignment: 5.16 (*se*: .27), for items with *kein* and DP-alignment: 4.51 (*se*: .31), and for items with *kein* and NP-alignment: 4.95 (*se*: .26). The results show a main effect for QUANTIFIER [$F1(1,19) = 28.385^{***}$; $F2(1,23) = 20.614^{***}$] as well as a significant interaction QUANTIFIER \times ALIGNMENT [$F1(1,19) = 4.987^*$; $F2(1,23) = 9.132^{**}$]. Paired comparisons revealed a significant difference between *ein* and *kein* with DP-alignment [$t1(19) = 4.466^{***}$; $t2(23) = 4.750^{***}$], and no difference between *ein* and *kein* with NP-alignment, as predicted. The mean rating for matching filler items was 5.47 and for the non-matching ones 3.21.

The study thus confirms the predicted better ratings for *ein*-DPs than for *kein*-DPs with DP-aligned co-speech gestures. The finding corroborates the hypothesis that temporal alignment of gesture and speech plays a major role in gesturally-driven semantic effects and confirms Ebert & Ebert's (2014) claim that gesture alignment with the whole DP triggers a comparison-based reading and alignment with the NP results in exemplification-based readings.

4 General Discussion

We adopted the general framework of Ebert & Ebert (2014), which seeks to account for the semantic effects of gesture-speech interaction. In this approach, the temporal alignment of gesture and speech is meaningful and the same gesture may contribute differently to the overall semantic representation, depending on whether it accompanies a DP or an NP. Based on work by Fricke (2012), Ebert & Ebert (2014) claim that a DP-accompanying gesture gives rise to a comparison-based reading, while NP-accompanying gestures yield exemplification-based readings. As a consequence of Ebert & Ebert's dynamic system, it is predicted that only externally dynamic quantifiers accompanied by co-speech gestures can give rise to comparison interpretations. This then makes the prediction that externally static quantifiers such as the German negative quantifier *kein* (English: *no*) cannot exhibit such a reading, which in effect means that they should be illicit with DP-aligned gestures. This paper presents a study that aimed at experimentally corroborating this prediction. The study reveals the expected interaction effect of quantifier type and alignment position and thus confirms Ebert & Ebert's (2014) analysis. DP-aligned gestures with the negative quantifier *kein* are perceived as

significantly less acceptable than DP-aligned gestures with *ein*-indefinites. No such difference is found with NP-aligned gestures for the same quantificational DPs.

One remaining question is what is responsible for the alignment effect proposed by Ebert & Ebert (2014) and confirmed in our experiment. A possible answer might be found in the organization of language processing. If a speaker starts the production of a gesture late and hence aligns it to the NP, it might be impossible for a hearer to still integrate the gesture material into the DP, thus excluding a comparison-based interpretation. However, the gesture material can still be combined with the NP semantics, which necessarily results in an exemplification reading. We leave an experimental validation of this speculation for future research.

Acknowledgments

This work has been conducted within the DFG project PSIMS, which is part of the Priority Program XPRAG.de. We thankfully acknowledge the support of the German Research Foundation (DFG).

Furthermore, we would like to thank the DIPF, Frankfurt (Leibniz-Institut für Bildungsforschung und Bildungsinformation) for giving us the opportunity to use their laboratories for the execution of a pilot study to our final experimental study. We also thank Gregor Brinkmeier and Stephan Zerlik from *studiumdigitale* of the University of Frankfurt for supplying us with a studio and all technical equipment for the recordings as well as an enormous amount of help with the postprocessing of the recordings.

We thank Yvonne Portele for volunteering to enact the experimental items in the videos for a pilot study and Lars Eichstaedt for the enactment of the experimental items of the final study. Moreover, we would like to thank Theresa Stender for help with the execution of the experiment. We furthermore thank Kathryn Barnes, Aleksandra Ówiek, Christian Ebert, Susanne Fuchs, Stefan Hinterwimmer, and Manfred Krifka for valuable discussions on the topics of this paper.

Finally, we would like to express our special gratefulness to Robin Hörnig for help with the statistical evaluation and for discussion on the experimental design and the results.

References

- AnderBois, S., Brasoveanu, A., & Henderson, R. (2015). At-issue proposals and appositive impositions in discourse. *Journal of Semantics*, 32(1), 93-138.
- Ebert, Ch., & Ebert C. (2014). Gestures, demonstratives, and the attributive/referential distinction. Talk presented at the conference 'Semantics and Philosophy in Europe 7', 26-28 June 2014, Leibniz-ZAS Berlin.
- Ebert, Ch., Ebert, C., & Hörnig, R. (2020). Demonstratives as dimension shifters. In M. Franke, N. Kompa, M. Liu, J. L. Mueller, J. Schwab (Eds.), *Proceedings of Sinn & Bedeutung 24* (pp. 161-178). Osnabrück: University of Osnabrück.
- Ebert, C., Evert, S., & Wilmes, K. (2011). Focus marking via gestures. In I. Reich, E. Horch & D. Pauly (Eds.), *Proceedings of Sinn & Bedeutung 15* (pp. 193-208). Saarbrücken: universaar.
- Esipova, M. (2019). *Composition and Projection in Speech and Gesture*. Dissertation, New York University.
- Fricke, E. (2012). *Grammatik multimodal: Wie Wörter und Gesten zusammenwirken*. Berlin & Boston, MA: Walter de Gruyter.
- Groenendijk, J., & Stokhof, M. (1991). Dynamic predicate logic. *Linguistics & Philosophy*, 14(1), 39-100.
- Kamp, H., & Reyle, U. (1993). *From Discourse to Logic*. Dordrecht: Springer.

- Kendon, A. (1972). Some relationships between body motion and speech: An analysis of an example. In A. Siegman & B. Pope (Eds.), *Studies in Dyadic Communication* (pp. 177-210). New York, NY: Pergamon Press.
- Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In M. R. Key (Ed.), *The Relationship of Verbal and Nonverbal Communication* (pp. 207-227). The Hague: Mouton.
- Kendon, A. (2004). *Gesture. Visible Action as Utterance*. Cambridge, UK: Cambridge University Press.
- Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, 48(1), 16-32.
- Kopp, S., Tepper, P., & Cassell, J. (2004). Towards integrated microplanning of language and iconic gesture for multimodal output. In R. Sharma, T. Darrell, M. P. Harper, G. Lazzari & M. Turk (Eds.), *Proceedings of the International Conference on Multimodal Interfaces 2004* (pp. 97-104). State College, PA: ACM Press.
- Lascarides, A., & Stone, M. (2009). A formal semantic analysis of gesture. *Journal of Semantics*, 26(4), 393-449.
- Leiner, D. J. (2019). SoSci Survey (Version 3.1.06) [Computer Software]. Available at: <https://www.soscisurvey.de>.
- Loehr, D. (2004). *Gesture and Intonation*. Dissertation, Georgetown University, WA.
- Lücking, A. (2013). *Ikonische Gesten. Grundzüge einer linguistischen Theorie*. Berlin & Boston, MA: Walter de Gruyter.
- McNeill, D., & Levy, E. T. (1982). Conceptual representations in language activity and gesture. In R. J. Jarvella & W. Klein (Eds.), *Speech, Place, and Action: Studies in Deixis and Related Topics* (pp. 271-295). Chichester: Wiley.
- McNeill, D. (1992). *Hand and Mind. What Gestures Reveal About Thought*. Chicago, IL: University of Chicago Press.
- Peirce, C. S. (1931-58). *Collected Papers*. C. Hawthorne & P. Weiss (Eds.), Vol. 1-6. A. W. Burks (Ed.), Vol. 7-8. Cambridge, MA: Harvard University Press.
- Pittenger, R. E., Hockett, C. F., & Danehy, J. J. (1960). *The First Five Minutes: A Sample of Microscopic Interview Analysis*. Ithaca, NY: Paul Martineau.
- Potts, C. (2005). *The Logic of Conventional Implicatures*. Oxford: Oxford University Press.
- Potts, C. (2015). Presupposition and implicature. In S. Lappin & C. Fox (Eds.), *The Handbook of Contemporary Semantic Theory* (pp. 168-202). Chichester: Wiley Blackwell Publishing.
- Schlenker, P. (2018). Gesture projection and cosuppositions. *Linguistics & Philosophy*, 41(3), 295-365.
- Shanon, B. (1976). On the two kinds of presupposition in natural Language. *Foundations of Language*, 14(2), 247-249.
- Syrett, K., & Koev, T. (2015). Experimental evidence for the truth conditional contribution and shifting information status of appositives. *Journal of Semantics*, 32(3), 525-577.
- Szabolcsi, A. (1997). *Ways of Scope Taking*. Dordrecht: Kluwer.
- Umbach, C., & Gust, H. (2014). Similarity demonstratives. *Lingua*, 149, 74-93.
- von Fintel, K. (2004). Would you believe it? The king of France is back! Presuppositions and truth-value intuitions. In A. Bezuidenhout & M. Reimer (Eds.), *Descriptions and beyond* (pp. 315-341). Oxford: Oxford University Press.