

ARTICLE

Crossmodal collostructional analysis of English [ADV and ADV] constructions: multimodal constructions or crossmodal collostructions?

Daiya Kuryu 

Center for Languages and Literature, Lund University, Lund, Sweden
Email: daiya.kuryu@gmail.com

(Received 23 July 2024; Revised 05 February 2025; Accepted 07 February 2025)

Abstract

This study investigates the relationships between a group of English constructions that exhibit similar formal and semantic properties and their accompanying gestures. The target constructions, collectively referred to as the [ADV and ADV] constructions, are divided into two types: the reduplicative adverbial constructions (RACs; *over and over, again and again, on and on*) and the oppositive adverbial constructions (OACs; *back and forth, up and down, in and out*). Using a statistical method called crossmodal collostructional analysis, the degree of associations between these constructions and their frequent gestural correlates is quantified to explore whether they are likely to form multimodal constructions. Consequently, it is concluded that some combinations of a target construction and its corresponding gesture are associated strongly enough to be regarded as potential multimodal constructions, whereas others are more like crossmodal collostructions, which are associated combinations of linguistic and gestural constructions but are nonetheless compositional. Finally, the embodied motivations for the potential multimodal constructions are also discussed by drawing on the notion of exbodiment.

Keywords: Construction Grammar (CxG); multimodal constructions; crossmodal collostructions; crossmodal collostructional analysis; gesture lexicon; exbodiment

1. Introduction

Less commonly recognized as part of language is the gestural channel, including manual gestures, facial expression, and bodily posture. These are, however, subject to conventionalization and coordination with other linguistic processes. (Langacker, 2008, p. 462).

As Langacker noted as early as 2008, recent developments in Construction Grammar (CxG) as well as the creation of multimodal corpora have led researchers

to consider the multimodal nature of language (Zima & Bergs, 2017). Consequently, a number of case studies (e.g., Andrén, 2014; Mittelberg, 2017; Zima, 2017) have explored the possibility that language users have entrenched *multimodal constructions*, that is, ‘constructions with verbal and gesture FORM elements that express a joint MEANING’ (Hoffmann, 2021, p. 81, his emphasis). The nature of such multimodal constructions, however, has been subject to debate (Hoffmann, 2021; Zima & Bergs, 2017). On the one hand, some scholars hold a critical position in favor of the strict definition of multimodal constructions, according to which kinesic features must be obligatory for a combination of linguistic and gestural patterns to be qualified as a multimodal construction (Ningelgen & Auer, 2017; Ziem, 2017). On the other hand, in line with the usage-based view of language that sees entrenchment as a gradual phenomenon, more researchers expect a continuum of constructions between those that are ‘infrequently and inconsistently associated with gestures (and other multimodal cues) and [those] that are frequently and systematically used with given gestures’ (Zima, 2017, p. 2). It is the latter position that this study takes, and thus aligns with the gradient notion of constructionhood (Ungerer, 2023) in the multimodal context.

As argued by Hoffmann (2017) and Schoonjans (2017), the constructionhood of certain multimodal patterns depends not only on the *frequency* with which certain gestures co-occur with particular verbal constructions, but also on how *salient* and *typical* the gestures are for the constructions, which can be measured by applying collostructional analysis (Stefanowitsch & Gries, 2003). Using an automatic computer-vision system, Uhrig (2021) conducted such analysis, or what he termed *crossmodal collostructional analysis* and investigated the well-known association between the words *yes/no* and corresponding head movements (vertical/horizontal). By defining a collostruction broadly as ‘the combination of two constructions on arbitrary levels of representation that occur significantly more frequently together than expected’ (Uhrig, 2021, p. 257), he argued for the important qualitative distinction between multimodal constructions and *crossmodal collostructions*. While multimodal constructions are gestalt-like units with both linguistic and gestural information in their formal poles, crossmodal collostructions are moderately associated combinations of linguistic and gestural constructions that are compositional. Consequently, Uhrig (2021, 2022) argued that crossmodal collostructional analysis can quantify the degree of association and therefore rank candidates on a cline ranging from free combinations (no association, e.g., air quotes) through crossmodal collostructions (moderate association, e.g., *yes/no* and corresponding vertical/horizontal head movements, in that they frequently occur together but can be used independently of one another)¹ to multimodal constructions (strong association, e.g., German deictic *so* (‘like this’) and its accompanying iconic/pointing gesture, Ningelgen & Auer, 2017).² With this in mind, any of the multimodal patterns examined

¹In contrast, Andrén (2014) argues for the constructional status of headshakes combined with negative response particles from a developmental perspective, based on longitudinal data from six Swedish children aged 18 to 30 months.

²Ningelgen and Auer (2017) do not provide a quantitative analysis of the German deictic *so*, as it occurred only three times in their data. Instead, they argue against the multimodal constructional status of the non-deictic *so* as a vagueness/focus marker. However, the claim that the German deictic *so* qualifies as a multimodal construction would not be controversial, as it can pass a deletion test (the deletion of a kinesic element makes the co-occurring deictic *so* uninterpretable), and thus it would correspond to what Ziem

quantitatively thus far in the CxG literature may not be qualified as multimodal constructions, since apart from Uhrig (2021), no other studies have ever examined ‘the frequency of the [target] gesture in other constructions, the frequency of other gestures in the [target] construction as well as the frequency of other multimodal gestures and constructions’ (Hoffmann, 2017, p. 4). It is this gap that the present study seeks to bridge, by creating a lexicon of gestures (Kipp, 2004). The notion of gesture lexicon and gesture lemmas (its lexical entries) will be explained in more detail in Section 2.1.

The choice of the target constructions in this study is motivated by Hinnell’s (2018) study, in which the multimodality of English aspect-marking auxiliary verbs was investigated. By introducing the notion of action phases, which were defined as ‘the number of separate segments within the gesture stroke’ (Hinnell, 2018, p. 784), Hinnell showed that the gestures accompanying open aspect auxiliary verbs (*continue, keep*) involved two or more action phases, which jointly denoted the ‘multiplex’ (Talmy, 2000) semantics of the open aspect constructions. However, contrary to her expectations, which stemmed from Ladewig’s (2011, p. 6) finding that the referential uses of cyclic gestures referred to ‘ongoing events in every instance’, cyclic gestures occurred with only 17% of open aspect verbs. This, subsequently, leads the present study to look into other English expressions that mark open aspects, which are, two types of conjoined adverbials that show constructional properties (i.e., formal or semantic idiosyncrasy, Goldberg, 1995; Hilpert, 2019). The first type, *reduplicative adverbial constructions* (RACs), refers to the expressions in which the same adverbs are coordinated, including *over and over, again and again, and on and on*. These conjoined adverbials are likely metonymic expressions of potentially endless repetitions of *over/again/on*, and therefore “carry different specialized meanings than their singleton counterparts” (Rice, 1999, p. 241). The second type, *opposite adverbial constructions* (OACs), consists of conjoined antonyms such as *back and forth, up and down, and in and out*. Contrary to the RACs, these irreversible adverbials inherently have spatial dimensions in their meanings. Hereafter, these six expressions will collectively be referred to as the [ADV and ADV] constructions,³ which mark, or even emphasize iterative, habitual, or durative aspect (Rice, 1999). Consider the following examples, in which the [ADV and ADV] constructions coerce the whole situation type of a sentence into an iterative or habitual interpretation.

- (1) a. The author revised the manuscript.
- b. The author revised the manuscript *again and again*.

(2017, p. 5) calls a “inherently multimodal construction.” Nevertheless, note that the formal features of the gestural component of this multimodal construction remain unspecified and are therefore highly schematic in nature.

³The term [ADV and ADV] construction is used for convenience to collectively refer to the six target constructions. However, this does not mean that this paper argues for the existence of such a highly schematic mental representation. Additionally, note that the term specifically refers to these six target constructions and does not encompass other possible instantiations of conjoined adverbials, including formally and semantically similar expressions such as *time and (time) again* for the RACs and *here and there* for the OACs. The exclusion of these similar expressions is based on the following criteria: for the RACs, the focus is on constructions that feature the reduplication of the same adverb as an *obligatory* element; for the OACs, the analysis is restricted to spatial antonyms that convey the meaning of DIRECTIONALITY (see also Section 4.2).

- (2) a. We sent letters to each other.
 b. We sent letters *back and forth* to each other.

The present study investigates the multimodality of these constructions, with the following principal research questions: (i) Are there any specific gestures that are frequently used with the [ADV and ADV] constructions? (ii) If any, how strong are the associations between those gestures and the constructions? That is, are they strong enough to be qualified as multimodal constructions, or more like crossmodal collostructions? These two research questions, respectively, correspond to the two criteria for identifying multimodal constructions, namely, (i) frequency and (ii) salience (Hoffmann, 2017; Schoonjans, 2017). It should also be noted that the present study follows Kendon (2004) in defining gestures as body movements that have ‘*features of manifest deliberate expressiveness*’ (pp. 13–14, emphasis original; cf. Müller, 2014), with a particular focus on hand gestures. Methodologically, the first research question is addressed by identifying gestures that occur frequently with the target constructions. With regard to the second research question, by annotating as reference data a moderate number of gestures that occur with other constructions, crossmodal collostructional analysis is conducted, whereby some combinations of gestures and target constructions are shown to be potential multimodal constructions while others to be more like crossmodal collostructions. With obtained results, Section 4 also discusses the embodied motivations for multimodal constructions, thereby providing implications for future research in multimodal CxG.

2. Materials and methods

This section provides detailed descriptions of the methodological parts of this study. Sections 2.1 and 2.2 present three methodological concepts that are essential for the present study. In Section 2.3, the description of a multimodal corpus used to collect data is provided. The following two sections, Sections 2.4 and 2.5 demonstrate how the data were collected and annotated. Subsequently, Section 2.6 reviews two inter-coder reliability experiments. As a summary, Section 2.7 provides an overview of the annotated data.

2.1. Gesture lexicon and gesture lemma

While gestures seem to be highly spontaneous and free from standards of form at first, various researchers have found that they exhibit fairly stable form-meaning relations, even when they are obviously not emblems (cf. *recurrent gestures*, Ladewig, 2024; Müller, 2017). Hence, similar to emblems, in a given cultural group, there appears to be a shared lexicon or inventory of conversational gestures, which can be referred to as a *gesture lexicon* (Kipp, 2004; Kipp et al., 2007).

For instance, the ‘cyclic gesture’ (Ladewig, 2011; Ruth-Hirrel, 2018), where a speaker makes a continuous rotational movement of the hand, seems to be performed when talking about continuity, repetition, or change (McNeill, 1992, pp. 159–162). Another fairly conventionalized gesture is the ‘Palm Up Open Hand’ (Müller, 2004, p. 234), which is ‘characterized by a specific hand shape and orientation: palm open, fingers extended more or less loosely, palm turned upwards’.

As Kipp et al. (2007, p. 331) have noted, however, ‘while such forms appear to be universal, there is still much inter-speaker and intra-speaker variation in terms of the

exact position of the hands and their ensuing trajectory'. The notion that subsumes such variation is the gesture lemma, which 'can be taken as prototypes of recurring gesture patterns where certain formational features remain constant over instances' (Kipp et al., 2007, p. 328). Importantly, a gesture lemma is a *formal* generalization of similar gestural patterns, divided by gesture classes in terms of their functions, that is, 'a lexicon does not include the meaning of its entries' (Kipp, 2004, p. 128).

While this particular focus on gestural forms may seem extreme, Kipp's approach is in line with Kendon and his followers' descriptive, form-based approach, which starts with 'a detailed analysis of gesture form – both regarding their articulatory (etic) and their meaningful (emic) features or clusters of features – as a point of departure to reconstruct meaning' (Müller, 2014, pp. 138–139).⁴ Thus, although a gesture lemma is not determined by its potential meanings, it is likely that a set of gesture lemmas sharing specific formational features also share similar (referential) functions, as demonstrated by previous studies on 'gesture families' (see Fricke et al., 2014, pp. 1633–1635; Müller, 2014, pp. 135–137 for overviews).⁵ The description of an annotation and identification of gesture lemmas will be presented in Section 2.5.

2.2. Crossmodal collostructional analysis

In unimodal CxG, *collostructional analysis* has gained widespread popularity and been extensively utilized for exploring the relationships between words and grammatical constructions. Collostructional analysis has three sub-types, of which simple collexeme analysis (Stefanowitsch & Gries, 2003) is applied in the current study to the co-occurrence of constructions and gestures. The constructions investigated in this simple collexeme analysis are typically syntactic constructions that have an open slot for lexemes to occur in, and lexemes that are attracted or repelled to the constructions are called collexemes. The associational strength between a given construction and a collexeme is called *collostruction strength*.

In *crossmodal collostructional analysis* (Uhrig, 2021, 2022) as introduced in Section 1, let us refer to target gesture lemmas that are attracted or repelled to a target construction as *co-gestures* of the construction, although the target constructions investigated in this study are phrasal constructions that do not have open slots. To conduct such crossmodal collostructional analysis, as with the unimodal collostructional analysis, the following four different frequencies of occurrence of a target construction (C) and a co-gesture (G) must be retrieved from the corpus under investigation:

- (3) a. The frequency of G synchronized with C
- b. The frequency of G synchronized with all other constructions
- c. The frequency of C synchronized with gesture lemmas other than G
- d. The frequency of all other constructions synchronized with gesture lemmas other than G

⁴As will be demonstrated in Section 2.5.2, the process of determining the formational features of a gesture lemma would correspond to identifying emic features as mentioned in this quote.

⁵The gesture families that turned out to be particularly relevant to the present study are cyclic gestures (for the RACs; Ladewig, 2011; Ruth-Hirrel, 2018) and bidirectional gestures (for the OACs). For more details, see Section 2.7 and Appendix A.

These frequencies are then entered into a 2×2 contingency table, from which collocation strength is calculated by means of the Fisher exact test. (However, as in Schmid and Küchenhoff (2013), reference measures for collocation strength are still subject to debate. We will return to this issue in Section 3.)

With regard to the *synchronization* of G with C, for a gesture to be qualified as synchronized with a target construction, either of the following conditions or both must be met:

- (4) a. The lexical affiliate (Schegloff, 1984) of G includes C.
- b. The expressive phase (Kita et al., 1998) of G, often a single stroke, overlaps C.

Regarding (4a), the coding of lexical affiliation (i.e., co-expressiveness) follows Kipp's (2004) NOVACO scheme. As for (4b), if there is no overlap between them – that is, if the offset of G's expression phase precedes the onset of C, or if the offset of C precedes the onset of G's expression phase – they are not coded as synchronized. Given that gestures can sometimes precede their lexical affiliates, this definition of synchrony may result in cases in which a target construction has two (or more) synchronized gestures, each satisfying at least one of the conditions.

2.3. TED corpus search engine (TCSE)

The present study uses data from TED Talks, accessed through the TED Corpus Search Engine (TCSE; <https://yohasebe.com/tcse>), an online corpus system that searches transcripts of over 5,300 TED Talks,⁶ 'allowing users to query surface text forms, lemmas, parts-of-speech, or their combinations' (Hasebe, 2015, p. 1). When running queries in TCSE, corresponding videos are retrieved for the matched texts, making it a publicly available multimodal corpus. As a resource for research on multimodality, TED Talks have already received scholarly attention, and several studies have in fact been conducted with a focus on the role of gestures in English TED Talk presentations (e.g., Harrison, 2021; Masi, 2020). However, as pointed out by Hasebe (2015, p. 181), TED Talk speakers also include non-native speakers of English, even though they are proficient. This may pose a critical problem for research like this study, which examines potential mental/social representations of constructions. Nonetheless, according to Hasebe's (2018) statistical analysis (Kendall's *tau*) comparing lexical data between TCSE and COCA, there is a substantial correlation between the two corpora (Kendall's *tau* = 0.62, $p < 0.001$), as shown in Figure 1. This demonstrates that while TCSE reflects a specific register of TED Talks, namely, presentations, it still preserves the basic characteristics of the English language (Hasebe, 2018, p. 168), making it valid for theoretical investigations of English. Accordingly, the results of the present study should also be considered valid, at least as a preliminary study that offers insights for research on multimodal constructions.

⁶As of February, 2025.

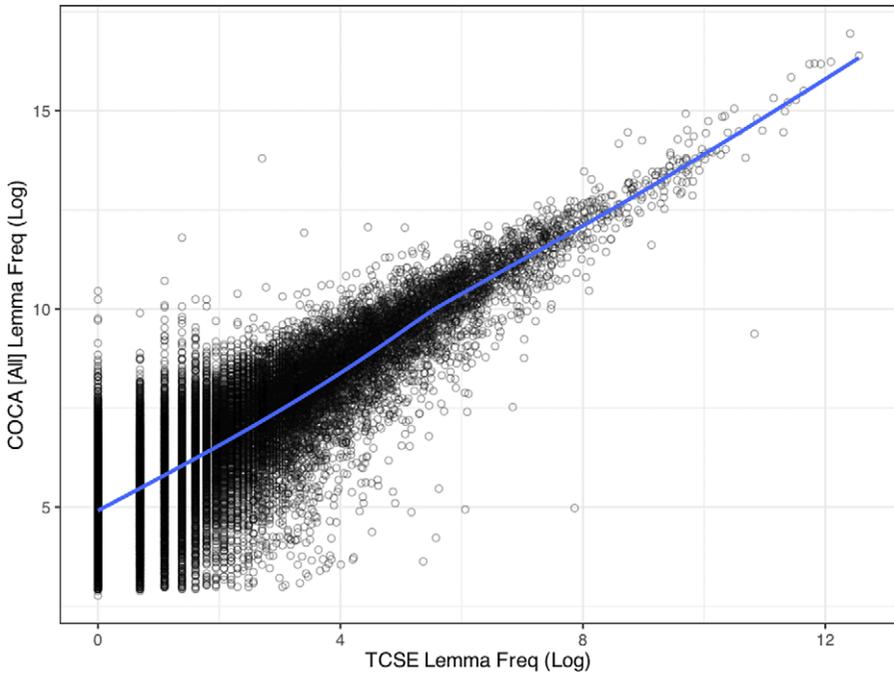


Figure 1. A plot of relative frequencies of lemmas in TCSE and COCA [All] (Adapted from Hasebe, 2018, p. 168).

2.4. Data collection

The data were collected from TCSE in October and November 2022. For each search string, POS tags were not assigned so that it could retrieve as many instances as possible including those that were not used as adverbs (e.g., *She went in and out of hospitals*). The retrieved instances were first viewed to determine whether they were valid or not; that is, whether the speaker's hands were visible on screen while uttering the target constructions, regardless of whether they were gesturing or not. In total, valid instances amounted to 576, and Table 1 shows how they were distributed across the constructions under scrutiny.

For each of these instances, the utterance containing the target construction was clipped. The approximate length per clip was 10 seconds: 3–5 seconds on either side of the target construction.

2.5. Data annotation

For each clipping, gestures synchronized with the target constructions were annotated in ELAN (ELAN, 2022). Such gesture annotation was conducted by following Kipp's (2004) NOVACO scheme. The annotation consists of the following three different procedures, each of which is described in detail in the following subsections: (i) segmentation of movement phases and identification of movement phrases (Section 2.5.1); (ii) identifying and creating gesture lemmas to construct a gesture

Table 1. Numbers of instances of each construction under investigation

Constructions	Number of instances
RACs	315
<i>Over and over</i>	160
<i>Again and again</i>	107
<i>On and on</i>	48
OACs	261
<i>Back and forth</i>	110
<i>Up and down</i>	92
<i>In and out</i>	59
Total	576

lexicon (Section 2.5.2), which is then used in (iii) the subsequent assignment of a gesture lemma to each movement phrase (Section 2.5.3).

2.5.1. Movement phases and phrases

The segmentation of movement phases and organization of movement phrases in the NOVACO scheme (Kipp, 2004) are based on a coding scheme proposed by Kita et al. (1998). As described in the scheme and also known widely, there are types of movement phases, including *preparation*, *stroke*, *retraction*, *partial retraction*, *dependent hold* (*pre-stroke* or *post-stroke*), and *independent hold*. Of particular relevance to the present study is the distinction between a *single-segment phase* and a *multi-segment phase* (Kita et al., 1998, pp. 29–30), both of which can constitute a stroke. According to Kita et al. (1998), a stretch of bodily movement is decomposed into two phases “if there is an abrupt change of direction in the hand movement, AND there is a discontinuity in the velocity profile of the hand movement before and after the abrupt direction change” (p. 29, emphasis original). Accordingly, even if there is an abrupt change in direction, multiple movement segments that do not have a discontinuity in between, such as a hold, are still regarded as constituting a *single phase*, namely, a *multi-segment phase*. Somewhat related to this distinction, the presence of a hold becomes especially significant when we consider beat gestures as well. While multiple up-and-down manual movements are seen as a *single beat gesture* if there is no hold in between, those with holds are considered a sequence of *several beat gestures*.

Movement phases as described in most of this section thus far concern the *internal organization* of gestures. Then, the *movement phrase*, into which several movement phases are grouped together, is a notion that corresponds to the term *gesture* in its most common usage. In annotation, movement phases are coded first. Subsequently, movement phases are grouped together into a movement phrase according to the following rule based on Kipp (2004, pp. 124–125) and Kita et al. (1998, p. 27).

Movement phrase = (preparation) *expressive phase* (*retraction*)
Expressive phase = (prestroke hold) stroke (poststroke hold) |
independent hold
Retraction = retraction | partial retraction

According to the rule, a gesture (namely, a movement phrase) must always include one expressive phase that consists at least of a single- or multi-segment stroke phase,

or an independent hold. Segmentation and identification of these phases were shown to be reliably coded by Kita et al. (1998).

2.5.2. Identifying and creating gesture lemmas

An equally important step as the one described in the last section is to construct a gesture lexicon by identifying and creating gesture lemmas from empirical data. This step is best explained in Kipp's own words:

For annotation, first a lexicon of gesture lemmas must be collected from the empirical material. This collection step is done by systematically sifting through the empirical data and cataloguing the gestures by comparing them to the already found ones. Once the lexicon is complete, annotation can begin. During annotation no new lemmas may be added. Otherwise inconsistencies could emerge. Instead, a rest category must serve as a container for gesture lemmas not yet located in the lexicon (Kipp, 2004, p. 128).

In the current study, this step was conducted by examining a randomly selected half of all 576 clippings, with reference to Kipp's (2004) already established lexicon of 63 gesture lemmas. Once a gesture lemma from Kipp's (2004) lexicon was attested in the data, the lemma was added to the lexicon for the present study. However, the need for more subdivisions and the creation of lemmas that were not present in Kipp's (2004) lexicon emerged during this step. Such procedure was conducted following the NOVACO scheme (Kipp, 2004, pp. 123–137), which starts with a high-level classification of gestures into four gesture classes. Figure 2 summarizes the classification and serves as a decision tree for coding which gesture class applies to the gesture being annotated.

The decision tree is used in such a way that the coder goes from top to bottom and considers each class from left to right with reference to criteria that are specified for each class. Accordingly, the coder first checks whether the hand/arm movement in question is communicative or not. Non-communicative

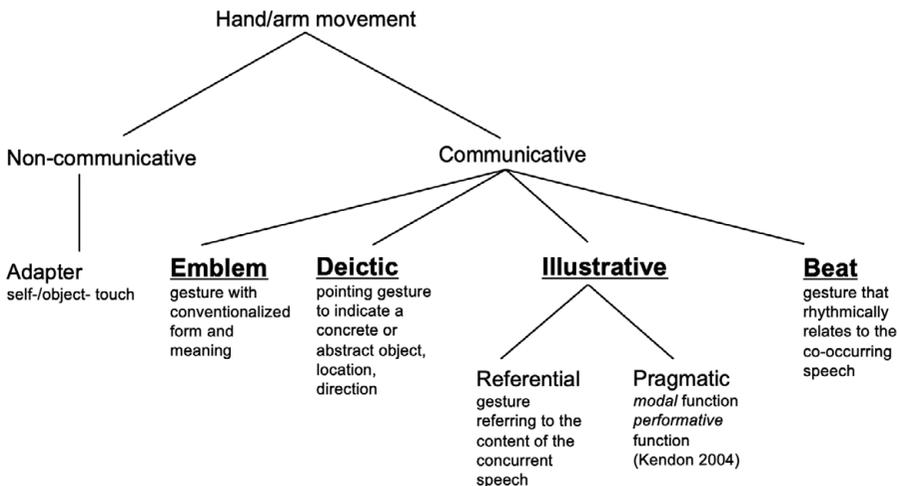


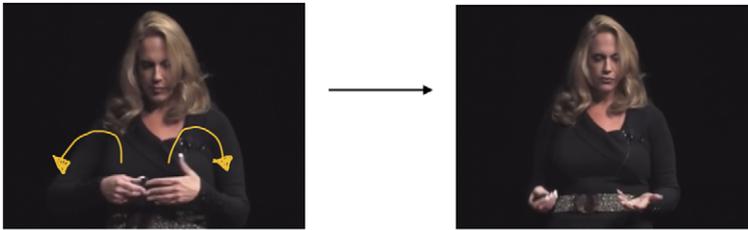
Figure 2. A decision tree for classifying gesture classes (based on Kipp, 2004, p. 125).

movements, such as adapters, are not a concern for the present study. Communicative movements, on the other hand, are further categorized into the following four classes: emblems, deictics, illustratives, and beats. This classification is a modified version of Kipp (2004, pp. 39–43), which is in turn based on traditional classifications such as Efron (1941), Ekman and Friesen (1969), and McNeill (1992). The key difference from Kipp (2004) is that the illustrative class is not further specified in this study so it broadly subsumes those that are not classified as emblems, deictics, or beats.

It should be noted that these classes are not mutually exclusive since different criteria are applied to different classes. However, as Kipp (2004) claimed, classifications are still possible, by identifying the *predominant* function and by defining how to categorize gestures in cases of ambiguity: the left branch is taken when in doubt. As has been mentioned just above, each class is checked from left to right, referring to criteria for each.⁷ Criteria differ on a scale of whether they are based on the form of gestures or on the meaning/function. The degree to which they are based on form increases when going to the left; and so does meaning/function when going to the right. Emblems are the category that relies most on form-based criteria. Essentially, emblematic gestures conform to specific standards of form and have corresponding word-like meanings. In addition, they have the potential to be used as a replacement for speech, which is another important criterion for emblems. Deictics also rely on form-based criteria, be they of the concrete or abstract type (McNeill, 1992, p. 173). Canonically, deictic gestures consist of a single index finger or a flat open hand being pointed toward a (virtual) object, thus having the deictic function. Illustratives, by contrast, are not subject to formal conditions, rendering them solely a functional category. The illustrative class is a superordinate category, leaving the room for more subdivisions, which nonetheless are not a concern here. It subsumes referential gestures (gestures related to the propositional content of the concurrent speech; cf. iconics and metaphoric, McNeill, 1992) and pragmatic gestures (gestures that convey ‘features of an utterance’s meaning that are not a part of its referential meaning or propositional content’ (Kendon, 2004, p. 158); however, the illustrative class only includes the modal and performative functions). Finally, beats are a rest class into which a gesture is placed if none of the categories in the left branches apply. As with illustratives, form is not a criterion for the class membership. Beat gestures relate to the co-occurring speech rhythmically, signaling that ‘the speaker feels [that the word/phrase the beat gestures accompany] is important with respect to the larger discourse’ (McNeill, 2005, p. 40).

After identifying the class of the gesture in question, the next step is to specify the formal features of the gesture lemma. As stated above, the notion of the gesture lemma is a formal generalization of gestures that exhibit similarity across instances, and thus, it does not refer, in principle, to the semantic aspect of the gestures (except for emblems). Following Kipp (2004, pp. 128–131), in identifying gesture lemmas, coders must take into consideration *form dimensions*, which determine whether a gesture is a new lemma altogether or merely a formal variant of an existing gesture lemma. A dimension is referred to as *formational* (Kipp, 2004, p. 129) if changing a gesture along the dimension changes the lemma. The form dimensions are the following:

⁷For more detailed descriptions of criteria for each class, see Kipp (2004, pp. 126–128).



Description	Open, cup-shaped hand is brought into palm-up position in a small arc.
Features	HS: open, Orient: PU, Loc: stomach, Move: small arc (movement from the wrist)
Frequency	OVER: 2 (1.4%), AGAIN: 4 (3.8%), ON: 1 (2.5%) BF: 1 (0.9%), UD: 1 (1.2%) Reference data: 33 (5.9%)

Figure 3. An example of gesture lemmas (*Illustrative.Cup-Flip*)⁸.

- (5) a. Hand shape (HS)
- b. Hand location (Loc)
- c. Hand orientation (Orient)
- d. Hand/arm movement (Move)
- e. Handedness (1H or 2H)
- f. Concomitant shoulder movement (Sh)
- g. Concomitant facial expression (Fe)

Shoulder movement and facial expression can be a formational dimension because some emblems can only be differentiated by them (Kipp, 2004, p. 129). In addition to the specification of formational features, a gesture lemma includes the verbal description of its form. An example of gestures lemmas is shown in Figure 3. The second row ('Features') in Figure 3 outlines the formational features of the gesture lemma *Illustrative.Cup-Flip*. While it specifies several features (e.g., the hand shape, hand orientation) as formational (i.e., essential for the gesture lemma), it does not include handedness. This means that the differences in handedness (one hand vs. both hands) are nothing but formal variants within this lemma. Importantly, handedness serves as a formational feature for certain other gesture lemmas (e.g., Figures 6 and 9).

This identification and collection step resulted in a lexicon of 62 gesture lemmas. Of these lemmas, 37 lemmas were identical to Kipp's (2004) ones. The rest of the 25 lemmas were either completely new for this study or modified versions of Kipp's (2004) lemmas.⁹ The list of all 62 gesture lemmas is presented in Appendix A.¹⁰

⁸Romero, B. (2012, April). *Gaming for Understanding*. TED Talk. https://www.ted.com/talks/brenda_romero_gaming_for_understanding.

⁹Of these 25 gesture lemmas, 13 belong to cyclic gestures and bidirectional gestures, which frequently co-occurred with the [ADV and ADV] constructions (see Appendix A). These new lemmas were created from systematic variations within what was previously a single lemma.

¹⁰While the idea that there are culturally specific repertoires of fairly conventionalized gestures is supported by recent studies on recurrent gestures, the gesture lexicon created in this study would not be empirically plausible, as it includes lemmas that are more of the singular gesture type (Müller, 2017) in nature

2.5.3. Annotating gesture lemmas

Having completed the collection of all gesture lemmas, the annotation of gestures in each clipping was conducted. First, for each clipping, gestures that were *synchronized* with the target constructions (for the definition of synchrony, see Section 2.2) were annotated with regard to movement phases, and the movement phases annotated were then grouped together to form movement phrases. In the total of all 576 clippings, 532 gestures (i.e., movement phrases) were annotated as synchronized with the target constructions, and each movement phrase was assigned a gesture lemma from the gesture lexicon constructed for the present study. These 532 gestures are necessary for conducting crossmodal collostructional analysis, and more specifically, for the first cell (3a in Section 2.2; the frequency of G synchronized with C) and the third cell (3c; the frequency of C synchronized with gesture lemmas other than G).

However, crossmodal collostructional analysis requires other frequencies, namely, the frequency of G synchronized with all other constructions (3b) and the frequency of all other constructions synchronized with gesture lemmas other than G (3d). Therefore, gestures in synchrony with other constructions than the [ADV and ADV] constructions were needed as a reference dataset. To this end, an additional annotation step was carried out: gestures that occurred on either side of the target constructions, which were not synchronized with them, were annotated from a randomly selected half of all 576 clippings. This additional annotation step resulted in 560 gestures (i.e., movement phrases), each of which was also assigned a gesture lemma. During this step, movement phrases that did not fit any of the gesture lemmas were put into the rest category named ‘unknown’ (Kipp, 2004, p. 130).

2.6. Coding reliability

To assess whether the lexicon of 62 gestures could be reliably coded, two intercoder reliability experiments were carried out.¹¹ The author of this paper and two different coders (Coder A and Coder B) participated in the two experiments, respectively. Section 2.6.1 provides the descriptions of the first intercoder reliability test, and Section 2.6.2 the second one.

2.6.1. First intercoder reliability experiment

In this experiment, two coders (the author and Coder A) independently annotated 150 clippings, which were randomly selected from the 497 clippings that involved gestures synchronized with the [ADV and ADV] constructions.¹² These clippings contained pre-annotated movement phases and movement phrases that were assembled from those movement phases. However, in each clipping, the movement phrase

(i.e., they occurred only once or a few times in the data). Thus, a compromise must be reached that the gesture lexicon for the present study is rather a theoretical construct, as it is necessary for conducting crossmodal collostructional analysis.

¹¹When the two experiments were conducted, the gesture lemmas I.2H-Alternation-Sagittal and I.1H-Small-To-Fro (see Figures 6 and 9 in Section 2.7) were subsumed under a single lemma. Additionally, unlike in the eventual study, Adapter was still counted as a lemma back then, as it was included in Kipp’s (2004) gesture lexicon.

¹²Out of all 576 clippings, 497 clippings involved gestures synchronized with the [ADV and ADV] constructions. This means that the rest of 79 clippings did not include gestures co-occurring with the target constructions.

Table 2. Results of the first intercoder reliability experiment

	%	Cohen's k
Trial 1	70.0	0.66
Trial 2	82.7	0.80

synchronized with the [ADV and ADV] construction was not allocated a corresponding gesture lemma. The task of the coders was to assign a gesture lemma to each of the 150 movement phrases.

This task resulted in a 70% agreement between the coders and a Cohen's k value of 0.66, which is very close to the score for "tentative conclusions to be drawn" (Carletta, 1996, p. 252). Examining the disagreement between coders, 56% were due to confusion between similar gestures in form. The remaining 44% were due to fundamental differences in interpreting functions of gestures. After defining stricter criteria for distinguishing similar gestures (see Appendix B), the task was conducted again. This resulted in an 82% agreement and a Cohen's k value of 0.80, which is good reliability according to Carletta (1996). Table 2 shows the results of the two trials.

2.6.2. Second intercoder reliability experiment

The second experiment replicated Kipp's (2004, pp. 140–141) coding reliability experiment. Two coders (the author and Coder B) independently annotated a 2:16-minute snippet of a TED Talk, in which the [ADV and ADV] constructions were not uttered. These data contained pre-annotated movement phases. In accordance with Kipp (2004, p. 140), the task of the coders was twofold: first, the coders were asked to assemble movement phases to movement phrases, which is called 'segmentation'. They were then asked to assign a gesture lemma to each phrase, which is called 'classification'.

For the classification task, only the movement phrases that matched between the two coders in the segmentation task were taken into consideration. The number of such movement phrases was 49.

As seen in the first column of Table 3, the reliability percentage of the segmentation task was very high. However, as columns two and three show, the classification task yielded critical results. A Cohen's k value of 0.43 indicates that the reliability was not very high. The difference between the results of the first and second experiments would probably be due to the contexts in which gestures were performed. In the first experiment, all gestures were synchronized with the target constructions, rendering it easier for the coders to interpret the functions of gestures. On the other hand, since the contexts of the second experiment were not limited to certain types of constructions, the coders could not agree on the gestures that were similar in form but functionally distinct. Therefore, after defining the criteria for coding formally similar gestures across different functional classes (see Appendix B), the classification task

Table 3. Results of the second intercoder reliability experiment

	Segmentation	Classification	
	%	%	Cohen's k
Trial 1	90.7	46.9	0.43
Trial 2	–	83.7	0.82

was conducted again. This resulted in an 83% agreement and a Cohen's k value of 0.82, which is considered good reliability according to Carletta (1996).

To conclude this section, the experiments show that gesture lemmas can be reliably coded not only in specific contexts, such as when all gestures are synchronized with certain kinds of constructions, but also in general contexts. In line with Kipp's results (2004, pp. 140–141), the improvement in reliability between trials 1 and 2 in both experiments indicates that more rigid definitions and documentation in the coding manual can reduce more sources of error.

2.7. Overview of annotated data

After the two intercoder reliability experiments, all the annotated gesture lemmas, including those in the reference data, were reviewed with reference to the criteria for distinguishing similar gestures and the revised descriptions of some gesture lemmas (see Appendices A and B). This section presents an overview of the annotated data.

The annotated clippings collectively form a small multimodal corpus in which 407 speakers are performing 1,092 gestures in total. Of all these gestures, 294 gestures are synchronized with the RACs, and 238 gestures with the OACs; the remaining 560 gestures are synchronized with other constructions, which occur on either side of the target constructions in a randomly selected half of all 576 clippings, forming a reference dataset for conducting crossmodal collostructional analysis. With regard to the RACs and OACs, a notable fact is that they were quite frequently accompanied by gestures: for the RACs, the gesture accompaniment rate is as high as 85.4%; for the OACs, it is even higher at 87.4%.¹³

Table 4 shows the distribution of gesture lemmas that co-occurred with the RACs. Of the total 294 gestures, the gesture lemma *I. Cyclic-Sagittal* (Figure 4) was used most frequently with the RACs, with a proportion of 36.4%. Overall, gestures with cyclic movements (highlighted in red) occurred quite frequently, totaling 131 instances (44.6%). Equally importantly, the Beat gestures (highlighted in green) also accompanied substantial instances of the RACs, amounting to 27.6% of gestures observed with the RACs. As we will see in Table 6, however, this may be a general tendency considering the environment in which the speakers were situated, that is, TED Talks.

On the other hand, Table 5 displays the distribution of gestures that occurred concurrently with the OACs. The five most frequent gestures all showed a bidirectional movement (Figures 5–9). The bidirectional gestures including these five (highlighted in blue) accounted for as much as 73.1% of the total 238 gestures, aligning with the semantics of the OACs. Another interesting point is that the proportion of Beat gestures is much smaller than that of the RACs (Table 4) and the reference data (Table 6). We will address this issue in Section 3.

Table 6 summarizes the gesture lemmas that occurred in the vicinity of the target constructions in half of all 576 clippings, which serve as reference data specifically used for conducting crossmodal collostructional analysis. The contexts in which the

¹³Out of all 315 instances of the RACs, there are 269 instances (85.4%) where at least one gesture occurs in synchrony with the RACs. Thus, the total number of 294 in Table 4 means that in 25 instances, two gestures are synchronized with one RAC (see the definition of synchrony in Section 2.2). Similarly, while the gesture accompaniment rate of the OACs is 87.4% (out of all 261 instances, in 228 instances at least one gesture occurs in synchrony with the OACs), the total number of 238 in Table 5 means that two gestures are synchronized with one OAC in 10 instances.

Table 4. Distribution of gesture lemmas co-occurring with the RACs¹⁴

	Gesture lemma	#	%		Gesture lemma	#	%
1	I.Cyclic-Sagittal	107	36.4	17	E.So-What	3	1.0
2	Beat	81	27.6	18	I.Chop	2	0.7
3	I.Cyclic-Frontal	15	5.1	19	I.Cyclic-Transverse	2	0.7
4	I.Frame	13	4.4	20	I.Bidirection-Sagittal	2	0.7
5	I.PUOH	8	2.7	21	I.Bidirection-Horizontal	1	0.3
6	I.2H-Alternation-Sagittal	7	2.4	22	I.Fling-Down	1	0.3
7	I.Cup-Flip	7	2.4	23	E.Wipe	1	0.3
8	D.Space	6	2.0	24	I.Away	1	0.3
9	E.Dismiss	5	1.7	25	E.Hand-Rub	1	0.3
10	I.1H-Small-To-Fro	4	1.4	26	I.Down	1	0.3
11	I.Location	4	1.4	27	E.Finger-As-Entry	1	0.3
12	I.Clockwork	4	1.4	28	I.Explode	1	0.3
13	I.Umbrella	3	1.0	29	E.Doubt-Shrug	1	0.3
14	I.Bidirection-Vertical	3	1.0	30	I.Dome	1	0.3
15	I.Emerge	3	1.0	31	I.Pluck-Apple	1	0.3
16	I.Inverse-Cyclic-Sagittal	3	1.0	32	E.Attention	1	0.3
			Total				
							294



Description	Hand moves in circles where in the upper arc the hand moves away from the body.
Features	Move: circular movement/parallel to the sagittal plane
Frequency	OVER: 55 (37.2%), AGAIN: 37 (34.9%), ON: 15 (37.5%) BF: 4 (3.8%), UD: 1 (1.2%), IO: 1 (2.0%) Reference data: 34 (6.1%)

Figure 4. I.Cyclic-Sagittal¹⁵.

reference data were collected are assumed to be more neutral than those in which the gesture lemmas in Tables 4 and 5 were collected. However, we must also acknowledge that the contexts of the reference data may be skewed in favor of the target constructions, as they were adjacent to those constructions. This might be a cause

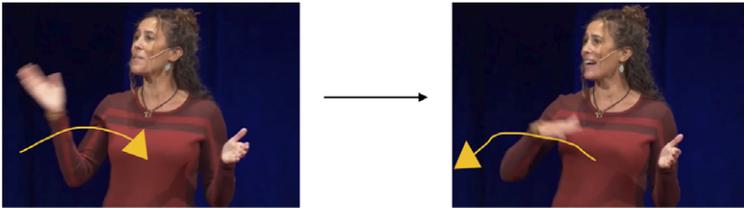
¹⁴The # column shows the absolute frequency of each gesture lemma. The % column shows the relative frequency of each gesture lemma. The capital letters “E”, “D”, and “I” stand for “Emblem”, “Deictic”, “Illustrative” respectively.

¹⁵For Footnotes 15–20, the provided links direct to the specific timestamps where the gesture lemma in question is used with the [ADV and ADV] constructions.

Van Buren, D. (2017, November). *What a World without Prisons Could Look like*. TED Talk. <https://www.youtube.com/watch?v=m6X1i8khmt8&t=486s>

Table 5. Distribution of gesture lemmas co-occurring with the OACs

	Gesture lemma	#	%		Gesture lemma	#	%
1	I.Bidirection-Horizontal	45	18.9	16	I.Frame	3	1.3
2	I.2H-Alternation-Sagittal	38	16.0	17	I.Cross	3	1.3
3	I.Bidirection-Vertical	35	14.7	18	I.Umbrella	3	1.3
4	I.Small-Bidirection-Vertical	17	7.1	19	I.PUOH	3	1.3
5	I.1H-Small-To-Fro	16	6.7	20	E.Wipe	2	0.8
6	Beat	12	5.0	21	I.Cyclic-Frontal	2	0.8
7	I.Location	8	3.4	22	I.Open	2	0.8
8	I.Bidirection-Sagittal	8	3.4	23	E.So-What	2	0.8
9	I.Cyclic-Sagittal	6	2.5	24	I.Cup-Flip	2	0.8
10	I.2H-Alternation-Vertical	5	2.1	25	I.Fling-Down	1	0.4
11	I.Away	5	2.1	26	I.Emerge	1	0.4
12	I.Vibrate	4	1.7	27	E.Anticipation	1	0.4
13	I.Approach-Away	4	1.7	28	E.Wave	1	0.4
14	I.Jump-Up	4	1.7	29	I.In	1	0.4
15	I.Play-Catch	3	1.3	30	I.Merge	1	0.4
	Total				238		



Description	Hand moves to one side, and then the other (and repeats the movement several times). If this gesture occurs with both hands, they move in the same direction.
Features	HS: open/forefinger, Loc: →side → other side Move: movement from the elbow
Frequency	OVER: 1 (0.7%) BF: 29 (27.4%), UD: 5 (6.0%), IO: 11 (22.4%) Reference data: 11 (2.0%)

Figure 5. I.Bidirection-Horizontal¹⁶.

for relatively high frequencies of gestures in the reference dataset that also frequently occurred with the target constructions. Nonetheless, this does not necessarily have an advantageous impact on the results of crossmodal collostructional analysis, as it makes it more difficult for those gestures to obtain low *p*-values in the Fisher exact test as well as high odds ratios.

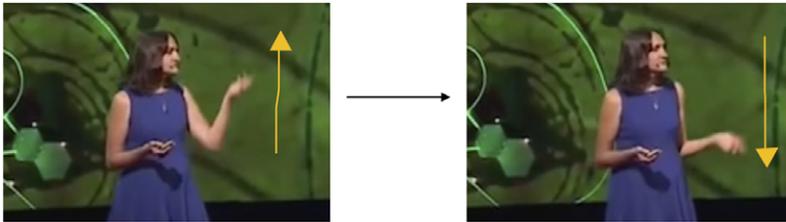
First, most notable among the reference data is the large number of Beat gestures, which is also consistent with the findings in Table 4. This can plausibly be considered a presentation strategy by the speakers of TED Talks to make it visible that they

¹⁶Hardt, M. J. (2019, November) *Are We Interrupting the Kinky Sex Lives of Fish?* TED Talk. https://www.youtube.com/watch?v=Va_8GmU9cmQ&t=122s



Description	Both hands alternate quickly in a back and forth movement on the sagittal axis.
Features	2H, HS: open/forefinger, Loc: in front of torso Move: typically movement from the wrist
Frequency	OVER: 5 (3.3%), AGAIN: 1 (0.9%), ON: 1 (2.5%) BF: 26 (24.5%), UD: 2 (2.4%), IO: 10 (20.8%) Reference data: 11 (2.0%)

Figure 6. I.2H-Alternation-Sagittal¹⁷.



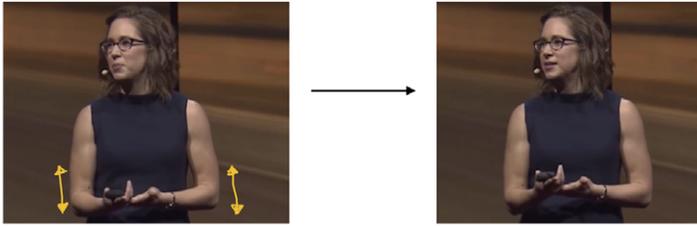
Description	Hand moves in an up-down movement. If this gesture occurs with both hands, they move in the same direction.
Features	Orient: PD/PU, Move: straight up-down motion (from the elbow)
Frequency	OVER: 2 (1.4%), AGAIN: 1 (0.9%) BF: 2 (1.9%), UD: 31 (37.3%), IO: 2 (4.1%) Reference data: 5 (0.9%)

Figure 7. I. Bidirection-Vertical¹⁸.

regard what they are referring to as ‘important with respect to the larger discourse’ (McNeill, 2005, p. 40). Following the Beat gesture is the gesture lemma I. *Cyclic-Sagittal*, with a proportion of 6.1%. This gesture, which was predominant in Table 4 (the RACs), can also be used in different contexts, serving a self-oriented function (word/concept search) and a performative function (e.g., request), as well as

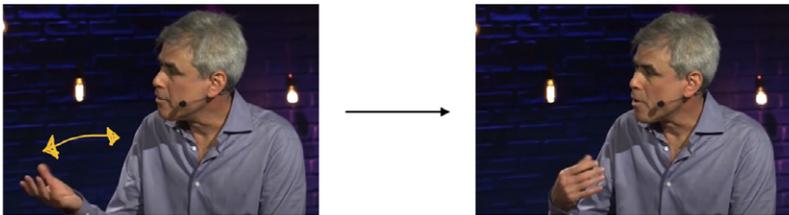
¹⁷Picciolini, C. (2017, November) *My Descent into America’s Neo-Nazi Movement – and How I Got out*. TED Talk. <https://www.youtube.com/watch?v=VM6HZqQKhok&t=978s>

¹⁸Aamodt, S. (2013, June). *Why Dieting Does not Usually Work*. TED Talk. <https://www.youtube.com/watch?v=jn0Ygp7pMBA&t=130s>



Description	Hand/finger moves in a small up-down movement
Features	Orient: PD/PU, Move: up-down motion (from the wrist)
Frequency	BF: 2 (1.9%), UD: 14 (16.9%), IO: 1 (2.1%) Reference data: 2 (0.4%)

Figure 8. I.Small-Bidirection-Vertical¹⁹.



Description	Hand moves quickly back and forth in front of the torso.
Features	lH, HS: open/forefinger, Loc: in front of torso, Move: AB → TB/side → other side (movement from the wrist)
Frequency	OVER: 1 (0.7%), AGAIN: 2 (1.9%), ON: 1 (2.5%) BF: 9 (8.5%), UD: 2 (2.4%), IO: 5 (10.4%) Reference data: 2 (0.4%)

Figure 9. I.lH-Small-To-Fro²⁰.

a referential function conveying the idea of continuity or progress (Ladewig, 2011, pp. 7–13). Including such cyclic gestures, most of those appearing frequently in the reference dataset would be regarded as *recurrent gestures* (Ladewig, 2024; Müller, 2017), having fairly stable formational cores paired with their semantic cores, which can allow for a variety of contextual variants (e.g., I.Away [cf. ‘Away gestures’, Bressen & Müller, 2014], I.Frame [cf. ‘bimanual PVOH’, Mittelberg, 2017], and I.PUOH [Müller, 2004]; see Appendix A). Thus, the findings in Table 6 can also be generally seen as supporting the idea that there are certain regularities in how we gesture and that there are culturally conventionalized repertoires of gestures apart from emblems.

¹⁹Lanier, H. (2017, December). “Good” and “Bad” Are Incomplete Stories We Tell Ourselves. TED Talk. <https://www.youtube.com/watch?v=XZiNVGA78kA&t=268s>

²⁰Haidt, J. (2016, November). Can a Divided America Heal? TED Talk. https://www.youtube.com/watch?v=D-_Az5nZBBM&t=710s

Table 6. Distribution of gesture lemmas co-occurring with other constructions (reference data)

	Gesture lemma	#	%		Gesture lemma	#	%
1	Beat	130	23.2	31	E.Wave	3	0.5
2	I.Cyclic-Sagittal	34	6.1	32	“unknown”	3	0.5
3	I.Cup-Flip	33	5.9	33	I.Merge	3	0.5
4	I.Location	25	4.5	34	I.Vibrate	2	0.4
5	E.So-What	25	4.5	35	I.Small-Bidirection-Vertical	2	0.4
6	I.Emerge	24	4.3	36	E.Finger-As-Entry	2	0.4
7	I.Away	24	4.3	37	E.Block	2	0.4
8	I.Frame	23	4.1	38	E.Refuse	2	0.4
9	I.PUOH	22	3.9	39	D.Self	2	0.4
10	I.Explode	20	3.6	40	I.Walls	2	0.4
11	E.Wipe	18	3.2	41	I.Thought-Grip	2	0.4
12	D.Space	13	2.3	42	I.1H-Small-To-Fro	2	0.4
13	I.Umbrella	12	2.1	43	I.2H-Alternation-Vertical	1	0.2
14	I.2H-Alternation-Sagittal	11	2.0	44	I.Clockwork	1	0.2
15	I.Bidirection-Horizontal	11	2.0	45	E.Anticipation	1	0.2
16	E.Number	11	2.0	46	I.In	1	0.2
17	E.Calm	10	1.8	47	E.Doubt-Shrug	1	0.2
18	E.Fling-Down	8	1.4	48	I.Pluck-Apples	1	0.2
19	I.Jump-Up	8	1.4	49	I.Cyclic-Transverse	1	0.2
20	I.Open	8	1.4	50	I.Bidirection-Sagittal	1	0.2
21	E.Dismiss	6	1.1	51	E.More-Or-Less	1	0.2
22	E.Attention	6	1.1	52	I.Front-Back-Distance	1	0.2
23	I.Bidirection-Vertical	5	0.9	53	E.Big	1	0.2
24	I.Dome	5	0.9	54	E.Indignation	1	0.2
25	D.Addressee	5	0.9	55	E.Small	1	0.2
26	I.Heart	5	0.9	56	I.Grab	1	0.2
27	I.Cyclic-Frontal	4	0.7	57	I.Snatch	1	0.2
28	I.Inverse-Cyclic-Sagittal	4	0.7	58	I.Push	1	0.2
29	I.Down	3	0.5	59	E.Air-Quotes	1	0.2
30	I.Chop	3	0.5	Total		560	

3. Results

With all the annotated data for each of the six variants of the [ADV and ADV] constructions and the reference data, crossmodal collostructional analysis was conducted. The four frequencies necessary for crossmodal collostructional analysis are as follows (as outlined earlier in Section 2.2):

- (3) a. The frequency of G synchronized with C
- b. The frequency of G synchronized with all other constructions
- c. The frequency of C synchronized with gesture lemmas other than G
- d. The frequency of all other constructions synchronized with gesture lemmas other than G

To exemplify the analysis, Table 7 presents a 2 × 2 contingency table for the target construction (C) *over and over* and the co-gesture (G) I.Cyclic-Sagittal (Figure 4), from which the Fisher exact test is performed.²¹

As introduced in Section 1, Hoffmann (2017) and Schoonjans (2017) argued that the constructionhood of a multimodal pattern depends not only on the mere

²¹The Fisher Exact test was conducted using R.

Table 7. A contingency table for crossmodal collostructional analysis (a target construction *over and over* and a co-gesture I.Cyclic-Sagittal)

	+G I.Cyclic-Sagittal	-G other gesture lemmas
+C <i>over and over</i>	1. (3a) +C+G = 55	3. (3c) +C-G = 93
-C Reference data	2. (3b) -C+G = 34	4. (3d) -C-G = 526

frequency with which the target construction co-occurs with the target gesture but also on how *salient* and prototypical the gesture is for the construction, assuming that constructions have prototype structures (Cienki, 2017). Building on this argument, the present study employed crossmodal collostructional analysis to measure the salience of a gesture for the target constructions. In doing so, only gestures that satisfied the frequency criterion, albeit tentatively – those that occurred in more than 10% of the instances of each target construction – were considered for the Fisher Exact test, which yields the collostruction strength.

Below, Tables 8 and 9 summarize the results of the crossmodal collostructional analyses for the RACs and the OACs, respectively. Before delving into these results, however, we need to consider which value should be taken as a reliable measurement of salience for this study. In traditional unimodal collostructional analysis (Stefanowitsch & Gries, 2003), the Fisher Exact *p*-value is used as a collostruction strength, a value that indicates the associational strength between a construction and a collexeme. However, as Uhrig (2022, p. 109) points out, such an approach is also problematic for the current study because, as a test of significance, the Fisher exact test exhibits a bias in favor of higher frequency. Consequently, in Table 8, the *p*-values of the gesture lemma I.Cyclic-Sagittal decrease (which suggests that the associational strengths increase) in inverse proportion to the absolute co-frequency

Table 8. Results of the crossmodal collostructional analysis for the RACs

<i>Over and over</i>				
Gesture lemma	Frequency with the target construction (N = 148)	Frequency in the reference data (N = 560)	<i>p</i> -value Fisher exact	Odds ratio (95% CI)
I.Cyclic-Sagittal	55 (37.1%)	34 (6.1%)	7.71E-20	9.10 (5.50, 15.27)
Beat	40 (27.0%)	130 (23.2%)	0.332	1.22 (0.78, 1.87)
<i>Again and again</i>				
Gesture lemma	Frequency with the target construction (N = 106)	Frequency in the reference data (N = 560)	<i>p</i> -value Fisher Exact	Odds ratio (95% CI)
I.Cyclic-Sagittal	37 (34.9%)	34 (6.1%)	2.56E-14	8.25 (4.70, 14.54)
Beat	33 (31.1%)	130 (23.2%)	0.085	1.49 (0.91, 2.40)
<i>On and on</i>				
Gesture lemma	Frequency with the target construction (N = 40)	Frequency in the reference data (N = 560)	<i>p</i> -value Fisher Exact	Odds ratio (95% CI)
I.Cyclic-Sagittal	15 (37.5%)	34 (6.1%)	5.05E-8	9.21 (4.11, 20.20)
Beat	8 (20.0%)	130 (23.2%)	0.845	0.82 (0.32, 1.89)

Table 9. Results of the crossmodal collostructional analysis for the OACs²²

<i>Back and forth</i>				
Gesture lemma	Frequency with the target construction (N = 106)	Frequency in the reference data (N = 560)	p-value Fisher Exact	Odds ratio (95% CI)
I.Bidirection-Horizontal	29 (27.3%)	11 (2.0%)	1.12E-16	18.64 (8.62, 43.13)
I.2H-Alternation-Sagittal	26 (24.5%)	11 (2.0%)	2.10E-14	16.10 (7.35, 37.58)
Beat	3 (2.8%)	130 (23.2%)	4.87E-8	0.09 (0.01, 0.29)
<i>Up and down</i>				
Gesture lemma	Frequency with the target construction (N = 83)	Frequency in the reference data (N = 560)	p-value Fisher Exact	Odds ratio (95% CI)
I.Bidirection-Vertical	31 (36.14%)	5 (0.9%)	2.20E-25	65.14 (23.77, 225.25)
I.Small-Bidirection-Vertical	14 (16.86%)	2 (0.4%)	1.24E-11	55.87 (12.44, 516.83)
Beat	6 (7.2%)	130 (23.2%)	4.60E-4	0.25 (0.08, 0.60)
<i>In and out</i>				
Gesture lemma	Frequency with the target construction (N = 49)	Frequency in the reference data (N = 560)	p-value Fisher Exact	Odds ratio (95% CI)
I.Bidirection-Horizontal	11 (22.4%)	11 (2.0%)	1.08E-7	14.29 (5.25, 39.03)
I.2H-Alternation-Sagittal	10 (20.4%)	11 (2.0%)	8.22E-7	12.67 (4.52, 35.16)
I.1H-Small-To-Fro	5 (10.2%)	2 (0.4%)	5.14E-5	31.25 (4.94, 334.29)
Beat	3 (3.3%)	130 (23.2%)	3.51E-3	0.21 (0.04, 0.68)

of each variant of the RACs with the gesture lemma, despite the nearly identical relative frequencies of the lemma across the three variants. For this very reason, the current study refers primarily to *odds ratio*, which is a measure of effect size and thus not dependent on the sample size of the data. As explained clearly by Schmid and Küchenhoff (2013, pp. 553–554), odds ratio scores are themselves easily interpretable in such a way that the odds ratio of 9.10 given for I.Cyclic-Sagittal co-occurring with *over and over* in Table 8 indicates that a construction which has the feature +TARGET CONSTRUCTION (= *over and over*) is 9.10 times more likely to also have the feature +TARGET GESTURE (= I.Cyclic-Sagittal) than a construction not having the feature +TARGET CONSTRUCTION (i.e., a construction in the reference dataset). With this focus on odds ratio as a reference measure of the associational strength between a construction and a gesture, the column for odds ratio in boldface in Tables 8 and 9.

As seen in Table 8, the three variants of the RACs exhibit very similar tendencies, with the gesture lemma I.Cyclic-Sagittal dominating over other gesture lemmas. The odds ratios for I.Cyclic-Sagittal in the three variants are

²²Table 9 includes the Beat gesture even though its frequencies are below 10%, because it exhibits significant repulsion to each variant of the OACs.

around 8 to 9, indicating that they are all highly attracted to the gesture lemma, despite the fact that it was also frequently observed in the reference dataset. Additionally, for each variant, the Beat gesture also accounts for a large proportion, ranging from 20% to 31%. When it comes to the odds ratio, however, the contrast becomes prominent between the gesture lemmas Beat and I.Cyclic-Sagittal. For each variant, the association with Beat is not significant, and especially for the construction *on and on* the odds ratio of 0.82 even suggests that it is slightly repelled to Beat; on the other hand, the scores of 1.22 and 1.49 for the other two variants indicate slight attractions. Obviously, this is because the gesture lemma Beat was the most frequent one in the reference dataset, amounting to 23% of all gesture lemmas (Table 6).

Regarding the OACs, as can be predicted by Table 5, gestures with bidirectional movements show significant attractions. As for *back and forth*, the gesture lemmas I.Bidirection-Horizontal (Figure 5) and I.2H-Alternation-Sagittal (Figure 6) account for a large proportion of the accompanying gestures, yielding high odds ratio scores. The same applies to *in and out*; however, in addition to the two lemmas, another lemma, I.1H-Small-To-Fro (Figure 9), features an even higher odds ratio since it was rarely found in the reference dataset. Concerning *up and down*, two bidirectional gestures on the vertical axis, I.Bidirection-Vertical (Figure 7) and I.Small-Bidirection-Vertical (Figure 8), both obtain odds ratios of over 55, reflecting the fact that these two lemmas were nearly exclusively used for *up and down* and not in the reference dataset. Finally, unlike the RACs, each variant of the OACs exhibits significant repulsion to the Beat gesture, which also typically involves bidirectional movement but lacks the referential function. This may be because, as a gesture with bidirectional movement, the Beat gesture was overshadowed by other bidirectional gestures with referential meanings which were in perfect alignment with the meanings of the OACs.

4. Discussion

With the obtained results, we are now back to the research question posed in Section 1: If there are any specific gestures that are frequently employed with the [ADV and ADV] constructions, do they form multimodal constructions, or are they more compositional crossmodal collocations (Uhrig, 2021)? In Section 4.1, in light of the proposed criteria for identifying multimodal constructions, namely, (i) frequency and (ii) salience (Hoffmann, 2017; Schoonjans, 2017), we discuss the ontological status of the multimodal patterns reviewed in the preceding section. Furthermore, Section 4.2 presents an account of the embodied motivations for multimodal constructions.

4.1. Multimodal constructions versus crossmodal collocations

First, we examine the fact that the gesture lemmas I.Cyclic-Sagittal and Beat were frequently observed in all three variants of the RACs, each making up 20% to 37% of the data. It is not surprising that I.Cyclic-Sagittal accounted for such a large proportion, as it refers to iteration (Ruth-Hirrel, 2018) or ongoing events (Ladewig, 2011), which are perfectly compatible with the aspectual meanings of the RACs. But what about Beat? Are the frequent co-occurrences with Beat an accidental feature of the RACs rooted in the specific register of this study, namely TED Talks

since the Beat gesture is essentially of the free combination type (such as air quotes; Uhrig, 2021) without any semantic values of its own? This possibility would be refuted by Lelandais' (2024) study, which also examines the co-speech gestures of *over and over*. Using an archive of American television news programs, the study found that cyclic gestures and beat gestures were frequently produced in combination with *over and over*, accounting for 42% and 38% of hand gestures co-occurring with the construction, respectively (Lelandais, 2024, pp. 21–24). This finding, albeit slightly higher in proportion, is in complete agreement with the results of the present study, making it more convincing that the frequent co-occurrence with I.Cyclic-Sagittal and Beat is a general feature of the RACs, or at least of *over and over*.²³ Therefore, it can be argued that the combinations of each variant of the RACs with I.Cyclic-Sagittal or Beat meet the first criterion for multimodal constructions, namely, that of frequency.²⁴ As we saw in Table 8, however, the two gesture lemmas differ with respect to their odds ratios. For each variant of the RACs, I.Cyclic-Sagittal features a high odds ratio; hence, in light of the second criterion of salience, it is highly likely that the combination of each variant of the RACs with I.Cyclic-Sagittal ceases to be merely compositional and instead becomes stored as an integrated whole, that is, a multimodal construction. By contrast, the much lower odds ratios for Beat in all three variants, around 0.8 to 1.5, suggest that the combination of the RACs and Beat would remain compositional, positioned somewhere along a continuum between free combinations and cross-modal collocations, especially given that the associations are not significant.

In the same vein as the combination of the RACs and I.Cyclic-Sagittal, it is highly likely that each variant of the OACs and its corresponding bidirectional gestures jointly form multimodal constructions. In particular for *up and down*, bidirectional gestures on the vertical axis (I.Bidirection-Vertical and I.Small-Bidirection-Vertical) yielded considerably high odds ratios, indicating a high degree of multimodal constructionhood. As for *back and forth* and *in and out*, the gesture lemmas I.Bidirection-Horizontal and I.2H-Alternation-Sagittal may well be part of multimodal constructions in terms of frequency and salience. However, the status of I.1H-Small-To-Fro for *in and out* remains subject to further debate. On the one hand, I.Bidirection-Horizontal and I.2H-Alternation-Sagittal were also observed in the reference dataset with substantial frequency, suggesting that they can arguably be regarded as recurrent gestures. On the other hand, I.1H-Small-To-Fro was rarely found in the reference dataset and thus is more likely a so-called singular gesture (Müller, 2017). This resulted in its much higher odds ratio in *in and out* than the two gesture lemmas in question, despite the lower frequency of co-occurrence with the construction. This apparently contradictory result seems to have stemmed from a methodological issue inherent in the statistical approach chosen for this study. However, it also raises a theoretical question about which condition – frequency or salience – has a greater impact on the multimodal constructionhood, necessitating further empirical investigation.

²³Still, Beat gestures may occur less frequently in spontaneous private conversations, as news programs and TED Talks might not entirely be distinct genres.

²⁴Apart from the threshold value of 10% set tentatively for conducting crossmodal collocation analysis in Tables 8 and 9, the present study does not further specify a threshold value for what can count as *sufficient* frequency: instead, it assumes a continuum where constructionhood is proportional to frequency.

While the argument presented so far in this section aligns more closely with the approach emphasizing lower-level representations of constructions (the exemplar-based model; e.g., Bybee, 2013), the results can also be interpreted as supporting the existence of higher-level multimodal constructional schemas, such as [the RAC (ADV_1 and ADV_1) + the cyclic gesture family] and [the OAC (ADV_1 and ADV_2) + the bidirectional gesture family]. For the RACs overall, the cyclic gestures (those highlighted in red in Tables 4–6) accounted for 44.6% of all gestures co-occurring with the RACs (see Table 4), yielding an odds ratio of 9.39 (95% CI: 6.32, 14.16, $p < 0.001$); for the OACs overall, the bidirectional gestures (those highlighted in blue in Tables 4–6) accounted for 73.1% of all gestures accompanying the OACs (see Table 5), resulting in an odds ratio of 43.02 (95% CI: 26.98, 70.35, $p < 0.001$). From the perspective that assumes redundant representations of constructions (Hilpert, 2019, pp. 67–68), both of these multimodal constructional schemas and the lower-level multimodal constructions can coexist in the construct-i-con. The advantage of assuming the higher-level multimodal constructional schemas, however, is that they can license *atypical* combinations of the RACs/OACs and the cyclic/bidirectional gestures. Those atypical combinations cannot be explained by the lower-level multimodal constructions discussed above, as they represent the (proto-)*typical* combinations of the target constructions and the gesture families in question. For example, in addition to I.Bidirection-Horizontal and I.2H-Alternation-Sagittal, the constructions *back and forth* and *in and out* were accompanied by I. (Small) - Bidirection-Vertical, in four and three instances, respectively. The multimodal constructional schema [the OAC (ADV_1 and ADV_2) + the bidirectional gesture family] is likely to have operated in the production of these atypical multimodal constructs, based on the speaker's online conceptualizations of the described bidirectional events (see also the next section).

4.2. *Exbodiment: the experiential basis for multimodal constructions*

As we have seen thus far, Uhrig's (2021) distinction between multimodal constructions and crossmodal constructions is based on the strength of association between a linguistic construction and a gesture, as well as the lack of compositionality resulting from the association. By contrast, Hoffmann's (2021) definition of multimodal constructions, as provided in Section 1, emphasizes the jointness of meaning or co-expressiveness of the linguistic and gestural parts, which seems to be the case given the potential multimodal constructions posited in the last section. This nature of multimodal constructions can be best explained by drawing on Mittelberg's (2013) notion of *exbodiment*, according to which gestures are thought to be 'bodily semiotic acts through which embodied image and action schemas may, to some degree at least, be externalized, made visible, and used for meaningful communication' (Mittelberg, 2008, p. 144). For the target constructions of the present study, the *image schemas* in which their meanings are conventionally grounded are particularly relevant here, with an image schema defined as 'a recurring, dynamic pattern of our perceptual interactions and motor programs that gives coherence and structure to our experience' (Johnson, 1987, p. xiv).

Regarding the RACs, their meanings are grounded in the ITERATION image schema, which can also be inferred from their forms that include lexical iterations. The ITERATION schema, in turn, is inherently connected with the CYCLE image schema (Ruth-Hirrel, 2018) because we experience iteration through cycles, such

as hands going around a clock and the sun moving across the sky, with the primary metaphor TIME IS SPACE operating in this conception (Fauconnier & Turner, 2008). As such, the image schemas ITERATION and CYCLE serve as embodied motivations for the RACs, and they become externalized, or *exbodied* (Mittelberg, 2013) into cyclic gestures.

Similarly, the OACs can also be seen as grounded in the ITERATION schema given the aspectual dimension of their meanings, but they are certainly more strongly grounded in the spatial relation image schemas (Lakoff & Johnson, 1999) such as FRONT-BACK, UP-DOWN, LEFT-RIGHT and so on. For *up and down*, it is obvious that the UP-DOWN image schema plays a crucial role in both the meaning of the construction and the manifestation of bidirectional gestures on the vertical axis. A more interesting fact, however, is that the constructions *back and forth* and *in and out* were accompanied by bidirectional gestures that were realized on any axis. As mentioned at the end of the previous section, in addition to I. *Bidirection-Horizontal* and I. *2H-Alternation-Sagittal*, these constructions also co-occurred with I. (Small)-*Bidirection-Vertical* several times. This indicates that they would arguably be grounded in a more general spatial image schema, something like BIDIRECTIONALITY, with a more specific image schema (such as FRONT-BACK, LEFT-RIGHT, etc.) activated and enacted in each usage event to produce a bidirectional gesture on a specific axis. This specification would plausibly depend on the speaker's construal, including perspective-taking, fictive motion, etc., which is also likely influenced by various contextual factors. Given this argument, the reason why the constructions *back and forth* and *in and out* were accompanied less by bidirectional gestures on the vertical axis but more by those on the horizontal and sagittal axes, may be that there was an alternative that prevailed, namely, the multimodal construction '*up and down* + I. *Bidirection-Vertical*/I. *Small-Bidirection-Vertical*', which is conventionally grounded in the more specified UP-DOWN image schema.

On a final note, there is another noteworthy observation concerning *in and out*. While its preference for co-speech gestures was quite similar to *back and forth*, its meaning, or more specifically, the image schema on which it bases its meaning, is different; the meaning of *in and out* inherently involves CONTAINMENT. Nevertheless, the fact that gestures accompanying *in and out* had few elements that could evoke the CONTAINMENT image schema, and instead profiled its BIDIRECTIONAL aspect more, suggests that there may be some kind of pattern or constraint on which aspect of an image schema can be materialized into a gesture.

5. Conclusion

Using a small multimodal corpus constructed by manual annotation (Section 2.7), the present study shows that the [ADV and ADV] constructions are significantly attracted to certain kinds of gesture lemmas, forming multimodal constructions. Each variant of the RACs, on the one hand, co-occurred quite frequently with I. *Cyclic-Sagittal* and Beat. However, subsequent crossmodal collostructional analysis revealed that in combination with the RACs, the former is highly likely to be stored as multimodal constructions, whereas the latter to be positioned somewhere between the free combination and the crossmodal collostruction zone. On the other hand, each variant of the OACs exhibited significant attractions to its

corresponding bidirectional gestures, and therefore they too are plausibly represented as multimodal constructions in a speaker's construct-i-con.

In Section 4.2, the embodied motivations for the potential multimodal constructions were also discussed, drawing on Mittelberg's (2008, 2013) notion of embodied. Building on that notion, it is also possible to predict in an *a priori* manner what kinds of gestures may co-occur with a target construction, by considering in advance the embodied image or action schemas in which the meaning of the construction is grounded. Thus, what Tables 4 and 5 showed us – namely, that the cyclic gestures and bidirectional gestures accounted for 44% and 73% of the gestures accompanying the RACs and the OACs, respectively – is not surprising, although the kinesic information of the gestures we can predict from the embodied schemes is quite schematic in nature. However, as has been shown in the present study, the realization of those gestures is patterned in certain ways. In other words, they have some specific kinesic features that are conventionalized (e.g., I.2H-Alternation-Sagittal, I.Cyclic-Sagittal), which in turn makes it plausible to assume that multimodal constructions have prototype structures (Cienki, 2017), such that some gestural configurations are more prototypical while others are more peripheral (compare 'back and forth + I.2H-Alternation-Sagittal' with 'back and forth + I.Bidirection-Vertical'). Therefore, one of the tasks of future research in multimodal CxG would be to identify such conventionalized features of gestures that externalize the embodied schemas of a target construction, which, in principle, have many possible ways of manifestation.

Acknowledgments. I would like to express my gratitude to Toshio Ohori for his supervision of the original thesis on which this paper is based. I am also grateful to Yoichiro Hasebe for creating and maintaining TCSE. Furthermore, this paper has greatly benefited from the comments from Beate Hampe, Hikaru Hotta, Ippai Inoue, Jordan Zlatev, Kyoko Ohara, Manon Lelandais, Masato Yoshikawa, Minoru Ohtsuki, and Yoshiki Nishimura. Lastly, I extend my appreciation to the two anonymous reviewers and the editor for their valuable feedback on the manuscript.

Appendices and data availability statement. Appendices A (a complete list of gesture lemmas) and B (coding criteria for distinguishing similar lemmas), R scripts for the Fisher exact tests, and spreadsheets summarizing the results of gesture annotations and the two intercoder reliability experiments are available on OSF (<https://osf.io/8t2pj/>).

Competing interest. The author declares none.

References

- Andrén, M. (2014). Multimodal constructions in children: Is the headshake part of language? *Gesture*, 14(2), 141–170. <https://doi.org/10.1075/gest.14.2.02and>.
- Bressem, J., & Müller, C. (2014). The family of Away gestures: Negation, refusal, and negative assessment. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & J. Bressem (Eds.), *Body, language, communication: An international handbook on multimodality in human interaction* (38.2) (pp. 1592–1604). Berlin: De Gruyter Mouton.
- Bybee, J. (2013). Usage-based theory and exemplar representations of constructions. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar* (pp. 49–69). New York: Oxford University Press.
- Carletta, J. (1996). Assessing agreement on classification tasks: The Kappa statistic. *Computational Linguistics*, 22(2), 249–254.
- Cienki, A. (2017). Utterance construction grammar (UCxG) and the variable multimodality of constructions. *Linguistics Vanguard*, 3(s1), 1–10. <https://doi.org/10.1515/lingvan-2016-0048>.

- Efron, D. (1941). *Gesture and environment*. New York: King's Crown Press.
- Ekman, P., & Friesen, W. (1969). The repertoire of non-verbal behavior: Categories, origins, usage, and coding. *Semiotica*, 1(1), 49–98.
- ELAN (Version 6.4) [Computer software]. (2022). Nijmegen: Max Planck Institute for Psycholinguistics, *The Language Archive*. Retrieved from <https://archive.mpi.nl/tla/elan>.
- Fauconnier, G., & Turner, M. (2008). Rethinking metaphor. In R. W. Gibbs (Ed.), *The Cambridge Handbook of Metaphor and Thought* (pp. 53–66). Cambridge: Cambridge University Press.
- Fricke, E., Bressemer, J., & Müller, C. (2014). Gesture families and gestural fields. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & J. Bressemer (Eds.), *Body, language, communication: An international handbook on multimodality in human interaction* (38.2) (pp. 1630–1640). Berlin: De Gruyter Mouton.
- Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: The University of Chicago Press.
- Harrison, S. (2021). Showing as sense-making in oral presentations: The speech-gesture-slide interplay in TED talks by Professor Brian Cox. *Journal of English for Academic Purposes*, 53, 1–19. <https://doi.org/10.1016/j.jeap.2021.101002>.
- Hasebe, Y. (2015). Design and implementation of an online corpus of presentation transcripts of TED Talks. *Procedia - Social and Behavioral Sciences*, 198(24), 174–182. <https://doi.org/10.1016/j.sbspro.2015.07.434>.
- Hasebe, Y. (2018). TED Corpus Search Engine: TED Talks wo kenkyu to kyoiku ni katsuyosuru tame no platform [TED Corpus Search Engine: The platform for employing TED Talks for research and education]. *Eigo Corpus Kenkyu [Research in English corpora]*, 25, 159–171.
- Hilpert, M. (2019). *Construction grammar and its application to English* (2nd ed.). Edinburgh: Edinburgh University Press.
- Hinnell, J. (2018). The multimodal marking of aspect: The case of five periphrastic auxiliary constructions in North American English. *Cognitive Linguistics*, 29(4), 773–806. <https://doi.org/10.1515/cog-2017-0009>.
- Hoffmann, T. (2017). Multimodal constructs – multimodal constructions? the role of constructions in the working memory. *Linguistics Vanguard*, 3(s1), 1–10. <https://doi.org/10.1515/lingvan-2016-0042>.
- Hoffmann, T. (2021). Multimodal Construction Grammar: From Multimodal Constructs to Multimodal Constructions. In X. Wen & J. R. Taylor (Eds.), *The Routledge Handbook of Cognitive Linguistics* (pp. 78–92). New York: Routledge.
- Johnson, M. (1987). *The body in the mind: The bodily basis of reason and imagination*. Chicago: Chicago University Press.
- Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.
- Kipp, M., Neff, M., & Albrecht, I. (2007). An annotation scheme for conversational gestures: How to economically capture timing and form. *Language Resources and Evaluation*, 41(3–4), 325–339. <https://doi.org/10.1007/s10579-007-9053-5>.
- Kipp, M. (2004). *Gesture generation by imitation: From human behavior to computer character animation*. Florida: Dissertation.com.
- Kita, S., van Gijn, I., & van der Hulst, H. (1998). Movement phases in signs and co-speech gestures, and their transcription by human coders. In I. Wachsmuth & M. Fröhlich (Eds.), *Gesture and sign language in human computer interaction: International gesture workshop, Bielefeld, Germany, Sept. 17–19, 1997; Proc* (pp. 23–35). Berlin: Springer.
- Ladewig, S. (2011). Putting the cyclic gesture on a cognitive basis. *CogniTextes*, 6, 1–26. <https://doi.org/10.4000/cognitextes.406>.
- Ladewig, S. (2024). Recurrent gestures: Cultural, individual, and linguistic dimensions of meaning-making. In A. Cienki (Ed.), *The Cambridge Handbook of Gesture Studies* (pp. 32–55). Cambridge: Cambridge University Press.
- Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to Western thought*. New York: Basic Books.
- Langacker, R. (2008). *Cognitive Grammar: A Basic Introduction*. Oxford: Oxford University Press.
- Lelandais, M. (2024). Multimodal marks of iteration in discourse. *Faits de Langues*, 53(2), 41–68. <https://doi.org/10.1163/19589514-53020003>.
- Masi, S. (2020). Exploring meaning-making practices via co-speech gestures in TED Talks. *Journal of Visual Literacy*, 39(3–4), 201–219. <https://doi.org/10.1080/1051144X.2020.1826223>.
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.

- McNeill, D. (2005). *Gesture and thought*. Chicago: University of Chicago Press.
- Mittelberg, I. (2008). Peircean semiotics meets conceptual metaphor: Iconic modes in gestural representations of grammar. In A. Cienki & C. Müller (Eds.), *Metaphor and gesture* (pp. 115–154). Amsterdam & Philadelphia: John Benjamins.
- Mittelberg, I. (2013). The embodied mind: Cognitive-semiotic principles as motivating forces in gesture. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & S. Tessendorf (Eds.), *Body, language, communication: An international handbook on multimodality in human interaction (38.1)* (pp. 750–779). Berlin: Mouton de Gruyter.
- Mittelberg, I. (2017). Multimodal existential constructions in German: Manual actions of giving as experiential substrate for grammatical and gestural patterns. *Linguistics Vanguard*, 3(s1), 1–14. <https://doi.org/10.1515/lingvan-2016-0047>.
- Müller, C. (2004). Forms and uses of the Palm Up Open Hand. A case of gesture family? In C. Müller (Ed.), *Semantics and pragmatics of everyday gestures* (pp. 234–256). Berlin: Weidler.
- Müller, C. (2014). Gesture as “deliberate expressive movement.” In M. Seyfeddinipur & M. Gullberg (Eds.), *From gesture in conversation to visible action as utterance*, (pp. 127–152). Amsterdam:
- Müller, C. (2017). How recurrent gestures mean: conventionalized contexts-of-use and embodied motivation. *Gesture*, 16, 278–306. <https://doi.org/10.1075/gest.16.2.05mul>.
- Ningelgen, J., & Auer, P. (2017). Is there a multimodal construction based on non-deictic so in German? *Linguistics Vanguard*, 3(s1), 1–15. <https://doi.org/10.1515/lingvan-2016-0051>.
- Rice, S. (1999). Aspects of prepositions and prepositional aspect. In L. de Stadler & C. Eyrych (Eds.), *Issues in Cognitive Linguistics: 1993 ICLC proceedings* (pp. 225–247). Berlin/New York: Mouton de Gruyter.
- Ruth-Hirrel, L. (2018). *Cyclic Gestures and Multimodal Symbolic Assemblies: an Argument for Symbolic Complexity in Gesture*. [Doctoral thesis]. Albuquerque, NM: University of New Mexico.
- Schegloff, E. A. (1984). On some gestures’ relation to talk. In J. M. Atkinson & J. Heritage (Eds.), *Structures of Social Action: Studies in conversation analysis* (pp. 266–298). Cambridge: Cambridge University Press.
- Schmid, H., & Küchenhoff, H. (2013). Collostructional analysis and other ways of measuring lexicogrammatical attraction: Theoretical premises, practical problems and cognitive underpinnings. *Cognitive Linguistics*, 24(3), 531–577. <https://doi.org/10.1515/cog-2013-0018>.
- Schoonjans, S. (2017). Multimodal construction grammar issues are construction grammar issues. *Linguistics Vanguard*, 3(s1), 1–8. <https://doi.org/10.1515/lingvan-2016-0050>.
- Stefanowitsch, A., & Gries, S. T. (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, 8(2), 209–243. <https://doi.org/10.1075/ijcl.8.2.03ste>.
- Talmy, L. (2000). *Toward a cognitive semantics* (Vol. II). Cambridge: MIT Press.
- Uhrig, P. (2021). *Large-scale multimodal corpus linguistics: The big data turn* [Habilitation thesis]. Erlangen: Friedrich-Alexander-Universität Erlangen-Nürnberg.
- Uhrig, P. (2022). Hand gestures with verbs of throwing: Collostructions, style, and metaphor. *Yearbook of the German Cognitive Linguistic Association*, 10(1), 99–120. <https://doi.org/10.1515/gcla-2022-0006>.
- Ungerer, T. (2023). A gradient notion of constructionhood. *Constructions*, 15(1), 1–20. <https://doi.org/10.24338/cons-543>.
- Ziem, A. (2017). Do we really need a multimodal construction grammar? *Linguistics Vanguard*, 3(s1), 1–9. <https://doi.org/10.1515/lingvan-2016-0095>.
- Zima, E., & Bergs, A. (2017). Multimodality and construction grammar. *Linguistics Vanguard*, 3(s1), 1–9. <https://doi.org/10.1515/lingvan-2016-1006>.
- Zima, E. (2017). On the multimodality of [all the way from X PREP Y]. *Linguistics Vanguard*, 3(s1), 1–12. <https://doi.org/10.1515/lingvan-2016-0055>.

Cite this article: Kuryu, D. (2025). Crossmodal collostructional analysis of English [ADV and ADV] constructions: multimodal constructions or crossmodal collostructions? *Language and Cognition*, 17, e39, 1–28. <https://doi.org/10.1017/langcog.2025.8>