Journal of Memory and Language 104 (2019) 56–69

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Information packaging in speech shapes information packaging in gesture: The role of speech planning units in the coordination of speech-gesture production

Isabella Fritz[a,*], Sotaro Kita[b], Jeannette Littlemore[a], Andrea Krott[c]

[a] Department of English Language and Applied Linguistics, University of Birmingham, Birmingham B15 2TT, UK
[b] Department of Psychology, University of Warwick, CV4 7AL Coventry, UK
[c] School of Psychology, University of Birmingham, B15 2TT Birmingham, UK

ARTICLE INFO

Keywords:
Gesture
Motion events
Syntax
Planning units
Event conceptualization

ABSTRACT

Linguistic encoding influences the gestural manner and path depiction of motion events. Gestures depict manner and path of motion events differently across languages, either conflating or separating manner and path, depending on whether manner and path are linguistically encoded within one clause (e.g., "rolling down") or multiple clauses (e.g., "descends as it rolls") respectively. However, it is unclear whether such gestural differences are affected by how speech packages information into planning units or by the way information is lexicalised (as verb plus particle or as two verbs). In two experiments, we manipulated the linguistic encoding of motion events in either one or two planning units while lexicalisation patterns were kept constant (i.e., verb plus particle). It was found that separating manner (verb) and path (particle) into different planning units also increased gestural manner and path separation. Thus, lexicalisation patterns do not drive gestural depiction of motion events. Rather gestures are shaped online by how speakers package information into planning units in speech production.

Introduction

When we speak, we often spontaneously produce gestures. Gestures are tightly linked to how we encode information linguistically at the temporal and semantic levels (McNeill, 1992, 2005). In terms of synchronisation, gestures co-occur with the element in speech that is closest to the gesture's content, often initiated before their semantic affiliates (McNeill, 1992; Morrel-Samuels & Krauss, 1992; Schegloff, 1984). From a semantic perspective speech-gesture coordination is reflected by linguistic choices. This coordination is evident on both a lexical and structural level. For instance, gestures have been found to adapt to fine-grained differences in verb semantics (Gullberg, 2011; Gullberg & Narasimhan, 2010; Kita & Özyürek, 2003), and it has been suggested that the content of a gesture is linked to how clauses package information in speech (Kita, 2000; Kita & Özyürek, 2003).

Evidence for the influence of linguistic packaging on gestural content stems from cross-linguistic studies on motion event gestures (Kita & Özyürek, 2003; Wessel-Tolvig & Paggio, 2016; Özyürek et al., 2008; Özyürek, Kita, Allen, Furman, & Brown, 2005; Özçalişkan, 2016). Motion events are linguistically encoded differently across languages. According to Talmy's typology (2000), languages generally fall into two different categories. In so-called satellite-framed languages (e.g., German and English) the manner component of a motion event is usually encoded within the verb, while the path component is encoded in a "satellite" (a particle or an affix). Both components together often form a so-called particle verb (e.g., "to roll down" or "to climb up"). In verb-framed languages (e.g., Spanish and Japanese), path is encoded in the verb, while manner, if encoded in speech at all, is encoded in a phrase, gerund or in a separate clause (e.g., "he descends the hill as he rolls").

Gestures that accompany descriptions of motion depict manner and path in different ways in satellite-framed and verb-framed languages. In satellite-framed languages, where the motion event is linguistically encoded within one clause, manner and path tend to be conflated in a single gesture. In contrast, in verb-framed languages, where manner and path are encoded in two separate clauses, speakers tend to separate manner and path gesturally (Kita & Özyürek, 2003; Kita et al., 2007; Wessel-Tolvig & Paggio, 2016; Özyürek et al., 2008; Özyürek et al.,

* Corresponding author at: Language Acquisition and Language Processing Lab, Department of Language and Literature, Norwegian University of Science and Technology, Trondheim, Norway.

E-mail address: isabella.fritz@ntnu.no (I. Fritz).

https://doi.org/10.1016/j.jml.2018.09.002
Received 21 October 2017; Received in revised form 11 September 2018
0749-596X/

2005; Özçalişkan, 2016; Özçalışkan, Lucero, & Goldin-Meadow, 2016). Thus, linguistic encoding in speech influences gestural content; but what is the relevant linguistic level?

These cross-linguistically varying gestural patterns may stem from differences in information packaging during speech production planning (Kita & Özyürek, 2003). We call this view the Planning Unit Account. More specifically, when manner and path are linguistically encoded in a single planning unit for speech production, a single gesture expresses both manner and path. When manner and path are linguistically encoded in two planning units, two gestures express manner and path separately. Clauses are assumed to play a vital role in the coordination of gesture and speech because from a speech production perspective they have been assumed to constitute a good proxy for planning units (Bock, 1982; Levelt, 1989). Thus, clausal packaging of manner and path (one clause in satellite-framed languages, two clauses in verb-framed languages) is related to gestural packaging of manner and path. However, Kita & Özyürek (2003, p. 17) do not claim that gesture-speech coordination is bound to a clausal scope. Rather they argue that gesture-speech coordination is linked to a more general processing unit[1] defined as a unit that "corresponds to what can be processed within one processing cycle for the formulation of speech".

Alternatively, the cross-linguistically varying gestural patterns might stem from differences in motion event conceptualisation based on lexicalisation of motion concepts and its implication for clausal structure. We call this view the Lexicalisation Account. This hypothesis is based on the argument that clauses are not just important units for speech production but also "conceptual units" (Pawley, 1987, 2010). This view can also account for Kita and Özyürek's (2003) cross-linguistic finding. According to this account, satellite-framed languages, which encode manner and path in a single clause, represent manner and path within a single conceptual unit. In contrast, verb-framed languages represent manner and path in two separate conceptual units. What is expressed in a single conceptual unit is expressed as a single gesture. The influence of lexicalisation patterns on motion event conceptualisation is also in line with Slobin's (2003, 2006) thinking-for-speaking hypothesis, which states that during speech production we have to filter our thoughts through linguistic encoding possibilities. Hence, lexicalisation patterns guide the speaker's attention to different aspects of the motion event (Slobin, 2000). Slobin (2000, 2003) argues that due to the obligatory encoding of manner in satellite-framed languages (within the main verb) in combination with the path component which is governed by the verb (i.e., particle verbs), speakers of these languages tend to perceive motion events as a "single conceptual event". Since it is not obligatory to encode manner linguistically in verb-framed languages, speakers of these languages do not perceive manner as inherent to the motion event. Manner is rather perceived as an activity that accompanies the path element of the motion event which is encoded in the main verb (e.g., exit, enter). These differences in conceptualisation would also explain the prominence of manner and path conflated gestures in satellite-framed languages and manner and path separated gestures in verb-framed languages.

Importantly, Kita et al. (2007) found that linguistic encoding only has an online effect on motion event conceptualisation and that gestural content is not bound to a habitual way of thinking based on how a language predominantly encodes motion events (satellite-framed versus verb-framed construction). In their study, Kita et al. (2007) compared gestures accompanying the two types of constructions within English: a satellite-framed construction (one verb framing, e.g., "he rolled down the hill") or a verb-framed construction (two verb framing, e.g., "he went down as he rolled"). They hypothesised that if habitual (dominant) language-specific event conceptualisation shapes motion event gestures, conflated manner and path gestures would be expected regardless of the construction type. However, in Kita et al.'s study (2007) the participants' gestures differed between satellite-framed and verb-framed constructions. When participants used satellite-framed constructions they accompanied speech with the expected conflated manner and path gesture. But when participants used verb-framed constructions manner and path were not only linguistically but also gesturally separated. Furthermore, essentially the same effect of construction types on gesture was also found in Dutch, another satellite-framed language (Mol & Kita, 2012). Hence, these studies suggest that gestures are shaped during speech production based on online linguistic choices and not on habitual language-specific event conceptualisations. Thus, conceptual events, which the Lexicalization Account associates with gestural information packaging, must be generated online at the moment of speaking.

One shortcoming of studies on motion events so far is that they have not been able to establish whether differences in gestural patterns stem from differences in how speech is packaged into planning units or differences in lexicalisation patterns (what information is encoded in a clause). Hence, different accounts could explain the gestural differences between verb-framed and satellite-framed constructions. Thus, the main aim of the present study is to provide unambiguous evidence for the Planning Unit Account. To this end, we manipulated the linguistic distance between manner and path components while keeping the lexicalisation pattern constant.

In Experiment 1 we tested whether increasing the linguistic distance between manner and path elements within the same clause of a satellite-framed construction can break up the planning unit into a manner and a path component, and consequently separate manner and path into two different gestures. Crucially, we asked German speakers to insert a sub-clause ("as seen in the video") between manner and path elements, which should make speakers plan manner and path in two different planning units. This assumption is based on previous research suggesting that clauses are a good proxy for planning units (Bock, 1982; Levelt, 1989) and thus inserting a sub-clause (i.e., another planning unit) between manner and path should lead the participants to process manner and path in separated planning units. Moreover, Wagner, Jescheniak, & Schriefers (2010) found evidence that the scope of planning changes when processing load increases. For the sentence structure in our experiment, we assume that the insertion of the sub-clause within a main clause will increase processing load which then results in the production of manner verb and path particle in different planning units.

It is possible in German to insert extra linguistic elements between manner and path elements. German particle verbs (e.g. "hinunterrollen" – "to down-roll") can be linguistically combined into one word or split up into two (potentially) distant words, depending, among other factors, on the clause type (main clause versus subordinate clause). German main clauses have an S-V-O structure where the verb always has to be placed in the second position of the clause and the particle comes in the final position. As seen in (1) and (2), the verb (e.g., *klettert*, "climbs") and the particle (e.g., *hinauf*, "up") can be separated by inserting elements such as prepositional phrases, direct objects (e.g., *einen Regenbogen*, "a rainbow") or even whole clauses (e.g., *wie im Video gesehen*, "as seen in the video").

(1) Der Elefant **klettert** einen Regenbogen **hinauf**.
"The elephant **climbs** a rainbow **up**."

(2) Der Elefant **klettert**, wie im Video gesehen, einen Regenbogen **hinauf**.
"The elephant **climbs**, as seen in the video, a rainbow **up**."

In German subordinate clauses, the verb and the particle are in reverse order compared to main clauses. Importantly, these two elements are contracted (e.g., *hinaufklettert*, "up-climbs") in the final position of the clause (3).

(3) Ich sehe, dass der Elefant einen Regenbogen **hinaufklettert**.
"I see that the elephant a rainbow **up-climbs**."

[1] In this paper we will use the term "planning unit" synonymously to Kita and Özyürek's (2003) "processing unit".

As a control, we also tested English native speakers in Experiment 1. Just like in German, English motion events are linguistically encoded with a particle verb. However, manipulating the clause type changes neither the distance nor the word order of the particle and the verb (4 and 5).

(4) The elephant is **climbing up** the rainbow. (Main Clause)
(5) I can see that the elephant is **climbing up** the rainbow. (Subordinate Clause)

The Lexicalisation Account and Planning Unit Account predict different patterns of results across the constructions in (1)–(5). First, the Lexicalisation Account (Pawley, 1987, 2010; Slobin, 2000, 2003) predicts that gestures should be shaped by the information that is encoded within a single clause (one clause construction versus two clause construction). Since German and English are both satellite-framed languages where manner and path are encoded within the same clause (one clause construction), very similar motion event gestures should occur in both languages and across clause types. Second, the Planning Unit Account predicts that co-speech gestures should be shaped by the way in which information is packaged into planning units in speech. In German main clauses, manner and path elements are separated from each other, thus it is more likely that manner and path are encoded in different planning units. Consequently, this account predicts that gesture separates manner and path more often in German main clauses compared to German subordinate clauses, where the particle verb is linguistically contracted. In English, by contrast, gestures should be similar for both main and subordinate clauses.

In Experiment 2, we sought further evidence for the Planning Unit account by inserting different linguistic elements between manner and path expressions to see if they influence gestural packaging of information differently. More specifically, we inserted either a clause ("as seen in the video") or a phrase ("in this short video"), while controlling for the length of the inserted element in terms of syllables. As mentioned above, clauses are assumed to constitute planning units (Bock, 1982; Levelt, 1989). Thus, manner and path are more likely to be in different planning units when a clause is inserted than when a phrase is inserted. According to the Planning Unit Account, we should find a higher rate of gestural separation of manner and path for inserted clauses than inserted phrases.

Methods – Experiment 1

Participants

25 native English speakers and 23 native German speakers took part in the study. Participants either received course credits or a £3 Voucher for participation. They all gave written consent to have their data included in the study. Two participants (one English and one German) were excluded from the coding and analysis because they did not use any iconic gestures which depicted the target motion events. Another two participants were excluded because their sentence structure differed from the target sentence structure in most of the responses. More specifically, one English participant used additional embedded clauses apart from the given ones ("as seen in the video") which might have influenced gesture production. One German participant was excluded because she used the past perfect tense in the main clause condition where the particle verb was not split but contracted in the final position of the clause, similar to the subordinate condition. The language and gestures produced by the remaining 23 English and 21 German speakers were coded. Because only participants that used a reasonable amount of both separated gestures and/or conflated gestures can inform us about the effect of syntactic structure on gesture separation, we included only participants who produced at least two gestures per experimental condition which depicted both manner and path gesturally (either separated or conflated). Based on these exclusion criteria, 17 English participants (average age 21.2, SD = 3.9) and 15 German participants (average age 23.3, SD = 2.6) were included in the analyses.

Material

13 short cartoons taken from the German children's series "Die Sendung mit der Maus" ("The programme with the mouse") were used as stimuli (WDR, 1974-2015). The cartoon sequences ranged from 3 to 8 seconds and all trials included a character (mouse, duck or elephant) which performed a motion event. To control speech output, participants were given a particle verb to describe the target motion event, for instance "roll into" (hineinrollen) and "spin up" (hinaufdrehen). All 13 particle verbs are listed in the Appendix (Table A1).

Design

The experiment had a 2 × 2 design with Language (English vs. German; between participant) and Clause Type (main vs. subordinate; within participant) as independent variables. The dependent variable was a binary variable as to whether gesture depicted manner and path in two separate gestures or in a single conflated gesture for a given response to a stimulus video (see the Coding and Analysis subsection for more details).

The clause type was manipulated so that the distance between the manner element and the path element in a sentence is different in German (but not in English). In the subordinate clause condition participants were instructed to begin their retellings with the element "I can see in the video that" (German: "Ich sehe im Video, dass") (see Examples in 6 and 7). Initiating a sentence with this clause forced the participants to continue with a subordinate clause. In both English and German, the manner element and the path element are adjacent with each other.

(6) Ich sehe im Video, dass der Elefant in eine Sandgrube **hineinrollt**. (German Subordinate Clause Condition)
"I see in the video that the elephant in a sandpit **in-rolls**."
(7) I can see in the video that the elephant is **rolling into** a sandpit. (English Subordinate Clause Condition)

In the main clause condition, German participants were instructed to insert the clause "wie im Video gesehen" ("as seen in the video") between verb and particle as in (8). To ensure that the participants produced this grammatical structure, they were instructed to start the sentence with the subject (the mouse, the elephant or the duck), followed by the verb in second position. Unlike in the subordinate condition, the manner element and the path element were separated by other words. The total distance between verb and satellite could vary, depending on how many other elements the participants chose to include (e.g., "with an umbrella"). Since it is not possible to insert a clause in between verb and satellite in English, English participants were instructed to place the clause "as seen in the video" at the end of their sentence as in (9), to keep the speech output and the overall complexity of the sentences in both languages as similar as possible.

(8) Die Maus **schwebt**, wie im Video gesehen, mit einem Regenschirm in den Pool **hinunter**. (German Main Clause Condition)
"The mouse **floats**, as seen in the video, with an umbrella into the pool **down**."
(9) The mouse is **floating down** into the pool, as seen in the video. (English Main Clause Condition)

Power calculation

Following Özçalişkan (2016), we estimated the power of two theoretically important pair-wise comparisons of gestural packaging of manner and path in the current study, based on previous similar studies. The first key comparison is between-participant: German Main Clauses vs. English Main Clauses. The power for this comparison was estimated, based on the observed power of an equivalent comparison in Özyürek et al. (2008). Özyürek et al. (2008) investigated how cross-linguistic (Turkish vs. English) differences in clausal packaging of manner and path in speech was associated with gestural packaging of manner and path. The second key comparison is within-participant: German Main Clauses vs. German Subordinate Clauses. The power for this comparison was estimated, based on the observed power of an equivalent comparison in Kita et al. (2007). Kita et al. (2007) investigated how the syntactic structure used in descriptions of manner and path (two-clause vs. one-clause descriptions) was associated with gestural packaging in a within-participant design. In these two studies, the items (i.e., animated clips), the procedures (i.e., narration), the dependent variable (i.e., gestural manner/path conflation vs. separation) and the independent variables (i.e., syntactic structure) were very similar to the current study.

We used the R package SIMR (Green & MacLeod, 2015) to calculate observed power for the previous studies and to estimate the necessary number of participants and items to achieve high power (80%) in the generalized mixed effect modelling. See Supplementary Material for further details of the power analyses. We also uploaded the relevant raw data from the previous studies and the R script to the Open Science Framework at https://osf.io/kf9pj/.

First, we estimated the power of the key between-participant comparison in the current study, based on the observed power of the language-effect (i.e., Turkish ($N = 20$) versus English ($n = 20$)) in Özyürek et al. (2008). Using the actual effect size found in their study, the results showed that with a reduced number of 32 participants (as in our Experiment 1), the chance of detecting the language effect is still 100% and thus well above the recommended 80% threshold.

Since the number of observations in Experiment 1 (264 observations, see results below) was lower than the estimated 398 observations for 32 participants based on Özyürek et al.'s (2008) study, we ran an additional calculation to determine how many observations per participant would be necessary to replicate their language effect. The results showed that the chance of finding the effect is already above 80% with only three observations per participant. In our Experiment 1, each participant provided at least four observations for the analysis.

Second, we estimated the power of the key between-participant comparison in the current study, based on the observed power of Kita et al.'s (2007) clause-type effect ($N = 20$). The results showed that with the 15 German participants included in our Experiment 1, the chance of detecting the effect is around 90% (note that we did not expect an effect for English participants).

Since the number of observations in Experiment 1 for German participants (132 observations, see results below) was lower than the estimated observations for 15 participants in our study (i.e., between 249 observations for 14 participants and 290 observations for 16 participants based on Kita et al. 2007)), we ran an additional calculation to determine how many observations per participant would be necessary to replicate their clause-type effect. The result showed that the chance of finding the effect is above 80% with four observations per participant, and all participants had at least four observations in our Experiment 1.

Taken together, these estimates suggest that the numbers of participants and observations in our Experiment 1 were sufficient to achieve high power.

Procedure

Participants were tested in a lab at the University of Birmingham. They were told that the purpose of the study was to investigate how different sentence structures of a language influence speech production in narrations. They were instructed to retell the cartoon clips within a single sentence using either a main or a subordinate clause construction. The participants retold the cartoons to a third person (confederate) in the room who was not able to see the cartoons. The participants were told that the confederate's task was to write down one keyword for each clip, which she/he thought was the most vital information of the story. Participants' responses were video and audio recorded for later analyses.

The experiment was set up in a PowerPoint presentation and presented to the participants on a laptop. Main clause and subordinate clause conditions were blocked and the order of conditions was counter-balanced across participants. The 13 stimuli were repeated in the two conditions. The task was explained to the participants by the experimenter walking through an example stimulus (see Fig. 1 for stimulus presentation) followed by a practice stimulus. Each condition was explained and introduced with the same example and practice stimulus. The second condition was not explained until the first condition was completed. Each trial started with a slide showing the particle verb and either the initial main clause in the subordinate condition or the embedded clause (German)/final clause (English) in the main clause condition. This slide was displayed for three seconds before the actual clip started. After the clip had been shown, the screen turned blank. In order to keep advanced sentence planning to a minimum, participants were told that they should start their retelling as soon as they saw the blank screen. For the example clip, an example answer was presented after the blank slide. This example answer illustrated the correct sentence structure to make it easier for the participants to reproduce it in their own retellings. Concerning the form of the verb (progressive form, tense), no limitations were given and the participants were told that it was up to them which form of the verb they used. Also, participants were instructed to use their hands while describing what the characters were doing, but it was not specified what type of hand movements they should produce. If participants asked how to use their hands or whether a certain type of gesture was correct, they were told that it was up to them what to do with their hands. If participants did not use their hands twice in a row, they were reminded to do so.

Although in McNeill's research tradition (cf. 2005, p. 259ff.), the experimenter does not instruct the participants to use their hands in order minimise the participants' awareness of their gesture production (see also Gullberg, 2010), we did instruct the participants to "use their hands" for several reasons. First, not all participants use gestures describing specific target events and some participants do not gesture at all (Chu, Meyer, Foulkes, & Kita, 2014). These individual differences in gesture frequency also leads to the problem of an unequal distribution of gestures across items (Gullberg, 2010). Thus, without encouraging participants to gesture, it can be particularly difficult to reach a sufficient sample size. The method of "gesture encouragement" has been used in studies investigating the role of gestures in problem solving tasks (Broaders, Cook, Mitchell, & Goldin-Meadow, 2007; Chu & Kita, 2011; Novack, Congdon, Hemani-Lopez, & Goldin-Meadow, 2014) and in a study on the production of metaphorical speech related to the use of gestures (Argyriou, Mohr, & Kita, 2017). In terms of speech-gesture

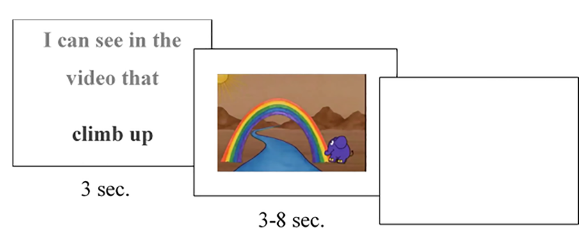

Fig. 1. Example stimulus for the English Subordinate Clause condition. During task explanation, a fourth slide illustrated the sentence that the participants were expected to produce, i.e. I can see in the video that the elephant is climbing up the rainbow. Participants started to describe the video once the slide turned blank.

production studies using motion events, this method has been employed by Özçalışkan et al. (2016). In particular, they instructed their participants to describe the stimuli scene "while using their hands as naturally as possible" (p. 12). Instructing participants to use their hands does lead to a higher frequency of gestures, but unlikely to a change in the type of gestures that participants produce, including how they gesturally express manner and path.

Data coding and analysis

The recordings were coded using the video annotator ELAN (Lausberg & Sloetjes, 2009). All speech was transcribed, but only responses where the participants produced the trained sentence structure were considered in the analyses. Hence, manner and path had to be linguistically encoded within one clause for the utterance to be included. The responses in which participants failed to follow the instructions (e.g., where they forgot to include the given clause "as seen in the video" or they produced a gesture after speech) were excluded from the analyses. Participants sometimes used other verbs than the ones presented on the slides. We only included responses with particle verbs in the analyses and then only if they were semantically similar to the given verb. These responses contained different particles (e.g. "emporkriechen" – "crawl upwards"; instead of "hervorkriechen" "crawl out") or different manner verbs ("climb out" instead of "crawl out").

The gestural coding of motion events was based on the "Cross-linguistic Motion Event Project" coding manual (used in: Kita et al., 2007; Özyürek et al., 2008) and was adapted and elaborated for the stimuli used in the current experiment. Only strokes depicting the target motion event, i.e. the gestural depiction of the given particle verb of each trial, were coded (Kita, van Gijn, & van der Hulst, 1998; McNeill, 1992).

In a first step all target event gestures were classified either as manner, path, conflated, or hybrid. Path gestures depicted only the direction of the event (e.g., for the motion "to float down" this might involve a downward movement with (an) open palm(s) but without any movements to the left or right which would indicate manner). Manner gestures depicted only the manner aspect of the motion event (e.g., for the motion "to climb up" this might involve the participant opening and closing their palm(s) without moving their arms upwards). Conflated gestures depicted motion and manner of the motion event in a single gesture (e.g., for the motion "to roll into", rotating one's wrist(s) with a simultaneous change of location away from the body). Hybrid gestures were combinations of either a path or manner gestures followed (or preceded) by a conflated gesture, all produced within a single stroke. For example, a gesture might have started out as a manner only gesture and then continued as a conflated manner and path gestures. This type of gesture occurred for example for the "spin up" stimulus. Hybrid gestures were rarely produced (i.e., 23 instances across the whole dataset).

For three of the 13 cartoon clips (slide down, jump over, jump into), it turned out that participants almost exclusively used ambiguous gestures, which can equally be classified as path only or conflated. It is difficult to gesturally depict sliding manner without showing the downward path. This makes it difficult to decide whether a diagonally downward hand movement is path only or conflated. For the stimuli jump over and jump into, the jumping movement was not repeated und thus participants indicated this motion event almost exclusively with an arc *movement*. This makes it ambiguous as to path only vs. conflated. When the event includes multiple repeated jumps (hops), "jumping gestures" could express manner by repeating the jumping movement. Such gestures were produced for example for the stimulus with the particle verb "jumping around". These three excluded items were in the stimulus set because they are useful for the research questions investigated in another study (Fritz, 2018).

Next, we classified each response to each stimulus video into three types, based on the types of gestures produced: Separated Responses, Conflated Responses, and Singleton Responses. In Separated Responses, both manner and path were expressed gesturally and at least one of them was expressed in a separated fashion. This included (a) responses with either one manner and one path gesture and (b) responses with a conflated gesture plus a manner or a path gesture, either produced separately or combined in hybrid gestures. Conflated Responses contained conflated gestures only. Finally, Singleton Responses included either manner or path gestures (but not both). These typically contained either a single manner gesture or a single path gesture, but also cases where participants produced two manner only or two path only gestures in one response. To ensure coding reliability, a second coder blind to the research question, coded 19% of the trials (including responses to all 13 stimuli movies). Only responses including a gesture were considered. The second coder was trained to identify target event gestures and to classify them as either Manner, Path, Conflated or Hybrid Gesture. Based on these annotations, interrater reliability was calculated on whether the gesture(s) identified for each trial fell into the three categories (Singleton, Conflation, Separation) used for the analyses. The two coders agreed in 88% of the responses (Cohen's $\kappa = 0.769$, $p < .001$).

Since we were interested in whether participants conflate or separate manner and path gesturally within one response, Singleton Responses were excluded from the analyses. As mentioned above, we included only participants who produced at least two gestures per experimental condition that depicted both manner and path gesturally; either separated or conflated (17 English participants and 15 German participants). 41% of all responses of these remaining participants depicted both manner and path of a target motion (39% for the English participants (i.e., 132 responses) and 44% for the German participants (i.e., 132 responses). It is not unusual for gesture production studies to lose a large proportion of the data in order to focus the analysis on the responses relevant to the research questions (Akhavan, Nozari, & Göksun, 2017; Wessel-Tolvig & Paggio, 2016; Özçalışkan et al., 2016).

The resulting data were analysed by fitting mixed effects logistic regression models in RStudio (R Core Team, 2018) using the glmer function (Bates, Mächler, Bolker, & Walker, 2015). Our dependent measure was a binary variable coding whether a participant gesturally encoded manner and path of a particular motion event in a Conflated (0) way or in a Separated (1) way, i.e. by encoding manner and path in separate gestures. We started by fitting a full model with Language (German, English) and Clause Type (Main Clause, Subordinate Clause) plus their interaction as fixed factors. The random effect structure was "maximal" (Barr, Levy, Scheepers, & Tily, 2013) for subjects and items, unless otherwise stated. The significance of a factor or an interaction between factors was determined by comparing models with and without these factors/interactions, using a maximum likelihood method (i.e., ANOVAs). We will report the chi-square statistics, degree of freedom and p-value for each model comparison.

In all analyses, we used the Nelder_Mead optimiser from the NLopt C library (Johnson, 2014) as the models converged better with this optimiser than the default optimiser in the lme4 package ("BOBQYA"). By doing this, all but one full model converged. For the model that did not converge, we slightly simplified the random effect structure for items (but not for subjects). Namely, we dropped covariance between random slopes and random intercepts for items for Clause Type. This random effect structure has been used in the literature (Baayen, Davidson, & Bates, 2008; Frisson, Koole, Hughes, Olson, & Wheeldon, 2014), and was shown to have excellent protection against Type 1 error, equivalent to the maximum random effect structure (Barr et al., 2013).[2]

Using the default optimiser ("BOBQYA") did not change the overall pattern of results. With the default optimiser, all analyses had to use a full model with the slightly simplified random effect structure for items and subjects, as described above.

All the raw data, the R scripts and the output are available on the Open Science Framework at https://osf.io/kf9pj/.

Results

Results are summarised in Fig. 2 and Table 1. We first tested whether the two groups (i.e., English and German) gestured differently for the two clause types. Response Type (Conflated vs. Separated) was entered as the dependent binary variable into a mixed effect logistic regression analysis and Language (English, German) and Clause Type (Main Clause, Subordinate Clause) and their interaction as independent variables. For this analysis, we fitted a model with a slightly simplified random effect structure for items (dropping the covariance of random slopes and random intercepts for Clause Type). The same optimiser (i.e., Nelder_Mead) was used for this model comparison as for the other analyses in Experiment 1. Model comparison indicated a significant interaction between Language and Clause Type ($\chi^2 = 4.70$, df = 1, p = .03). To explore the nature of the interaction further, we tested the effect of Clause Type on Response Type in English and in German separately, using the same logistic regression analysis, but with the maximal random structure for Clause Type for both items and subjects. In English, no main effect of Clause Type was found ($\chi^2 = 0.11$, df = 1, p = .741) (see Table 2). This is as predicted because English expresses manner and path in the same way in both subordinate and main clauses. In German, however, the main effect of Clause Type was significant ($\chi^2 = 4.79$, df = 1, p = .029) (see Table 3). The mean proportions of Separated Responses (i.e., responses with separated gestures) were higher in Main Clauses, where manner and path were separated linguistically, than in Subordinate Clauses, where manner and path were linguistically expressed together.

Discussion

The results show that clause type affected gestural depiction of manner and path of motion events in German but not in English. Gestures separated manner and path more often in German main clauses where manner and path are also linguistically separated by an inserted clause and other elements, compared to German subordinate clauses where manner and path are adjacent with each other (the path particle is a prefix to the manner verb). In English, where manner and path are always adjacent with each other independent of clause type, gestures separated manner and path equally often in the two clause types, and at about the same frequency as in German subordinate clauses. Because a clause is a good proxy for a planning unit in speech (Bock, 1982; Levelt, 1989), inserting a clause into German main clauses is likely to have separated the manner verb and the path particle into separate planning units (see Wagner et al., 2010, for evidence that the scope of planning changes when processing load increases). Thus, the findings in Experiment 1 support the Planning Unit Account; that is, gestures package information that is encoded within each planning unit for speech production.

[2] R code for the maximal model (did not converge):glmer(MP ~ condition * L1 + (1 + condition|subject) + (1 + condition|item), data = dat1, control = glmerControl(optimizer = "Nelder_Mead"), family = binomial); R code for the model with a simplified random effect structure for item (used in the model comparison): glmer(MP ~ condition * L1 + (1 + condition|subject) + (1|item) + (1|condition:item), data = dat1, control = glmerControl(optimizer = "Nelder_Mead"), family = binomial).

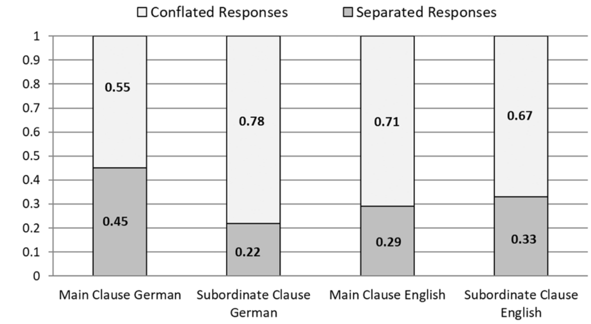

Fig. 2. Mean Proportions of Separated Responses (i.e., responses in which participants produced a separate Manner and/or Path gesture) and Conflated Responses (i.e., responses in which participants produced only Manner-Path conflated gestures) in Main Clauses versus Subordinate Clauses in English and German in Experiment 1. See Appendix B (Table B1) for standard deviations and standard errors.

Table 1
Summary of the fixed effects of the final mixed logistic model for gestural depiction in Experiment 1.

Predictor	Coefficient	SE	Wald Z	p
Intercept	−0.9301	0.4314	−2.156	.03109*
Clause Type	0.1294	0.4491	0.288	.77323
Language	1.1460	0.4158	2.756	.00586**
Clause Type * Language	−1.2814	0.5910	−2.168	.03014*

Note: N = 264, log-likelihood = −150.7.
** < .01.
* < .05.

Table 2
Summary of the fixed effects in the mixed logistic models for gestural depiction for English in Experiment 1.

Predictor	Coefficient	SE	Wald Z	p
Intercept	−1.0864	0.3855	−2.818	.00483**
Clause Type	0.1935	0.5839	0.331	.74034

Note: N = 132, log-likelihood = −75.5.
** < .01.

Table 3
Summary of the fixed effects in the mixed logistic models for gestural depiction for German in Experiment 1.

Predictor	Coefficient	SE	Wald Z	p
Intercept	0.2554	0.4978	0.513	.6079
Clause Type	−1.2733	0.5266	−2.418	.0156*

Note: N = 132, log-likelihood = −73.3.
* < .05.

These results do not support the Lexicalisation Account, which predicted no difference in the usage of gestures across clause types and languages. According to the Lexicalisation Account the lexical items which are used to describe an event influence how we conceptualise this event (Pawley, 1987, 2010; Slobin, 2000, 2003), and in turn how gesture depicts manner and path. Because German and English express manner and path in a single clause (manner verb + path particle), Slobin (2000, p.132) argues that speakers of these languages tend to conceptualise manner and path as a "single conceptual event". The current study did not provide evidence that such conceptual events shape gestural expressions.

Though the current results support the Planning Unit Account, the manipulation of main vs. subordinate clauses in German may involve some confounding. First, it is possible that mere surface distance between the manner element and the path element may be causing separation of path and manner in gestures. Second, the main vs. subordinate clauses may have different degrees of processing demands. Therefore, in Experiment 2, we focused on German and manipulated planning units within the main clause that contained manner and path, while controlling for the distance between the manner element and the path element.

Experiment 2

In Experiment 2, we compared how often participants gesturally separated manner and path in three German main clause construction: (a) a present tense clause with a clause ("wie im Video gesehen" – "as seen in the video") embedded between the manner verb (after the subject NP) and the path particle (at the end of the sentence) (see Example 10), (b) a present tense clause with an embedded prepositional phrase ("in diesem kurzen Video" – "in this short video") instead of an embedded clause (see Example 11), (c) a present perfect clause without any inserted clause or phrase, which forces the path particle to be prefixed to the manner verb at the end of the sentence (e.g., see Example 12).

(10) *Inserted Clause Condition*
Der Elefant **klettert** *wie im Video gesehen* einen Regenbogen **hinauf**.
"The elephant **climbs** *as seen in the video* a rainbow **up**."
The elephant is climbing up a rainbow, *as seen in the video*.

(11) *Inserted Phrase Condition*
Der Elefant **klettert** *in diesem kurzen Video* einen Regenbogen **hinauf**.
"The elephant **climbs** *in this short video* a rainbow **up**."
In this short video the elephant is climbing up a rainbow.

(12) *Verb Final Condition*
Der Elefant ist in diesem kurzen Video einen Regenbogen **hinaufgeklettert**.
"The elephant is in this short video a rainbow **up-climbed**.
In this short video the elephant has climbed up a rainbow.

If a clause is a good proxy for a planning unit, then the manner verb and the path particle should be separated into two planning units in the Inserted Clause Condition but not in the Inserted Phrase condition or the Verb Final condition. Thus, the Planning Unit Account predicts that gestures should separate manner and path *more often* in the Inserted Clause Condition than the other two conditions. If mere distance between the manner verb and the path particle determines how often gestures separate manner and path, then gestures should separate manner and path *less often* in the Verb Final Condition than the other two conditions.

Besides clauses, also pauses have been found to reflect higher level planning processes if they are placed at grammatical boundaries, i.e. between clauses (e.g., Goldman-Eisler, 1972; Pawley & Syder, 1983) and phrases (Ferreira, 1991; Ramanarayanan, Bresch, Byrd, Goldstein, & Narayanan, 2009). Thus, we coded all pauses (filled and unfilled) in the responses that occurred at clause/phrase boundaries in Experiment 2 in the Inserted Clause and the Inserted Phrase Condition. Coding the pauses allows us to infer whether our planning unit manipulation was effective. In particular, we would expect more pauses to occur in the Inserted Clause Condition than in the Inserted Phrase Condition.

Methods – Experiment 2

Participants

26 German native speakers took part in Experiment 2. Participants were tested at the Natural Media Lab at the RWTH Aachen University and at the University of Innsbruck. For participating in the study they received compensation in form of a €5 Voucher. All participants gave written consent to have their data included in the study. Three participants were excluded because they had learned German sign language (Deutsche Gebärdensprache). We excluded those participants because previous studies found that learning a sign language has an influence on co-speech gesture production in terms of gesture frequency (Casey, Emmorey, & Larrabee, 2012) and in terms of the production of signs while speaking (Casey & Emmorey, 2008; Casey et al., 2012). Another two participants were excluded because they did not use any iconic gestures depicting the target motion events, and one participant had to be excluded due to technical problems. The language and gestures produced by the remaining 20 participants were coded for analyses. Like in Experiment 1, the analyses only included participants who produced at least two gestures per condition which depicted both manner and path (conflated or separated). Consequently, all analyses below include 15 participants (mean age = 27.2, years, SD = 3.5).

Material

We increased the number of experimental stimuli to 15 to reduce the number of participants that needed to be excluded for lack of sufficient number of trials with relevant gesture responses. In all experimental stimuli, manner and path could be easily gesturally separated. Ten clips were taken from the German children's series "Die Sendung mit der Maus" ("The programme with the mouse") (WDR, 1974-2015). Nine of those had also been used in Experiment 1. In addition, five new experimental stimuli were taken from the "Tomato Man movies" (Özyürek, Kita, & Allen, 2001). An additional stimulus from The Programme with the Mouse movies ("slide down") was used as an example to explain the task, and two additional stimuli were used as practice clips ("climb up" and "ride around"). All particle verbs used in Experiment 2 are listed in Appendix (Table A2).

Design

The experiment manipulated construction type (inserted clause, inserted phrase, verb final; within participant). See Examples 10–12. The dependent variable was a binary variable as to whether gesture depicted manner and path in two separate gestures or in a single conflated gesture for a given response to a stimulus video (as in Experiment 1).

The construction type manipulated the likely number of planning units: two planning units for the Inserted Clause Condition, one planning unit in the Inserted Phrase Condition and the Verb Final Condition. The construction type also manipulated the distance between the manner verb and the path particle: far in the Inserted Clause Condition and the Inserted Phrase Condition, and adjacent in the Verb Final Condition. The inserted clause and the inserted phrase separated the manner verb and the path particle by the same number of syllables. The number of syllables was kept the same across conditions in order to keep the duration of producing the insertions as similar as possible. We did not control the insertions for the number of words or morphemes because they can have very different production durations (in terms of time). In particular, we wanted to avoid that producing the phrase (instead of the clause) would take less time and therefore would mean less demand on working memory. In previous studies on working memory, stimuli length has also been matched in the number of syllables rather than the number of words (e.g., Braun, Marton, & Schwartz, 2005; Marton, G., Farkas, & Katsnelson, 2006).

The Verb Final Condition also included the same inserted phrase as in the Inserted Phrase Condition (i.e., "in this short video") to keep the planning complexity equivalent. Note that the Inserted Clause Condition is the same as the Main Clause condition in Experiment 1.

Power calculation

We estimated the power of the theoretically important pair-wise comparisons, namely Inserted Clause vs. Inserted Phrase and Inserted Clause vs. Verb Final Structure, based on a previous similar study, namely Kita et al. (2007). For the calculations the R package SIMR (Green & MacLeod, 2015) was used. As in Experiment 1, we used the observed power of Kita et al.'s (2007) clause-type effect (20 participants) to estimate the power for the generalized mixed effect modelling of the current study. The results showed that with the 15 participants included in our Experiment 2, the chance of detecting the effect is around 90%. Furthermore, the number of observations included in our Experiment 2 (311, see results below) exceeds the number of observations estimated in the power analysis for 15 participants (between 249 and 290 observations). Thus, we concluded that the number of participants and observations in our Experiment 2 is sufficient to reach high power. See the Supplementary Material for further details.

Procedure

Participants came to the lab and they were told that the study was about sentence production. For Experiment 2, we decided to create a more natural speech production task. We therefore did not present the verb that participants had to use right before each video clip as in Experiment 1, but participants were familiarised with all of the stimuli clips together with the particle verbs to be used before the experimental task. This way, participants had to retrieve the particle verbs from their mental lexicon during the retelling of the cartoons as during natural speech production. In the familiarisation phase, stimuli were presented in a PowerPoint presentation. All participants saw the stimuli in the same order. Then, participants were tested to see if they had memorised the particle verbs to be used. In this test, all screenshots of the cartoon clips were shown to the participant on an A4 paper. The order of the screenshots differed from the order of the stimuli in the PowerPoint presentation. Participants were asked to produce for each screenshot the particle verb they had seen before. If a participant's response was different from the expected one, the experimenter reminded them of the particle verb to be used before the participant continued with the next screenshot.

After this familiarisation phase, the participants were instructed as to what their retellings should look like. They were told that there would be three different ways of retelling the cartoons and that the first one will be introduced shortly. Firstly, the participants were shown how the stimuli were presented on the screen. A booklet was created to illustrate how the stimuli would be presented. They were told that every trial would start with a fixation cross for 1000 ms followed by the stimulus. Then certain linguistic elements would be shown for two seconds to ensure that the participant would use the correct phrase or clause. Note that the given elements occurred after instead of before the cartoon in Experiment 2 to reduce the error in participants' construction choice. To further control the linguistic outcome, the participants were told during the instructions how to refer to the characters (the Mouse, the Elephant, the Tomato and the Triangle). Participants were also instructed as to how to use the correct sentence structure. As in Experiment 1, participants were instructed to use their hands while retelling what the characters are doing.

Two practice trials followed the instructions to assure that the participants would be able produce the correct sentence structure and, if necessary, to enable the experimenter to repeat some of the instructions. Fig. 3 shows the example stimulus in Experiment 2.

Stimuli were presented using E-Prime. Construction type conditions were blocked and instructions for every new condition were given after the previous condition was completed. The order of the three conditions was counterbalanced across participants and the trials within each block were presented in a randomised order. After the participants finished retelling one cartoon clip, the experimenter initiated the next trial with a button press. In this experiment, the experimenter was sitting opposite the participant and functioned as a listener.

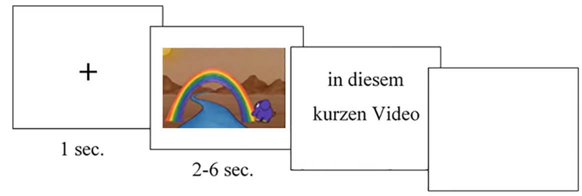

Fig. 3. Example stimulus for the Inserted Phrase Condition in Experiment 2.

Further cues were given to make it easier for participants to produce desired constructions. In the Verb Final condition, the present perfect tense was used. In German the present perfect tense is formed with the auxiliary "sein" ("to be") or "haben" ("to have") in its conjugated form (e.g. ist – 3rd Person Singular of "sein" – to be). Whether "sein" or "haben" is used depends on the particular verb. To help the participants, the correct auxiliary was displayed together with the given phrase that appeared after the cartoon clip (e.g., "ist in diesem kurzen Video" – "is in this short video"). The reflexive pronoun "sich" (oneself) was also given if it was obligatory to form a grammatically correct sentence (e.g., "hat sich in diesem kurzen Video" – "has itself in this short video").

Data coding and analysis

As in Experiment 1, responses which could not be clearly categorised as either Conflated Response, Separated Response or Singleton Response, were not included in the analysis. This made up 2.5% of the responses produced by the 15 participants who were included in the analyses. Gesture and speech coding was done in the same way as in Experiment 1. 46% of the responses of the 15 participants gesturally depicted both manner and path of a target motion (i.e., 311 trials).

As in Experiment 1, trials where participants did not produce the particles and/or verbs which were presented to them in the familiarisation phase, the trials were included if the particles and verbs were similar to the given ones. This happened rather frequently (55 trials where different verbs were used and 122 different particles across all participants who took part in the study; 900 trials in total). This mainly included particles with a very similar meaning, e.g. hinunter (down – towards the speaker) or the perspective neutral "runter" instead of herunter (down – away from the speaker). Making this difference in perspective taking was possible for all of the given particles. Importantly, these particles did not change the verb's meaning substantially (Dewell, 2011).

Inter-coder reliability was assessed as in Experiment 1. 19% of the all trials which included the trained sentence structure and also included a gesture were considered. The second coder was trained to identify target event gestures and to classify them as either Manner, Path, Conflated or Hybrid Gesture. Based on these annotations, inter-rater reliability was calculated on whether the gesture(s) identified for each trial fell into the three categories (Conflated Response, Separated Response, Singleton Response) used for the analyses. The two coders agreed in 87% of the responses (Cohen's $\kappa = 0.775$, $p < .001$).

For the pause analysis, only pauses that were longer than 200 ms (Smith & Wheeldon, 1999) and that were produced either before the inserted clause/phrase or after were considered. For each response, we determined whether or not a pause, as defined above, was produced (a binary variable).

In all analyses, we used the nloptwrap2 optimiser in the package "nloptr" (Johnson, 2014) as the models converged better with this optimiser than the default optimiser in the lme4 package ("BOBQYA"). By doing this, all full models converged.

Using the default optimiser ("BOBQYA") did not change the overall pattern of results. With the default optimiser, all analyses had to use a full model with the slightly simplified random effect structure for items and subjects, as described in the Method section of Experiment 1.

All the raw data, the R scripts and the output are available on the Open Science Framework at https://osf.io/kf9pj/.

Results

We first confirmed whether our manipulation of Construction Type affected planning units by comparing how likely it was that participants produced pauses at the boundaries of inserted clauses versus inserted phrases. We then tested the effect of Construction Type on gesture separation. As we assume that Construction Type and pauses at the boundary of inserted elements are both indicators of planning units, we tested whether Pause (yes/no) also predicted gesture separation.

To confirm our manipulation of planning units, we entered Pauses as binary variable (Pause, No Pause) in a logistic mixed effect model with Inserted Element (Inserted Clause, Inserted Phrase) as the independent variable and a maximum random effect structure for items and subjects. Model comparison indicated a significant main effect of Inserted Element ($\chi^2 = 11.073$ df = 1, $p < .001$). Participants produced significantly more pauses in the Inserted Clause Condition than the Inserted Phrase Condition (Fig. 4 and Table 4). This is in line with our assumption that inserted clauses are more likely to lead to planning unit boundaries, some of which manifest themselves as a pause.

To test for effects of our manipulation of Construction Type on gestural separation of manner and path, Response Type (Conflated, Separated) was entered into a mixed effect logistic regression with Construction Type (Inserted Clause, Inserted Phrase, Verb Final) as the independent variable. The results are summarised in Fig. 5 and Table 5. Model comparison indicated an overall effect of Construction Type ($\chi^2 = 7.725$, df = 2, p = .021).

We then tested, whether the Inserted Clause Condition differed from the other two levels (Inserted Phrase Condition, Verb Final Condition). To compare the Inserted Clause Condition against the other two levels, we fitted the same model while using Helmert coding for contrasts in "Construction Type". We found that responses of the Inserted Clause Condition differed significantly from the combined Inserted Phrase and Verb Final Conditions ($\beta = 0.88$, p = .004), indicating more separated gestures in the Inserted Clause Condition than in the two other conditions. No statistical difference was found between the Inserted Phrase Condition and the Verb Final Condition ($\beta = -0.02$, p = .959).

Because we assume a pause at the boundaries of inserted elements indicates two separate planning units, we ran the previous analysis with Pause (present or absent) instead of Construction Type as the independent variable, and we obtained essentially the same result, i.e. separated gestures are more likely when there is a pause present. More specifically, we analysed only the data from the Inserted Clause Condition and Inserted Phrase Condition where pauses were coded. Response Type (Conflated, Separated) was entered into a mixed effect logistic regression with Pause (Pause, No Pause) as the independent variable. We found a significant effect of Pause ($\chi^2 = 4.50$, df = 1, p = .034), indicating more separated gestures when a pause was present. Results are summarised in Table 6 and Fig. 6.

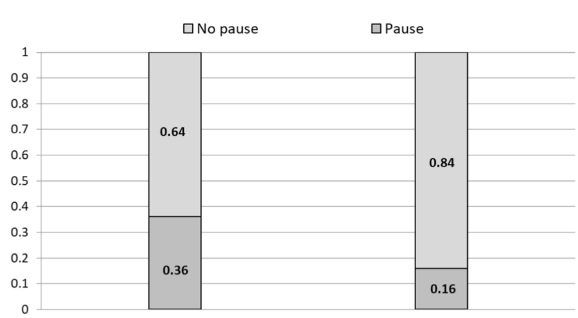

Fig. 4. Mean proportions of responses with a pause produced either before or after the Inserted Element (phrase/clause) versus responses without any pauses in Experiment 2. See Appendix B (Table B2) for standard deviations and standard errors.

Table 4
Summary of the fixed effects in the mixed logistic model for the occurrence of pauses in the Inserted Clause Condition versus the Inserted Phrase Condition in Experiment 2.

Predictor	Coefficient	SE	Wald Z	p
Intercept	−0.7517	0.3110	−2.417	.01564*
Inserted Element	−2.1309	0.7465	−2.854	.00431**

Note: N = 212, log-likelihood = -100.6.
* < .05.
** < .01.

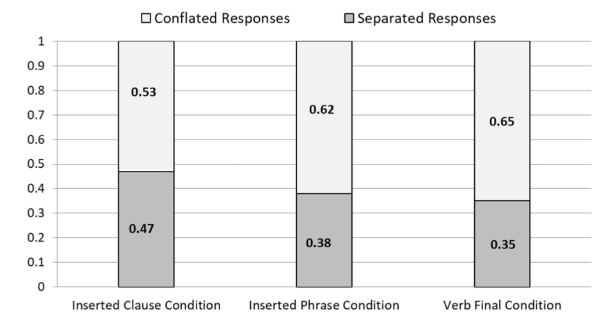

Fig. 5. Mean Proportions of Separated Responses (i.e., responses in which participants produced a separate Manner and/or Path gesture) and Conflated Responses (i.e., responses in which participants produced only Manner-Path conflated gestures) across conditions of Experiment 2. See Appendix B (Table B3) for standard deviations and standard errors.

Discussion

Participants produced more pauses between the manner verb and the path particle when they inserted a clause than when they inserted a phrase. Based on the assumption that pauses at grammatical boundaries reflect processing unit boundaries (Corley & Stewart, 2008; Ferreira, 1991; Ramanarayanan et al., 2009), these results confirm our assumption that embedding a clause makes it more likely that the manner verb and path particle fall into different planning units.

The key finding was that participants produced significantly more separated gestures when they had to embed a clause ("as seen in the video") (Inserted Clause Condition) compared to when they had to embed a phrase ("in this short video") (Inserted Phrase Condition) or when they produced the manner verb and path particle together at the end of a sentence (Verb Final Condition). The significant difference between the Inserted Clause Condition and the Verb Final Condition dovetails with the results of Experiment 1; but unlike in Experiment 1, in this experiment, the manner verb and the path particle were always in the same main clause. The significant difference between the Inserted Clause Condition and the Inserted Phrase Condition showed that what matters was the type of linguistic elements separating the manner verb and the path particle (not the lengths of the elements separating the two). Furthermore, when there was a pause at the boundary of inserted clause or phrase, which also indicates a boundary of planning units, the gesture tended to separate manner and path. The findings overall indicate that when a manner verb and a path particle are produced in different planning units, gestures are likely to depict manner and path in separate gestures. Furthermore, the likelihood of manner-path separation in gestures was comparable across the Inserted Phrase

Table 5
Summary of the fixed effects (Construction Type) in the mixed logistic model for gestural depiction in Experiment 2. Inserted Clause Condition is compared against the combined Inserted Phrase and Verb Final Condition.

Predictor	Coefficient	SE	Wald Z	p
Intercept	−0.549	0.376	−1.463	.143
Inserted Clause versus Inserted Phrase & Verb Final	0.882	0.304	2.898	.004**
Inserted Phrase versus Verb Final	−0.023	0.441	−0.051	.959

Note: N = 311, log-likelihood = −172.7.
** < .01.

Table 6
Summary of the fixed effects (Pause) in the mixed logistic model for gestural depiction in Experiment 2 (Inserted Clause and Inserted Phrase Condition only).

Predictor	Coefficient	SE	Wald Z	p
Intercept	−0.6967	0.4099	−1.700	.08922
Pause	1.1231	0.4333	2.592	.00954 **

Note: N = 212, log-likelihood = −118.3.
** < .01.

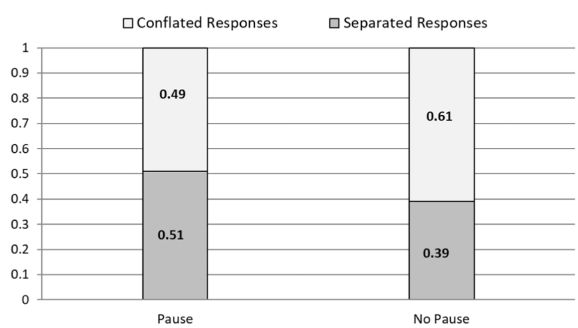

Fig. 6. Mean Proportions of Separated Responses (i.e., responses in which participants produced a separate Manner and/or Path gesture) and Conflated Responses (i.e., responses in which participants produced only Manner-Path conflated gestures) with and without a pause in the Inserted Clause and Inserted Phrase Condition of Experiment 2. See Appendix B (Table B4) for standard deviations and standard errors.

Condition and the Verb Final Condition. This suggests that mere surface distance between verb and particle does not influence gestural content. Thus, the results from Experiment 2 provide further evidence for the Planning Unit Account.

General discussion

This study aimed to provide evidence that supports the Planning Unit Account of how speech and gesture coordinate their contents, which cannot be explained by the Lexicalisation Account. In Experiment 1, manner and path were more likely to be depicted in separate gestures (as opposed to in a single gesture) when German speakers produced the manner verb and the path particle in a main clause that were separated by an inserted clause than when they produced the manner verb and the path particle together in a contracted form in a subordinate clause. In English, no difference in gestural depiction was found between the Main Clause Condition and the Subordinate Clause Condition because the verb and the particle were always produced one after the other. In Experiment 2, the manner verb and the path particle were always in the same main clause. Manner and path were more likely to be depicted in separate gestures when the manner verb and the path particle were separated by an inserted clause (Inserted Clause condition) than when they were produced together in a contracted form (Verb Final condition) or separated by an inserted phrase (Inserted Phrase condition). Because a clause is a good proxy for a planning unit (Bock, 1982; Levelt, 1989), the verb and the particle were more likely to be processed in two separate planning units in the Inserted Clause condition than in the other two conditions. Thus, the results of both Experiments 1 and 2 support the Planning Unit Account: a gesture tends to depict spatial information contained within a single planning unit for speech production. In both Experiments, the lexicalisation of manner and path (manner expressed as a verb and path expressed as a particle within the same clause) was kept constant; thus, the Lexicalisation Account cannot explain the current findings.

What are the implications of these results for the existing literature on motion event gestures? The current study, for the first time, provided evidence that unambiguously support the Planning Unit Account for coordination of speech and gesture production. Studies so far (Kita & Özyürek, 2003; Kita et al., 2007; Wessel-Tolvig & Paggio, 2016; Özyürek et al., 2008; Özyürek et al., 2005) have always compared satellite-framed constructions (i.e., path expressed as a particle) versus verb-framed constructions (i.e., path expressed as a verb). Hence, these studies could not disentangle the effect of lexicalisation patterns (satellite-framed vs. verb-framed) and that of planning units (one planning unit vs. two planning units). That is, the current study provided unambiguous evidence for the key component of the Interface Model for speech-gesture production (Kita & Özyürek, 2003). Furthermore, the pause analysis of Experiment 2, for the first time, provided evidence that clauses are a good proxy for planning units for speech production for the type of sentences used in this line of previous similar studies. This is in line with the literature on speech production (Bock, 1982; Levelt, 1989).

The findings from the current study do not directly refute the Lexicalisation Account; however, the Planning Unit Account is the most parsimonious account for the coordination of speech and gesture production. The Planning Unit Account can explain all results in the literature regarding manner and path depiction in co-speech gesture, along with similar results regarding the impact of the richness of verb meaning on gesture (Gullberg, 2011; placement verbs: Gullberg & Narasimhan, 2010; the motion verb “swing” Kita, 1993; Kita & Özyürek, 2003), and there is no finding that can only be explained by the Lexicalisation Account.

The idea that the information packaging in speech influences information packaging in gesture is further supported by comparison of co-speech and silent gestures (Özçalişkan, 2016; Özçalışkan et al., 2016). Özçalışkan et al. (2016) investigated how speakers of English (satellite-framed) and Turkish (verb-framed) depict manner and path in co-speech gestures and silent gestures. The gestural depiction differed cross-linguistically as in earlier studies (e.g., Kita & Özyürek, 2003) for co-speech gestures, but not for silent gestures. Speakers of both languages conflated manner and path in one gesture for a large majority of the time, when producing silent gestures. These results indicate that conflated gestures may reflect the “default setting” (a term previously used by Gullberg, 2011 in the domain of placement gestures) for event conceptualisation and more importantly that the speech production process triggers reconceptualisation of events. What are the theoretical implications of such a reconceptualisation process for the gesture-speech production literature? This interpretation contradicts a strong modular view of speech production processes which assumes that there is no online interaction between the preverbal message produced in the Conceptualiser and the Formulator which is responsible for grammatical encoding and for building the surface structure (Levelt, 1989). This modular view is also incorporated in a gesture-speech production model put forward by (de Ruiter, 2000). According to his Sketch Model which is based on Levelt's (1989) speech production model, speech and gesture are produced in parallel but independently after the preverbal message has been created in the Conceptualiser. Hence, this modular view on speech-gesture production cannot explain reconceptualisation of motion events based on packaging into speech-gesture planning units. If motion events are reconceptualised due to information packaging in speech, this would suggest that during speaking, event representations are generated interactively on both a conceptual and a syntactic level. Hence, this would allow online interaction between the planning of the pre-verbal message and the planning of the surface structure (i.e., syntactical planning) as suggested by Kita (1993), Kita & Özyürek (2003) and Vigliocco & Kita (2006).

Finally, we would like to point out that the current study focussed on the semantics of motion event gestures and the theories discussed in this paper, both in favour of the Planning Unit Account and the Lexicalisation Account, do not make any predictions about gesture-speech synchronisation. Thus, the current study did not investigate gesture-speech synchronisation (but see Fritz, 2018). However, we acknowledge that speakers aim to synchronise the gesture with their semantic affiliate (see McNeill, 1992, 2005), and this might have influenced gestural manner and path expressions in the current study.

Our study deployed a novel method of gesture-speech elicitation where we controlled for speech outcome and encouraged the participants to gesture. This method elicited a large number of motion event gestures produced with a given sentence structure. More specifically, 90% of all participants in Experiment 1 and 86% of all participants in Experiment 2 produced motion event gestures (including Singleton gestures) while using the specified sentence structure in more than half of the trials (M = 78% of trials, Experiment 1; M = 79% of trials in Experiment 2). The instructions regarding speech was not too difficult for the participants; only three participants across the two experiments failed to follow the instructions.

The type of instructions about gesture and speech used in our study were in line with the recent literature. Our instruction about gesture was similar to those used in some of the recent studies on gesture production, in which they also encouraged participants to produce gestures (without specifying how to or what to gesture) (e.g., Broaders et al., 2007; Chu & Kita, 2011; Novack et al., 2014) including a study on motion event gestures (Özçalışkan et al., 2016). Our instruction about speech was similar to those used in some of the recent studies on speech production, in which participants were also asked to start relatively complex sentences in a particular way, and to use specific words in the sentences (Konopka, 2012; Wheeldon, Smith, & Apperly, 2011). Nevertheless, future studies with a more naturalistic task should replicate our findings to ensure generalisability of our results.

In sum, our study provides evidence that planning units for speech production play an important role in how speech and gesture production are coordinated; more specifically, what information is encoded in each gesture tends to correspond to what information is linguistically encoded within each speech planning unit. This highlights the interactive nature of gesture-speech production – probably starting from a pre-verbal event conceptualisation all the way to the linguistic and gestural planning processes.

Acknowledgments

This research was supported by a University of Birmingham College of Arts and Law Doctoral Scholarship. We thank Effie Pearson for her help with the inter-coder reliability. Irene Mittelberg (Gesture Lab at the RWTH Aachen University) and the Department of Language and Literatures: Linguistics at the University of Innsbruck, we would like to thank for providing their facilities for testing for Experiment 2.

Appendix A

See Tables A1 and A2

Table A1
Particle Verbs from Experiment 1. German and English particle verbs given to describe the motion events depicted in the cartoon clips in Experiment 1. The particle verb "bore through" was given to describe two different video stimuli.

English	German
Example Item	
climb up	hinaufklettern
Practice Item	
ride around	herumfahren
Items	
bore through	durchbohren
climb up	hinaufklettern
crawl out	hervorkriechen
dance around	herumtanzen
drill down	hineindrehen
float down	hinunterschweben
jump around	herumhüpfen
jump into	hineinspringen
jump over	drüberspringen
roll into	hineinrollen
slide down	hinunterrutschen
spin up	hinaufdrehen

Table A2
Particle Verbs from Experiment 2. German particle verbs given to describe the motion events depicted in the cartoon clips in Experiment 2. The particle verbs "bore through" and "jump around" were both given to describe two different video stimuli.

German	English Translation
Example Item	
hinunterrutschen	slide down
Practice Item 1	
herumfahren	ride around
Practice Item 2	
hinaufklettern	climb up
Items	
durchbohren	bore through
herumhüpfen	jump around
herumtanzen	dance around
hinaufdrehen	spin up
hinaufhüpfen	jump up
hinaufklettern	climb up
hinaufrollen	roll up
hinaufsteigen	walk up
hineindrehen	drill down
hineinrollen	roll into
hinunterhüpfen	jump down
hinunterrollen	roll down
hinunterschweben	float down

Appendix B

See Tables B1–B4

Table B1
Mean Proportions, standard deviations and standard errors of Separated Responses (i.e., responses in which gesture separated Manner and/or Path) and Conflated Responses (i.e., responses that include only gestures that conflated Manner and Path) in Main Clauses and Subordinate Clause in English and German in Experiment 1.

	Conflated responses			Separated responses		
	Mean	*Stdv.*	*SE*	*Mean*	*Stdv.*	*SE*
English						
Main Clause	0.71	0.32	0.077	0.29	0.32	0.077
Subordinate Clause	0.67	0.25	0.060	0.33	0.25	0.060
German						
Main Clause	0.55	0.30	0.077	0.45	0.30	0.077
Subordinate Clause	0.72	0.21	0.054	0.28	0.21	0.054

Table B2
Mean Proportions, standard deviations and standard errors of responses with Conflated Gestures and Separated Gestures with and without a pause in the Inserted Clause and Inserted Phrase Condition of Experiment 2.

	Pause present			Pause absent		
	Mean	*Stdv.*	*SE*	*Mean*	*Stdv.*	*SE*
Construction Type						
Inserted Clause	0.36	0.22	0.056	0.64	0.22	0.056
Inserted Phrase	0.16	0.27	0.069	0.84	0.27	0.069

Table B3
Mean Proportions, standard deviations and standard errors of responses with Conflated Gestures and Separated Gestures across conditions of Experiment 2.

	Conflated responses			Separated responses		
	Mean	*Stdv.*	*SE*	*Mean*	*Stdv.*	*SE*
Construction Type						
Inserted Clause	0.53	0.31	0.080	0.47	0.31	0.080
Inserted Phrase	0.62	0.31	0.080	0.38	0.31	0.080
Verb Final	0.65	0.27	0.070	0.35	0.27	0.070

Table B4
Mean Proportions, standard deviations and standard errors of responses with Conflated Gestures and Separated Gestures with and without a pause in the Inserted Clause and Inserted Phrase Condition of Experiment 2.

	Conflated response			Separated response		
	Mean	*Stdv.*	*SE*	*Mean*	*Stdv.*	*SE*
Pause at Phrase Boundaries						
Pause present	0.49	0.37	0.095	0.51	0.37	0.095
Pause absent	0.61	0.26	0.067	0.39	0.26	0.067

C. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.jml.2018.09.002.

References

Akhavan, N., Nozari, N., & Göksun, T. (2017). Expression of motion events in Farsi. *Language, Cognition and Neuroscience, 32*(6), 792–804. https://doi.org/10.1080/23273798.2016.1276607.

Argyriou, P., Mohr, C., & Kita, S. (2017). Hand matters: left-hand gestures enhance metaphor explanation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 874–886*. https://doi.org/10.1037/xlm0000337.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. 2015, 67(1), 48. doi: https://doi.org/10.18637/jss.v067.i01

Bock, J. K. (1982). Toward a cognitive-psychology of syntax - information-processing contributions to sentence formulation. *Psychological Review, 89*(1), 1–47. https://doi.org/10.1037/0033-295x.89.1.1.

Braun, A., Marton, K., & Schwartz, R. G. (2005). The effect of age and language structure on working memory performance. In Paper presented at the XXVII Annual Conference of the Cognitive Science Society, Stresa (Italy). https://escholarship.org/content/qt6725487j/qt6725487j.pdf.

Broaders, S. C., Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2007). Making children gesture brings out implicit knowledge and leads to learning. *Journal of Experimental Psychology: General, 136*(4), 539–550. https://doi.org/10.1037/0096-3445.136.4.539.

Casey, S., & Emmorey, K. (2008). Co-speech gesture in bimodal bilinguals. *Language and Cognitive Processes, 24*(2), 290–312. https://doi.org/10.1080/01690960801916188.

Casey, S., Emmorey, K., & Larrabee, H. (2012). The effects of learning American Sign Language on co-speech gesture. *Bilingualism, 15*(4), 677–686. https://doi.org/10.1017/S1366728911000575.

Chu, M., & Kita, S. (2011). The nature of gestures' beneficial role in spatial problem solving. *Journal of Experimental Psychology: General, 140*(1), 102–116. https://doi.org/10.1037/a0021790.

Chu, M., Meyer, A., Foulkes, L., & Kita, S. (2014). Individual differences in frequency and saliency of speech-accompanying gestures: The role of cognitive abilities and empathy. *Journal of Experimental Psychology: General, 143*(2), 694–709. https://doi.org/10.1037/a0033861.

Corley, M., & Stewart, O. W. (2008). Hesitation disfluencies in spontaneous speech: The meaning of um. *Language and Linguistics Compass, 2*(4), 589–602. https://doi.org/10.1111/j.1749-818X.2008.00068.x.

de Ruiter, J. P. (2000). The production of gesture and speech. In D. McNeill (Ed.). *Language and Gesture* (pp. 248–311). Cambridge: Cambridge University Press.

Dewell, R. B. (2011). *The meaning of particle / prefix constructions in German.* Amsterdam, Philadelphia: John Benjamins.

Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language, 30*(2), 210–233. https://doi.org/10.1016/0749-596X(91)90004-4.

Frisson, S., Koole, H., Hughes, L., Olson, A., & Wheeldon, L. (2014). Competition between orthographically and phonologically similar words during sentence reading: Evidence from eye movements. *Journal of Memory and Language, 73*, 148–173. https://doi.org/10.1016/j.jml.2014.03.004.

Fritz, I. (2018). *How gesture and speech interact during production and comprehension* (PhD Thesis)Birmingham (UK): University of Birmingham.

Goldman-Eisler, F. (1972). Pauses, clauses, sentences. *Language and Speech, 15*(2), 103–113. https://doi.org/10.1177/002383097201500201.

Green, P., & MacLeod, C. J. (2015). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution, 7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504.

Gullberg, M. (2010). Methodological reflections on gesture analysis in second language acquisition and bilingualism research. *Second Language Research, 26*(1), 75–102.

Gullberg, M. (2011). Language-specific encoding of placement events in gesture. In J. Bohnemeyer, & E. Pederson (Eds.). *Event Representation in Language and Cognition* (pp. 166–188). Cambridge: Cambridge University Press.

Gullberg, M., & Narasimhan, B. (2010). What gestures reveal about how semantic distinctions develop in Dutch children's placement verbs. *Cognitive Linguistics, 21*(2), 239–262. https://doi.org/10.1515/COGL.2010.009.

Johnson, S. G. (2014). The NLopt nonlinear-optimization package. http://ab-initio.mit.edu/nlopt.

Kita, S. (1993). *Language and thought interface: A study of spontaneous gestures and Japanese mimetics* (PhD)Chicago: University of Chicago.

Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.). *Language and gesture* (pp. 162–185). Cambridge: Cambridge University Press.

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, September 17-19, 1997, Proceedings. Lecture Notes in Artificial Intelligence (Vol. 1317, pp. 23–35).

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. https://doi.org/10.1016/S0749-596X(02)00505-3.

Kita, S., Özyürek, A., Allen, S., Brown, A., Furman, R., & Ishizuka, T. (2007). Relations between syntactic encoding and co-speech gestures: Implications for a model of speech and gesture production. *Language and Cognitive Processes, 22*(8), 1212–1236. https://doi.org/10.1080/01690960701461426.

Konopka, A. E. (2012). Planning ahead: How recent experience with structures and words changes the scope of linguistic planning. *Journal of Memory and Language, 66*(1), 143–162. https://doi.org/10.1016/j.jml.2011.08.003.

Lausberg, H., & Sloetjes, H. (2009). Coding gestural behavior with the NEUROGES-ELAN system. *Behavior Research Methods, 41*(3), 841–849. https://doi.org/10.3758/brm.41.3.841.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: The MIT Press.

Marton, K., Schwartz, R. G., Farkas, L., & Katsnelson, V. (2006). Effect of sentence length and complexity on working memory performance in Hungarian children with specific language impairment (SLI): A cross-linguistic comparison. *International Journal of Language & Communication Disorders, 41*(6), 653–673. https://doi.org/10.1080/13682820500420418.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought.* Chicago: The University of Chicago Press.

McNeill, D. (2005). *Gesture and thought.* Chicago: The University of Chicago Press.

Mol, L., & Kita, S. (2012). Gesture structure affects syntactic structure in speech. In Paper presented at the 34th Annual Conference of the Cognitive Science Society, Austin, TX.

Morrel-Samuels, P., & Krauss, R. M. (1992). Word familiarity predicts temporal asynchrony of hand gestures and speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(3), 615–622. https://doi.org/10.1037/0278-7393.18.3.615.

Novack, M. A., Congdon, E. L., Hemani-Lopez, N., & Goldin-Meadow, S. (2014). From action to abstraction: Using the hands to learn math. *Psychological Science, 25*(4), 903–910. https://doi.org/10.1177/0956797613518351.

Özçalişkan, S. (2016). Do gestures follow speech in bilinguals' description of motion? *Bilingualism, 19*(3), 644–653. https://doi.org/10.1017/S1366728915000796.

Özçalışkan, Ş., Lucero, C., & Goldin-Meadow, S. (2016). Does language shape silent gesture? *Cognition, 148*, 10–18. https://doi.org/10.1016/j.cognition.2015.12.001.

Özyürek, A., Kita, S., & Allen, S. (2001). *Tomato Man movies: Stimulus kit designed to elicit manner, path and causal constructions in motion events with regard to speech and gestures.* Nijmegen, the Netherlands: Max Planck Institute for Psycholinguistics, Language and Cognition group.

Özyürek, A., Kita, S., Allen, S., Brown, A., Furman, R., & Ishizuka, T. (2008). Development of cross-linguistic variation in speech and gesture: Motion events in English and Turkish. *Developmental Psychology, 44*(4), 1040–1054. https://doi.org/10.1037/0012-1649.44.4.1040.

Özyürek, A., Kita, S., Allen, S., Furman, R., & Brown, A. (2005). How does linguistic

framing of events influence co-speech gestures?: Insights from crosslinguistic variations and similarities. *Gesture, 5*(1–2), 219–240.
Pawley, A. (1987). Encoding events in Kalam and English: Different logics for reporting experience. In R. S. Tomlin (Ed.). *Coherence and grounding in discourse* (pp. 329–360). Amsterdam: Benjamins.
Pawley, A. (2010). Event representation in serial verb constructions. In J. Bohnemeyer, & E. Pederson (Eds.). *Event Representation in language and cognition* (pp. 13–42). Cambridge: Cambridge University Press.
Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. *Language and Communication, 191*, 191–225.
R Core Team. (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. http://www.R-project.org.
Ramanarayanan, V., Bresch, E., Byrd, D., Goldstein, L., & Narayanan, S. S. (2009). Analysis of pausing behavior in spontaneous speech using real-time magnetic resonance imaging of articulation. *The Journal of the Acoustical Society of America, 126*(5), https://doi.org/10.1121/1.3213452.
Schegloff, E. A. (1984). On some gesture's relation to talk. In M. A. J. Heritage (Ed.). *In Structures of social action: Studies in conversation analysis* (pp. 266–296). Cambridge: Cambridge University Press.
Slobin, D. I. (2000). A dynamic approach to linguistic relativity and determinism. In S. Niermeier, & R. Dirven (Eds.). *Evidence for linguistic relativity* (pp. 107–138). Amsterdam/Philadelphia: John Benjamin Publishing Company.
Slobin, D. I. (2003). Language and though online: Cognitive consequences of linguistics relativity. In D. Gentner, & S. Goldin-Meadow (Eds.). *Language in mind* (pp. 157–192). Massachusetts: The MIT Press.
Slobin, D. I. (2006). What makes manner of motion salient? Explorations in linguistic typology, discourse, and cognition. In M. Hickmann, & S. Robert (Eds.). *Space in languages: Linguistic systems and cognitive categories* (pp. 59–81). Amsterdam/Philadelphia: John Benjamins.
Smith, M., & Wheeldon, L. (1999). High level processing scope in spoken sentence production. *Cognition, 73*(3), 205–246. https://doi.org/10.1016/S0010-0277(99)00053-0.
Talmy, L. (2000). *Toward a cognitive semantics, Vol II: Typology and process in concept structuring.* Cambridge: MIT Press.
Vigliocco, G., & Kita, S. (2006). Language-specific properties of the lexicon: Implications for learning and processing. *Language and Cognitive Processes, 21*(7–8), 790–816. https://doi.org/10.1080/016909600824070.
Wagner, V., Jescheniak, J. D., & Schriefers, H. (2010). On the flexibility of grammatical advance planning during sentence production: Effects of cognitive load on multiple lexical access. *Journal of Experimental Psychology: Learning, Memory and Cognition, 36*(2), 423–440. https://doi.org/10.1037/a0018619.
WDR (1974-2015). Die Sendung mit der Maus.
Wessel-Tolvig, B., & Paggio, P. (2016). Revisiting the thinking-for-speaking hypothesis: Speech and gesture representation of motion in Danish and Italian. *Journal of Pragmatics, 99*, 39–61. https://doi.org/10.1016/j.pragma.2016.05.004.
Wheeldon, L. R., Smith, M. C., & Apperly, I. A. (2011). Repeating words in sentences: Effects of sentence structure. *Journal of Experimental Psychology: Learning, Memory and Cognition, 37*(5), 1051–1064. https://doi.org/10.1037/a0023435.