



ARTICLE

Multimodal encoding of motion events in speech, gesture and cognition

Ercenur Ünal¹ , Ezgi Mamus^{2,3}  and Aslı Özyürek^{2,3,4}

¹Department of Psychology, Ozyegin University, Istanbul, Türkiye; ²Multimodal Language Department, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands; ³Centre for Language Studies, Radboud University, Nijmegen, The Netherlands; ⁴Donders Institute for Brain, Cognition and Behavior, Radboud University, Nijmegen, The Netherlands

Corresponding author: Ercenur Ünal; Email: ercenur.unal@ozyegin.edu.tr

(Received 30 March 2023; Revised 21 November 2023; Accepted 21 November 2023)

Abstract

How people communicate about motion events and how this is shaped by language typology are mostly studied with a focus on linguistic encoding in speech. Yet, human communication typically involves an interactional exchange of multimodal signals, such as hand gestures that have different affordances for representing event components. Here, we review recent empirical evidence on multimodal encoding of motion in speech and gesture to gain a deeper understanding of whether and how language typology shapes linguistic expressions in different modalities, and how this changes across different sensory modalities of input and interacts with other aspects of cognition. Empirical evidence strongly suggests that Talmy's typology of event integration predicts multimodal event descriptions in speech and gesture and visual attention to event components prior to producing these descriptions. Furthermore, variability within the event itself, such as type and modality of stimuli, may override the influence of language typology, especially for expression of manner.

Keywords: motion events; multimodal language; gesture; event integration; event cognition; cross-linguistic differences

1. Introduction

Our experience of the world consists of continuous streams of actions involving movements of people, objects and entities. Organizing these continuous streams into discrete event units and communicating about them with others is a core aspect of human cognition. How people communicate about motion events and how this is shaped by linguistic diversity are mostly studied with a focus on linguistic encoding in speech (Slobin, 1996; Talmy, 1985). Yet, human communication typically occurs in face-to-face settings with an interactional exchange of multimodal signals (Holler & Levinson, 2019; Perniss, 2018). One of these multimodal signals is the hand gestures



that accompany speech (Kendon, 2004; McNeill, 2005; Özyürek, 2017). People express motion events and their components using spontaneous gestures that have different affordances for representing and packaging event components, such as relying on iconic links between form and meaning to varying degrees (Kita et al., 2017). In this article, we review recent empirical evidence on multimodal encoding of motion to gain a deeper understanding of whether and how language typology shapes linguistic expressions in different modalities (i.e., in verbal and visual channels), and how this changes across different sensory modalities of input (e.g., information perceived through auditory versus visual modality) and interacts with other aspects of cognition (e.g., event apprehension, memory). Our goal is to bring these lines of work together for the first time to enhance our understanding of event language and cognition from a multimodal perspective and to discuss how these expand the seminal work of Leonard Talmy on typology of event integration (Talmy, 1975, 1985, 2000).

Motion events (e.g., *a woman walking towards a bus stop*) are central to everyday life and involve the displacement of an object known as the *figure* (the woman), with respect to a reference object known as the *ground* (the bus stop), along a trajectory or *path* (towards) and a *manner* through which the motion unfolds (walking). Yet, languages differ in how they encode these semantic components of motion events. Talmy's (2000) typology of event integration provides a useful framework for explaining how languages map complex event structures onto different syntactic categories.¹ This framework classifies languages based on whether the core schematic event component – path in the case of motion events – is expressed in the main verb or in a satellite. Satellite-framed languages (e.g., English, Dutch) typically express manner of motion in the main verb and use satellites such as particles or prepositional phrases to express path of motion. As a result, in satellite-framed languages, path and manner of motion are mostly conflated in a single clause (e.g., *she ran into the house*). By contrast, verb-framed languages (e.g., Turkish, Spanish, Greek, Japanese) typically express path of motion in the main verb and supporting event components, such as manner of motion, in adverbial phrases or subordinate verbs. Therefore, in verb-framed languages, path and manner of motion are typically distributed across separate clauses (e.g., *she entered the house running*) and manner is more likely to be omitted from the event description (Slobin, 2003). Although Talmy's (2000) typology of event integration captures both (intransitive) spontaneous motion events (e.g., *walk, run, jump*) and (transitive) caused motion events (e.g., *hit, push, put*), here we focus on the former as the majority of the co-speech gesture work is on spontaneous motion events.

In the recent years, a substantial amount of work has shown that speakers of these typologically different languages indeed express motion events adhering to

¹In this review, we relied on Talmy's (2000) definition to classify motion events as most research on co-speech gesture has relied on this definition. However, we should acknowledge that different definitions of motion events exist and it may be difficult to characterize which verbs are considered as motion verbs. For example, Slobin (2008) argues that motion events should have translocational meaning (e.g., *he jumped over the fence*) and distinguishes them from constructions that refer to static location (e.g., *he jumped up and down*). Slobin further distinguishes between transitive verbs indicating spontaneous motion (e.g., *walk, run, jump*) from intransitive verbs indicating caused-motion (e.g., *hit, push, put*). These distinctions are based on the differences in argument structure and other grammatical features of these verbs and constructions. By contrast, others propose that both transitive and intransitive manner verbs are potentially translocational as they can lead to a change of location (Zlatev et al., 2010). For empirical purposes, it is important to recognize these differences and clarify the criteria used for classifying motion events and motion verbs.

these attested differences (Allen et al., 2007; Bohemeyer et al., 2007; Gennari et al., 2002; Naigles et al., 1998; Papafragou et al., 2002, 2006; Slobin, 1996, 2006; see also articles in Bylund & Athanasopoulos, 2015; Ibarretxe-Antuñano, 2017). This work has shown that the typologies described above reflect the most frequent and typical patterns of linguistic expression used in narratives or short event descriptions across satellite-framed and verb-framed languages. However, deviations from these patterns have been also reported, since, for example, satellite-framed languages can also express path in the main verbs (e.g., English: *car exited the garage*) and verb-framed languages can express manner in the main verbs (e.g., Turkish: *kız içeri koştu* – a girl run inside; *koş*, corresponding to run; (Özçalışkan, 2015; Özçalışkan & Slobin, 2003). Yet, the use of such deviant patterns is limited, and speakers of these languages tend to conform to the typological patterns as reflected in their most frequent usage patterns.

In addition to systematic differences in cross-linguistic encoding in speech, motion events are also an ideal test bed for investigating whether and how these differences are reflected in gestural encoding. This is because motion events involve rich visuospatial information and gestures have modality-specific advantages for conveying visuospatial information – such as iconic gestures representing the similarity between the gesture form and the meaning of a referent. In fact, a core assumption that is shared by different models of gesture production is that gesture derives from visuospatial imagery (Sketch Model, de Ruiter, 2000; Postcard Model, de Ruiter, 2007; Gesture as Simulated Action Framework, Hostetter & Alibali, 2008, 2019; Information Packaging Hypothesis, Kita, 2000; Interface Model, Kita & Özyürek, 2003; Lexical Retrieval Hypothesis, Krauss et al., 2000; Growth Point Theory, McNeill, 1992; McNeill & Duncan, 2000). However, what is interesting from a linguistic encoding of motion events perspective is that according to the Interface Model of co-speech gesture production (Kita & Özyürek, 2003), gestures are generated through interactions between the linguistic conceptualization underlying speech production and the visuospatial imagery underlying gesture production. Through these interactions, co-speech gestures represent information following language-specific constraints on information packaging in the speech that they accompany. That is, each co-speech gesture is likely to express semantic information that is encoded within one processing unit (i.e., verbal clause) in speech.

This view is supported by cross-linguistic work showing that gestural encoding of event components differs in ways tightly linked to linguistic encoding in speech (Akhavan et al., 2017; Gullberg et al., 2008; Kita et al., 2017; Kita & Özyürek, 2003; Özçalışkan et al., 2016a, 2016b; Özyürek et al., 2005). However, the strict effect of language typology on multimodal encoding of motion events does not persist under all circumstances. In the sections that follow, we discuss empirical evidence on how multimodal expressions of motion in speech and gesture might interact with the effect of language typology and other aspects of cognition (e.g., visual attention and memory).

2. Cross-linguistic variability in encoding of motion events in speech and co-speech gesture

As mentioned already, one important consequence of the typological patterns in motion event encoding predicted by the event integration framework (Talmy, 2000)

is the packaging of semantic components at the clausal level in speech. In satellite-framed languages, path and manner are tightly packaged in a single clause, whereas in verb-framed languages, path and manner are separated across two clauses, and in most cases, manner might be omitted due to being expressed outside of the verb in a subordinate verb. In a seminal work that cross-linguistically tested the consequence of these typological patterns on gestural representations of motion events, speakers of a satellite-framed language, English, and two verb-framed languages, Turkish and Japanese, were asked to describe motion events (Kita & Özyürek, 2003). The results revealed that English speakers typically encoded path and manner of motion in a single verbal clause in speech and also tended to conflate path and manner in a single gesture. On the other hand, speakers of Turkish and Japanese typically distributed path and manner information across different clauses in speech and also tended to produce separate gestures for path and manner. Similar findings have been replicated across different languages and language pairs, such as English (Kita et al., 2007), Farsi (Akhavan et al., 2017), French (Gullberg et al., 2008), Turkish (Mamus et al., 2022, 2023; Ünal et al., 2022), Dutch-Turkish (ter Bekke et al., 2022), Turkish-English (Özçalışkan et al., 2016a, 2016b; Özyürek et al., 2005), Korean-English (Choi & Lantolf, 2008) and Japanese-English-Turkish (Kita & Özyürek, 2003). Thus, co-speech gesture often follows the typological patterns in motion event encoding defined by Talmy (2000).

Although the tight semantic link between speech and gesture is well established now, their semantic relation has been initially depicted in a different way. McNeill and Duncan (2000) claimed that gestures serve a compensatory purpose for speech. They examined Spanish and English speakers' motion event descriptions of Tweety cartoons and found that Spanish speakers – in line with their language typology – often omit manners in speech but depict them in gestures. Thus, gestures encode additional information to speech when this information is difficult to encode linguistically. However, later empirical work provided abundant evidence against the claims of McNeill and Duncan (2000), and showed that speech and gesture typically express similar content as co-speech gestures appear with the element expressed in the main verb (Akhavan et al., 2017; Gullberg et al., 2008; Kita & Özyürek, 2003; Özyürek et al., 2005; ter Bekke et al., 2022; Ünal et al., 2022). For example, when English and Turkish speakers encode both path and manner, they package them in syntactically different ways both in speech and in co-speech gesture – that is, English speakers producing conflated gestures more frequently and Turkish speakers producing separate path or manner gestures more frequently. Nevertheless, when the same English and Turkish speakers encode only path or only manner of motion, that is when they produce syntactically similar descriptions, importantly, their co-speech gestures also look similar (Özyürek et al., 2005). Crucially, in cases when one motion event component is omitted from speech, the same semantic element is also omitted from co-speech gesture (see also Sümer & Özyürek, 2022). Furthermore, in verb-framed languages where path is expressed in the main verb, even in cases when both path and manner are expressed in speech, people may express only path in gesture (French: Gullberg et al., 2008; Turkish: Mamus et al., 2022, 2023; Özçalışkan et al., 2016b, 2018; ter Bekke et al., 2022). A similar pattern is observed in Farsi that has a mixed verb-framed and satellite-framed typology (Akhavan et al., 2017). In the study with Farsi speakers, participants typically encoded path in light verbs plus prepositions and manner in adverbs (e.g., corresponding to *the girl came towards the tree in a*

running fashion) and produced gestures that only expressed path. These findings extend previously shown links between language-specific encoding in speech and co-speech gesture by showing that the semantic elements that can be packaged within the main verb in speech guide co-speech gesture production. Thus, these findings also provide counter evidence for a mechanism where gesture organizes in such a way that it expresses conventionally what is not expressed in motion event components in speech unlike the claims of McNeill and Duncan (2000).

3. Within-language variability and susceptibility of manner in multimodal encoding of motion events

While there is a tight semantic link between speech and co-speech gesture when cross-linguistic variability is concerned, studies also show some within-language variability in motion event encoding in speech, which has further consequences for co-speech gesture. These studies indicate that the variation in the type of manner in the event might trigger variation in speech and gesture patterns.

One study with speakers of English has shown that the syntactic encoding of path and manner varies depending on how manner relates to path of motion (Kita et al., 2007). In that study, English speakers were more likely to tightly package path and manner in a single clause in speech and a single gesture when the manner is *inherent* to the change of location (e.g., a triangle is *jumping* while going *up* on an inclined surface). On the other hand, the same participants were more likely to use separate clauses that express either path or manner of motion together with path-only or manner-only gestures when the manner was *incidental* to the change of location (e.g., a triangle *rotating* on its horizontal axis while going *down* into the water).

Further evidence on the influence of manner type on the mention of manner comes from cross-linguistic work. In a study comparing motion event descriptions of English and Greek speakers, both language groups described events involving predictable and unpredictable manners (Papafragou et al., 2006). When manners are predictable from the context (e.g., *a man is walking down the stairs*), Greek speakers frequently omitted the manner of motion from their event descriptions, in line with the verb-framed typology, as these manners could be inferred even if not explicitly expressed in speech. However, when manner was rather unusual and not easily predictable (e.g., *a man is sliding down the stairs*), Greek speakers were twice as likely to mention it compared to when it was a predictable manner. These findings converge with recent cross-linguistic evidence from speakers of Turkish and Dutch, showing that when describing spontaneous motion events involving a person changing location in non-default ways (e.g., *twirling, skipping*), both language groups mentioned manner more often than path (ter Bekke et al., 2022).² Furthermore, in the same study, both Turkish and Dutch speakers were equally likely to gesture about path and manner. Thus, for Turkish speakers, these

²These findings converge with recent evidence from other domains such as numerical expressions (Alcaraz-Carrión et al., 2022) and time expressions (Cánovas et al., 2020) showing that gestures were more likely to accompany low-frequency and less predictable linguistic expressions. This suggests that apart from typicality or predictability of the event itself, the typicality and predictability of the linguistic structure in the accompanying speech may influence expressions in the gestural modality.

atypical manners have possibly increased the frequency of mention of manner in speech, which further increased the frequency of manner gestures, eliminating previously shown manner omissions in verb-framed languages. This interpretation is corroborated by findings of another study (Ünal et al., 2022) that used similar motion event stimuli with the exception that manner of motion were rather typical ways of changing location (e.g., walking, running). In that study, Turkish speakers mentioned both path and manner of motion in speech but often produced path-only gestures. On an interesting note, recent evidence suggests that expression of manner gestures is more susceptible to the influence of social context than path gestures in Korean and Catalan – both verb-framed languages – for example, speakers produced fewer manner gestures when they interacted with an unknown superior compared to interaction with a friend (Brown et al., 2023).

Together, the studies reviewed in this and the previous sections suggest that spoken descriptions of typical motion events are most likely to be characterized by the patterns defined by Talmy's typology of event integration (2000). These patterns in speech further influence the frequency, form and content of co-speech gestures, as in the case of path-only gestures accompanying path verbs in speech in verb-framed languages. However, other factors such as the saliency or the typicality of the manner or pragmatic requirements can interact with the lexical or syntactic constraints on the expression of motion event components. For example, type of manner can influence syntactic choices within speakers of a satellite-framed language, which could override typological patterns (Kita et al., 2007). Furthermore, in verb-framed languages – such as Greek, Turkish, Korean and Catalan – manner expressions in event descriptions may be more sensitive and open to variation than path expressions. This may arise from the fact that manners are often optional and omissible in verb-framed languages (Slobin, 2003; Sümer & Özyürek, 2022). These patterns conform to but also extend Talmy's (2000) typology of events and demonstrate the need to take into account the variability within the event itself.

4. Role of sensory modality and visual experience in multimodal encoding of motion events

Most studies reviewed so far have used visual stimuli to examine motion event expressions and their patterns by using video-clips, cartoons, line drawings and so on (Akhavan et al., 2017; Gennari et al., 2002; Gullberg et al., 2008; Kita & Özyürek, 2003; Papafragou et al., 2002; Slobin et al., 2014; ter Bekke et al., 2022; Ünal et al., 2022). These studies have not taken into account whether these patterns might change depending on the modality of the input. The sensory modality of input may influence multimodal encoding of motion events as each sensory modality has different perceptual affordances – for example, vision dominates in spatial perception despite the fact that auditory and haptic channels can also provide spatial information through cross-modal integration (Alais & Burr, 2004; Eimer, 2004; Thinus-Blanc & Gaunet, 1997).

One exception to the previous literature using visual stimuli is the work of Özçalışkan et al. (2016b, 2018). They conducted cross-linguistic studies to examine differences in packaging of motion event elements in congenitally sighted, blind and blindfolded (sighted with covered eyes during the experiment) speakers of Turkish and English. They created haptic static scenes consisting of landmark

objects (e.g., a toy house) and dolls in different postures to indicate the motion (e.g., *a girl running into a house*). Sighted participants observed the scenes without touching, whereas blind and blindfolded participants explored the scenes through touch. The main goal of the study was to investigate how sighted and blind participants would package motion event elements at the syntactic level. The findings revealed that Turkish and English speakers packaged path and manner according to their language typology and there were no differences between blind and non-blind speakers within a language group. Thus, Turkish blind and non-blind participants separated path and manner, while English blind and non-blind participants conflated them both in speech and in co-speech gesture, as in previously mentioned cross-linguistic work (e.g., Kita & Özyürek, 2003; Özyürek et al., 2005). These findings indicate similarities in the syntactic form of co-speech gestures across participants with different visual experience (i.e., blind versus non-blind participants). Thus, they suggested that language typology is the main factor that determines speech and gesture patterns. However, the role of input modality on multimodal descriptions was not an interest of Özçalışkan et al. (2016b). Therefore, they did not report a direct comparison between descriptions of blindfolded and sighted participants.

A recent study that systematically investigated the role of input modality in sighted individuals, however, showed that sensory modality of input matters for encoding of motion events in speech (Mamus et al., 2022). In that study, everyday motion events (e.g., someone running to an elevator) were created as audio-only, visual-only or multimodal (audio+visual) stimuli. Turkish speakers who only listened to the events produced more path and fewer manner descriptions in speech compared to another group of Turkish speakers who watched the event with or without the audio. Therefore, compared to auditory input, visual input elicited manner more than path expressions, in a verb-framed language. Interestingly though, the change in the speech patterns were not reflected in co-speech gestures. Speakers dominantly produced path-only gestures regardless of the sensory modality of input. Thus, speech appears to be more sensitive to the input modality compared to gesture that adhered to the verb-framed typology of the language. These findings suggest that the sensory modality of input influences speakers' encoding of motion events to some extent, apart from the language typology (Slobin, 1996; Talmy, 2000).

Another study by Mamus et al. (2023) examined the role of (lack of) visual experience on motion event descriptions of blind, blindfolded and sighted speakers of Turkish using auditory stimuli only. They compared how often blind and non-blind speakers encoded path and manner in their speech and co-speech gesture. Blind participants had more path than manner expressions in their speech compared to sighted participants. In co-speech gestures, blind participants overall produced less path and less manner gestures than sighted participants, but path-only gestures were dominant across all groups. This suggests that lack of exposure to vision for long term might in fact change expression of path and manner both in speech and in gesture.

Taken together, these findings reveal some similarities and differences between the role of sensory modality of input and visual experience on expression of motion in speech and gesture, at least in speakers of a verb-framed language. Both temporary changes in the input modality of sighted participants and lack of visual experience for long term influenced manner expression in speech. Participants

expressed manner less frequently when they experienced motion events through audio as opposed to visual stimuli and when they lacked a visual experience of the world in general. This concurs the idea that linguistic encoding of manner is more susceptible to visual input than path is – at least in a verb-framed language. Motion expressions in gesture were modulated by long-term visual experience – possibly by changing how sensory information was mapped onto the event construal over time – but not by temporary changes in the sensory modality of input. However, the packaging of path and manner in speech and co-speech gesture seems resistant to change regardless of the effect of (lack of) long-term visual experience. This highlights the importance of comparing different sensory modalities of input, as well as different aspects of multimodal event descriptions (e.g., as frequency of mention in addition to packaging) to better understand whether and how sensory modality of input and long-term visual experience influences multimodal motion event descriptions. Furthermore, as in the work reviewed in the previous section, manner expression seems to be more susceptible to variability in input modality than path expression.

5. Interface between multimodal encoding motion and cognition

So far, we have focused on the variations in speech and gesture across speakers of different languages as well as within speakers of a single language driven by various factors such as type of manner in which the motion event unfolds, the sensory modality of stimuli and (lack of) long-term visual experience. We now turn to the interface between multimodal encoding of motion in speech and gesture and on other aspects of cognition, such as visual attention during event apprehension and event memory.

The link between speech production and event apprehension is well established. While or before describing visual events, speakers attend to those aspects of the events that they (plan to) speak about (Gleitman et al., 2007; Griffin & Bock, 2000; Konopka & Meyer, 2014; Meyer et al., 1998; van de Velde et al., 2014). Such findings have been taken as evidence for a speech production model according to which speaking begins with a preverbal apprehension of an event that includes the people, objects, entities and spatial-temporal features involved in the event (Levelt, 1989). A cross-linguistic extension of this model that is predominantly supported by empirical work in the domain of motion is known as the *thinking for speaking hypothesis* (Slobin, 1996). In this view, speakers attend to the aspects of experience they plan to communicate about in ways consistent with how their language packages information at the lexical and syntactic level. Consistent with this possibility, cross-linguistic eye-tracking studies show that speakers of English and Greek attend to motion events differently in ways that parallel how their language expresses motion, but only prior to describing the events in speech (Papafragou et al., 2008; see also Bunker et al., 2012, 2021; Flecken et al., 2014; Sakarias & Flecken, 2019; Trueswell & Papafragou, 2010).

If language specificity of gestures accompanying speech is an outcome of an interface between linguistic conceptualization and visuospatial imagery during message preparation (Kita & Özyürek, 2003), one would expect similar links between event apprehension and co-speech gesture production. This possibility was recently tested in an eye-tracking study with Turkish-speaking adults (Ünal

et al., 2022). Participants watched videos of motion events while their eye-movements were recorded. Once the video ended, they described the event to an addressee sitting across them. The videos were constructed in such a way that the path and manner information relevant for linguistic descriptions of motion could be defined as separate areas of interest. As a first step, motion descriptions in speech and gesture were investigated. As predicted by Talmy's (2000) typology of event integration, the majority of the descriptions included both path and manner of motion, with path of motion mostly expressed through path verbs. Furthermore, consistent with the predictions of the Interface Model (Kita & Özyürek, 2003), such spoken descriptions were frequently accompanied by gestures that only expressed path of motion.

Next, the relation between visual attention allocated to event components and the encoding of event components in speech and gesture was examined. Of interest was whether the additional encoding of path in gesture would be linked to even more visual attention allocated to path of motion during message preparation. The results were in line with this prediction: Turkish speakers allocated more attention to path of motion when their speech was accompanied by a path gesture compared to when they did not express any motion information in gesture. Crucially, these differences were found after controlling for the content of the motion descriptions in speech such that all descriptions expressed both path and manner. This suggests that the links between visual attention and gesture production emerged in addition to the links found between visual attention and speech production.

The study reviewed above demonstrates links between motion event descriptions in speech and gesture and event cognition *before* producing these descriptions. Another way in which such links could be seen is *after* producing multimodal motion event descriptions. A number of studies have explored this possibility by investigating whether encoding motion events in speech and gesture has consequences for memory for motion event components and whether this relation is modulated by language typology. We begin with studies testing the relation between motion event speech and memory and then discuss whether encoding motion event components in gesture has additional benefits for motion event memory.

In one study, adult English speakers had better motion event memory when their memory was tested after describing events compared to when they did not describe the events (Bunger et al., 2012). However, this study did not test whether the gains in memory accuracy come from mentioning specific motion event components in the description of that event. In another study examining the relation between motion event descriptions and memory more closely (Skordos et al., 2020), speakers of English and Greek viewed a set of motion event clips and were asked to produce a single verb describing each event. Immediately after this task, they saw another set of motion event clips and indicated whether or not these clips were the same as the ones they saw in the production task. Verb production patterns were consistent with Talmy's (2000) typology: English speakers were less likely to produce path verbs than Greek speakers, and Greek speakers were less likely to produce manner verbs than English speakers. In the memory task, participants had worse memory for manner of motion when they used a path verb to describe the event. However, memory for path of motion was not predicted by the type of verb produced. Importantly, the relation between verb production and memory was not modulated by language typology.

The findings above are suggestive of a relation between expressing motion event components in speech and memory for those components; however, this work has an important limitation. Participants were required to use a single verb to describe the event. Nevertheless, people typically describe events in complete utterances rather than in single verbs. This limitation was addressed by a recent study comparing another pair of typologically different languages: Turkish and Dutch (ter Bekke et al., 2022). Participants viewed and described videos of motion events to an addressee seated across them. They were not given any instructions that would constrain the kind of descriptions they would produce. Later, participants saw another set of motion events and indicated whether or not they had seen them before. Half of the events in this second set were the same as the ones they saw before. Of the remaining events, half had a different path and the other half had a different manner. Neither Turkish nor Dutch speakers had manner memory that were above chance level. Therefore, the relation between motion event speech and memory could only be tested for path of motion. The results revealed that both Turkish and Dutch speakers were better at recognizing that the path of motion had changed when they had mentioned path of motion in speech. Furthermore, as in the previous study, the relation between motion event descriptions in speech and motion event memory was not modulated by language typology.

Together, these findings on the relation between motion event encoding in speech and memory indicate that encoding motion event components in speech is associated with better memory for those components. However, this relation between speech and memory is not further modulated by language typology. Instead, this seems to be characterized by some patterns that generalize across speakers of typologically different languages. For example, memory for manner is worse than memory for path for both speakers of verb-framed and satellite-framed languages – Greek-English (Skordos et al., 2020), Turkish-Dutch (ter Bekke et al., 2022), English (Bunger et al., 2012), Greek-English (Papafragou et al., 2002) and Spanish-English (Gennari et al., 2002). All in all, these studies suggest that in addition to some language-general tendencies, whether or not an event component is encoded in speech seems to be more critical for motion event memory than how it is encoded with regard to a language typology.

The mechanism behind the relation between motion event encoding in speech and memory could be better understood when the relation between motion event encoding in speech and event apprehension is considered. Since the event components mentioned in the spoken descriptions are allocated more visual attention and construed as part of the conceptualization of this event (Levelt, 1989; Papafragou et al., 2008), they may also be remembered better. As discussed above, gesturing about motion event components guides visual attention to motion event components (Ünal et al., 2022). Furthermore, these effects emerge in addition to the effects of speech production on visual attention. An important question is whether similar links exist between motion event encoding in gesture and memory.

Contrary to this possibility, the study by ter Bekke et al. (2022) showed that gesturing about path of motion was not related to improvement in path memory. Importantly, in this study, gestures were spontaneously produced and when participants produced a path gesture, they typically described path of motion in their speech. Therefore, the findings of this study speak to (the lack of) an additional advantage of gesture production on top of speech production on motion event memory rather than the relative benefits of motion event expressions in speech

versus gesture. In order to evaluate the latter possibility, further research focusing on non-redundant gestures that express motion event components not expressed in the accompanying speech is necessary. Finally, similar to the findings on the relation between motion event descriptions in speech and memory, language typology did not interact with the relation between motion event descriptions in gesture and motion event memory. Together, these findings indicate that even though language typology shapes multimodal event descriptions in speech and gesture, and visual attention to event components prior to producing those descriptions, it does not further modulate the relation between these event descriptions and the cognitive processes that follow these descriptions, such as subsequent memory.

6. Discussion and conclusions

In this article, we reviewed a growing body of research on multimodal encoding of motion events in speech and gesture. We asked to what extent and under which conditions language typology shapes event descriptions, whether this changes across the modality of expression or sensory modality of input, and how event descriptions in different modalities interface other aspects of cognition such as visual attention to events and memory for events. Our goal was to draw on the evidence from these different lines of research to broaden our understanding of language as a multimodal and multisensory phenomenon and to expand on the insights provided by Talmy's typology of event integration.

Cross-linguistic evidence on the multimodal expression of same motion events shows that language typology shapes motion event descriptions in speech and co-speech gesture in ways paralleling the patterns formulated by Talmy's typology of event integration (1975, 1985, 2000). According to this typology, the informational content of a unit of processing in speech (i.e., a clause) differs in descriptions of motion events in satellite-framed and verb-framed languages. Similarly, the semantic elements expressed in gestures accompanying spoken motion event descriptions also differ across speakers of satellite-framed and verb-framed languages. This similarity is uniquely predicted by the Interface Model (Kita & Özyürek, 2003), according to which gestures are generated through an interface between visuospatial imagery and how information is packaged within a processing unit in the accompanying speech. Importantly, these cross-linguistic differences in co-speech cannot be explained by other models of gesture production proposing that gestures planned prior to and independent of linguistic formulation in speech (de Ruiter, 2000, 2007; Krauss et al., 2000).

Converging evidence for this view comes from recent cross-linguistic work examining descriptions of motion events only with gesture in the absence of speech (i.e., silent gesture). For example, when asked to describe motion events with silent gestures, speakers of both Turkish and English conflated path and manner in a single gesture unlike the typological patterns in Turkish (Özçalışkan et al., 2016a). In the same study, gestures produced along with speech differed cross-linguistically, following typological patterns (see also Özçalışkan et al., 2018, 2023). Thus, gestural representations of events reflect language typology but only when gestures accompany speech. These findings corroborate the idea that language specificity of co-speech gestures arise from interactions between linguistic conceptualization and the visuo-spatial imagery during online language production.

Nevertheless, motion event expressions in speech and gesture are not only shaped by language typology. Several studies demonstrated that factors related to event structure, such as the type of manner in which the motion unfolds, the sensory input through which the event to be described is perceived, speakers' lifetime experience with the visual world, and pragmatic requirements can possibly modulate how people express motion event components in speech. Moreover, under certain circumstances, variations in motion descriptions in speech further modulate the expression of the same events components in co-speech gestures. In both modalities, expression of manner seems to be more susceptible to these influences than path, possibly because manner is a more peripheral event component.³ There has been a growing body of work investigating how linguistic expression of core versus peripheral event components is modulated by linguistic, conceptual, or pragmatic factors (Do et al., 2020, 2022; Grigoroglou & Papafragou, 2019; Ünal et al., 2021). The work reviewed here extends this line of work to motion expressions in the visual modality (i.e., co-speech gestures) and to the path and manner components of motion events.

Another conclusion emerging from the work reviewed here concerns the relation between multimodal motion event descriptions and other aspects of cognition. Prior to language production, language-specific encoding of motion event components in both speech and gesture guide allocation of attention to event components. These findings support influential theories of speech production (Levelt, 1989) and the thinking for speaking hypothesis (Slobin, 1996) by showing that the way speakers pick up information from the visual world takes into account the lexical and structural constraints on how those aspects are expressed in language. These findings also provide evidence for the Interface Model (Kita & Özyürek, 2003) from the planning phase of multimodal language production by showing that gesture production interfaces with event conceptualization in ways similar to speech production.

However, after language production, how event descriptions interface with cognition is different for speech and gesture – at least for motion event memory. Motion event descriptions in speech but not gesture predicts memory for those event components mentioned in the descriptions. Even though the evidence reviewed here only concerns the domain of motion events, there is converging evidence from the domain of object locations (Karadöller et al., 2021, 2022). This work shows that memory for object locations is predicted by whether or not object locations are expressed in the descriptions of scenes. But whether object locations are described in the verbal/auditory modality (i.e., speech) or visual modality (i.e., gesture or sign) does not matter for memory.

Unlike the relation between multimodal motion event descriptions and visual attention, the relation between speech and memory does not seem to be modulated by language typology. Furthermore, memory for motion events seems to be characterized by some cognitive biases that generalize across speakers of different languages. Memory for manner might be more fragile (Gennari et al., 2002; Papafragou et al., 2002; ter Bekke et al., 2022) and even affected by how path is expressed (Skordos et al.,

³Alternatively, this variation in manner expression might not be an outcome of event structure and which event components are core versus peripheral. Instead, this pattern might be due to language modality and specific to spoken languages. There is evidence that in sign languages manner expression is less susceptible to variability (Sümer & Özyürek, 2022), possibly because the iconic and analogue expressions in sign provides the means for easily encoding a visually grounded event. This contrasts with the categorical forms in speech used for expressing event components (for a related discussion see also Talmy, 2003).

2020), while the opposite is not true. These findings are consistent with the idea that path is a more central aspect of motion and manner is more peripheral – a distinction that is also central to Talmy’s typological classification of languages. Importantly, these findings on the relation between motion event speech and memory also cohere with cross-linguistic evidence on motion event descriptions in speech and gesture, for example, the case of path bias in gestures and manner expressions being susceptible to the influence of factors other than language typology.

6.1. Future directions and open questions

At present, several questions remain open for further research on this topic. One critical issue is the definition of path and manner verbs in motion events. As introduced previously, path verbs of motion specify a direction of motion without detailing how the manner occurs whereas manner verbs of motion specify a manner of motion without specifying its precise direction. Manner also encompasses a wide range of aspects including motor pattern (e.g., crawl, walk, run), attitude (e.g., stroll, amble, saunter), and rate (e.g., hurry, dash) (e.g., Slobin, 2004; Slobin et al., 2014). However, it is not always straightforward to decide whether a motion verb is a manner or path verb as there are verbs encoding manner and path together (e.g., Cifuentes-Férez, 2008; Slobin et al., 2014). For example, the verb *fall* (as well as *sink*) is considered to be path verb in some studies but a manner verb in others. Support for *fall* as a path verb comes from that *fall* is a change-of-location verb and there is no particular motor pattern, rate, or attitude involved. Yet, it is also possible to claim that *fall* includes information about manner in a broad sense (such as descending with increasing momentum). Others argue that it is an in-between case which contains manner (in a broad sense) and path (downwards motion) together. As this categorization may vary depending on the subject of interest, researchers should provide operational definitions of path and manner verbs in a particular work, which is often not the case in the work discussed above.

The majority of the empirical evidence reviewed in this article comes from studies conducted with adults. In terms of development, it is widely accepted that children’s early motion event descriptions in speech are both characterized by language-specific and language-general patterns (Allen et al., 2007). However, the development of language-specific co-speech gestures is less well understood. Some findings indicate language specificity in speech and gesture around the same time (Özçalışkan, 2007; Özçalışkan et al., 2023), and some findings suggest that language-specific patterns develop later in gesture than in speech (Özyürek et al., 2008). Furthermore, the relation between speech and gesture may change throughout development, since children are typically more likely to express non-redundant information in co-speech gestures, for example, when describing caused motion events (Furman et al., 2014; Niu et al., 2022). Thus, developmental work can provide novel insights into language-specific influences on multimodal event descriptions in speech and gesture.

This article mainly focuses on motion event descriptions in first language and how they interact with broader cognition. However, the majority of language users around the world are multilingual. There has been a growing body of research investigating how people adjust their event descriptions when speaking a typologically different second language (e.g., Aktan-Erciyas et al., 2020; Cadierno, 2008; Emerson et al., 2021; Hohenstein et al., 2006; Soroli et al., 2012), how speaking typologically different

second language interacts with motion descriptions in first language (A. Brown & Gullberg, 2010, 2011; Emerson et al., 2021), and how bilingual experience interacts with other aspects of motion event cognition (e.g., Aktan-Erciyes et al., 2022). Future research should focus on how the factors discussed in the present article interact with bilingualism in shaping motion event expressions in speech and gesture.

The present discussion focuses on multimodal expression of literal motion where people describe a physical change of location. However, motion verbs are also used to refer to non-physical motion when describing a metaphorical change of location (Özçalışkan, 2003, 2004; Lakoff & Johnson, 2008; see also Caballero, 2007, 2017; Johnson & Larson, 2003). A growing body of research strongly suggests that the typological patterns in verb-framed and satellite-framed languages in the expression of literal motion are also extended to metaphorical uses of motion expressions (Özçalışkan, 2003, 2004, 2005). Nevertheless, existing work on expression of metaphorical motion is based on written descriptions in magazines (Caballero, 2007, 2017) and novels (Özçalışkan, 2003, 2004) or based on elicited descriptions in written form (Özçalışkan, 2005). Future work based on naturalistic face-to-face interactions can reveal whether these typological patterns are manifested in both spoken expressions and multimodal expressions in speech and co-speech gesture.

As discussed above, Talmy's (2000) typology of event integration offers a binary classification of languages as verb- or satellite-framed based on whether path of motion is encoded within or outside the main verb. However, this classification fails to characterize languages where path and manner are expressed by equivalent linguistic elements. For example, languages with pervasive serial constructions allow two or more main verbs slots in a single clause (e.g., *pǎ chū* - *run exit* in Mandarin Chinese) and cannot be classified as either verb- or satellite-framed based on the semantics of the main verb. This limitation of Talmy's typology is addressed by proposing a third category, known as equipollently-framed languages (Slobin, 2004), and empirical studies have provided empirical support for the presence of this third typology (Chen & Guo, 2009; Guo & Chen, 2009). Nevertheless, little is known about gestural representations of motion events in equipollently-framed languages. One study by Brown and Chen (2013) has shown that speakers of Mandarin Chinese frequently encoded manner in speech but did not tend to gestures that highlight manner and instead encode path in gestures (though this study did not focus on the expression of path in speech). However, evidence from co-speech gestures accompanying descriptions of other types of events shows that serial verb constructions tend to co-occur with single rather than multiple gestures (Defina, 2016; Niu et al., 2022). It remains to be seen how the factors discussed above such as type of manner or sensory modality shape motion event descriptions in equipollently-framed across spoken and gestural modalities.

Another direction for future research that would have important implications for application is in the area of translation studies. A number of studies have documented the challenges in translating motion information across typologically different languages (Filipović, 2008; Hijazo-Gascón, 2019; Ibarretxe-Antuñano, 2004; Rojo López & Cifuentes-Férez, 2015; Slobin, 1996). For instance, when translating motion descriptions in satellite-framed languages to verb-framed languages, information about manner tends to get omitted or expressed less elaborately. However, since the available evidence comes from text-based translations, little is known about how such omissions shape expressions in gesture or how the information conveyed in the

gestural modality gets translated. To address these issues, further work on in multimodal translation of motion descriptions in naturalistic settings is required.

Finally, empirical investigations of Talmy's (2000) typology of event integration that considers the multimodal nature of language is primarily in the domain of spontaneous motion events. There is also growing evidence on caused motion events (Furman et al., 2014; Niu et al., 2022). Future research should investigate whether the empirical conclusions drawn based on the current lines of work generalize to other classes of events included in Talmy's event integration framework, such as change of state events and caused motion events, as well as other core schematic features, such as the temporal contours of the event.

6.2. Final conclusions

In conclusion, the present review strongly suggests that Talmy's typology of event integration predicts multimodal event descriptions not only in speech but also in co-speech gesture. This typology also predicts visual attention to event components prior to producing these language-specific event descriptions. However, these influences of language typology may be overridden by variability within the event itself, such as type and modality of stimuli, especially for expression of peripheral event components, such as manner. Together, the empirical evidence reviewed here confirms but also extends Talmy's event integration framework.

Data availability statement. Data sharing is not applicable to this article as no datasets were generated or analyzed during the preparation of this article.

Competing interest. The authors declare no competing interests.

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>
- Aktan-Erciyes, A., Akbuğa, E., Kızılder, E., & Göksun, T. (2022). Motion event representation in L1-Turkish versus L2-English speech and gesture: Relations to eye movements for event components. *International Journal of Bilingualism*, 27, 61–86. <https://doi.org/10.1177/13670069221076838>
- Aktan-Erciyes, A., Göksun, T., Tekcan, A. İ., & Aksu-Koç, A. (2020). Children's thinking-for-speaking. *Linguistic Approaches to Bilingualism*, 11(5), 669–699. <https://doi.org/10.1075/lab.19027.akt>
- Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, 14(3), 257–262. <https://doi.org/10.1016/j.cub.2004.01.029>
- Alcaraz-Carrión, D., Alibali, M. W., & Valenzuela, J. (2022). Adding and subtracting by hand: Metaphorical representations of arithmetic in spontaneous co-speech gestures. *Acta Psychologica*, 228, 103624. <https://doi.org/10.1016/j.actpsy.2022.103624>
- Allen, S., Özyürek, A., Kita, S., Brown, A., Furman, R., Ishizuka, T., & Fujii, M. (2007). Language-specific and universal influences in children's syntactic packaging of manner and path: A comparison of English, Japanese, and Turkish. *Cognition*, 102(1), 16–48. <https://doi.org/10.1016/j.cognition.2005.12.006>
- Bohnemeyer, J., Enfield, N. J., Essegbey, J., Ibarretxe-Antuñano, I., Kita, S., Lüpke, F., & Ameka, F. K. (2007). Principles of event segmentation in language: The case of motion events. *Language*, 83(3), 495–532. <https://doi.org/10.1353/lan.2007.0116>
- Brown, A., & Chen, J. (2013). Construal of manner in speech and gesture in Mandarin, English, and Japanese. *Cognitive Linguistics*, 24(4), 605–631. <https://doi.org/10.1515/cog-2013-0021>
- Brown, A., & Gullberg, M. (2010). Changes in encoding of path of motion in a first language during acquisition of a second language. *Cognitive Linguistics*, 21(2), 263–286. <https://doi.org/10.1515/COGL.2010.010>

- Brown, A., & Gullberg, M. (2011). Bidirectional cross-linguistic influence in event conceptualization? Expressions of Path among Japanese learners of English. *Bilingualism: Language and Cognition*, 14(1), 79–94. <https://doi.org/10.1017/S1366728910000064>
- Brown, L., Kim, H., Hübscher, L., & Winter, B. (2023). Gestures are modulated by social context: A study of multimodal politeness across two cultures. *Gesture*, 21, 167–200. <https://doi.org/10.1075/gest.20034.bro>
- Bunger, A., Skordos, D., Trueswell, J. C., & Papafragou, A. (2021). How children attend to events before speaking: Crosslinguistic evidence from the motion domain. *Glossa: A Journal of General Linguistics*, 6(1), Article 1. <https://doi.org/10.5334/gjgl.1210>
- Bunger, A., Trueswell, J. C., & Papafragou, A. (2012). The relation between event apprehension and utterance formulation in children: Evidence from linguistic omissions. *Cognition*, 122(2), 135–149. <https://doi.org/10.1016/j.cognition.2011.10.002>
- Bylund, E., & Athanasopoulos, P. (2015). Introduction: Cognition, motion events, and SLA. *Modern Language Journal*, 99(S1), 1–13. <https://doi.org/10.1111/j.1540-4781.2015.12175.x>
- Caballero, R. (2007). Manner-of-motion verbs in wine description. *Journal of Pragmatics*, 39(12), 2095–2114. <https://doi.org/10.1016/j.pragma.2007.07.005>
- Caballero, R. (2017). Buildings that move: Motion metaphors in architectural reviews. *Ibérica, Revista de La Asociación Europea de Lenguas Para Fines Específicos*, 34, 89–110.
- Cadierno, T. (2008). Expressing motion events in a second language: A cognitive typological perspective. In M. Achard & S. Niemier (Eds.), *Cognitive linguistics, second language acquisition, and foreign language teaching* (pp. 13–50). De Gruyter Mouton. <https://doi.org/10.1515/9783110199857.13>
- Cánovas, C. P., Valenzuela, J., Carrión, D. A., Olza, I., & Ramscar, M. (2020). Quantifying the speech-gesture relation with massive multimodal datasets: Informativity in time expressions. *PLoS One*, 15(6), e0233892. <https://doi.org/10.1371/journal.pone.0233892>
- Chen, L., & Guo, J. (2009). Motion events in Chinese novels: Evidence for an equipollently-framed language. *Journal of Pragmatics*, 41(9), 1749–1766. <https://doi.org/10.1016/j.pragma.2008.10.015>
- Choi, S., & Lantolf, J. P. (2008). Representation and embodiment of meaning in L2 communication: Motion events in the speech and gesture of advanced L2 Korean and L2 English speakers. *Studies in Second Language Acquisition*, 30(2), 191–224. <https://doi.org/10.1017/S0272263108080315>
- Cifuentes-Férez, P. (2008). *Motion in English and Spanish: A perspective from cognitive linguistics, typology and psycholinguistics*. Universidad de Murcia.
- de Ruiter, J. P. (2000). The production of gesture and speech. In D. McNeill (Ed.), *Language and gesture* (pp. 284–311). Cambridge University Press. <https://doi.org/10.1017/cbo9780511620850.018>
- de Ruiter, J. P. (2007). Postcards from the mind: The relationship between speech, imagistic gesture, and thought. *Gesture*, 7(1), 21–38. <https://doi.org/10.1075/gest.7.1.03rui>
- Defina, R. (2016). Do serial verb constructions describe single events?: A study of co-speech gestures in Avatime. *Language*, 92(4), 890–910. <https://doi.org/10.1353/lan.2016.0076>
- Do, M. L., Papafragou, A., & Trueswell, J. (2020). Cognitive and pragmatic factors in language production: Evidence from source-goal motion events. *Cognition*, 205, 104447. <https://doi.org/10.1016/j.cognition.2020.104447>
- Do, M. L., Papafragou, A., & Trueswell, J. (2022). Encoding motion events during language production: Effects of audience design and conceptual salience. *Cognitive Science*, 46(1), e13077. <https://doi.org/10.1111/cogs.13077>
- Eimer, M. (2004). Multisensory integration: How visual experience shapes spatial perception. *Current Biology*, 14(3), R115–R117. <https://doi.org/10.1016/j.cub.2004.01.018>
- Emerson, S. N., Limia, V. D., & Özçalışkan, Ş. (2021). Cross-linguistic transfer in Turkish–English bilinguals’ descriptions of motion events. *Lingua*, 264, 103153. <https://doi.org/10.1016/j.lingua.2021.103153>
- Filipović, L. (2008). Typology in action: Applying typological insights in the study of translation. *International Journal of Applied Linguistics*, 18(1), 23–40. <https://doi.org/10.1111/j.1473-4192.2008.00189.x>
- Flecken, M., Stutterheim, C. V., & Carroll, M. (2014). Grammatical aspect influences motion event perception: Findings from a cross-linguistic non-verbal recognition task. *Language and Cognition*, 6(1), 45–78. <https://doi.org/10.1017/langcog.2013.2>
- Furman, R., Küntay, A. C., & Özyürek, A. (2014). Early language-specificity of children’s event encoding in speech and gesture: Evidence from caused motion in Turkish. *Language, Cognition and Neuroscience*, 29(5), 620–634. <https://doi.org/10.1080/01690965.2013.824993>

- Gennari, S. P., Sloman, S. A., Malt, B. C., & Fitch, W. T. (2002). Motion events in language and cognition. *Cognition*, 83(1), 49–79. [https://doi.org/10.1016/S0010-0277\(01\)00166-4](https://doi.org/10.1016/S0010-0277(01)00166-4)
- Gleitman, L. R., January, D., Nappa, R., & Trueswell, J. C. (2007). On the give and take between event apprehension and utterance formulation. *Journal of Memory and Language*, 57(4), 544–569. <https://doi.org/10.1016/j.jml.2007.01.007>
- Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, 11(4), 274–279. <https://doi.org/10.1111/1467-9280.00255>
- Grigoroglou, M., & Papafragou, A. (2019). Children’s (and adults’) production adjustments to generic and particular listener needs. *Cognitive Science*, 43(10), e12790. <https://doi.org/10.1111/cogs.12790>
- Gullberg, M., Hendriks, H., & Hickmann, M. (2008). Learning to talk and gesture about motion in French. *First Language*, 28(2), 200–236. <https://doi.org/10.1177/0142723707088074>
- Guo, J., & Chen, L. (2009). Learning to express motion in narratives by Mandarin-speaking children. In J. Guo, E. Lieven, N. Budwig, S. Ervin-Tripp, K. Nakamura, & Ş. Özçalışkan (Eds.), *Crosslinguistic approaches to the psychology of language: Research in the tradition of Dan Isaac Slobin* (pp. 193–208). Psychology Press.
- Hijazo-Gascón, A. (2019). Translating accurately or sounding natural?: The interpreters’ challenges due to semantic typology and the interpreting process. *Pragmatics and Society*, 10(1), 72–94. <https://doi.org/10.1075/ps.00016.hij>
- Hohenstein, J., Eisenberg, A., & Naigles, L. (2006). Is he floating across or crossing afloat? Cross-influence of L1 and L2 in Spanish–English bilingual adults. *Bilingualism: Language and Cognition*, 9(3), 249–261. <https://doi.org/10.1017/S1366728906002616>
- Holler, J., & Levinson, S. C. (2019). Multimodal language processing in human communication. *Trends in Cognitive Sciences*, 23(8), 639–652. <https://doi.org/10.1016/j.tics.2019.05.006>
- Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, 15(3), 495–514. <https://doi.org/10.3758/PBR.15.3.495>
- Hostetter, A. B., & Alibali, M. W. (2019). Gesture as simulated action: Revisiting the framework. *Psychonomic Bulletin & Review*, 26(3), 721–752. <https://doi.org/10.3758/s13423-018-1548-0>
- Ibarretxe-Antuñano, I. (2004). Language typologies in our language use: The case of Basque motion events in adult oral narratives. *Cognitive Linguistics*, 15(3), 317–349. <https://doi.org/10.1515/cogl.2004.012>
- Ibarretxe-Antuñano, I. (2017). *Motion and space across languages: Theory and applications*. John Benjamins Publishing Company. <https://doi.org/10.1075/hcp.59>
- Johnson, M. L., & Larson, S. (2003). “Something in the way she moves”–Metaphors of musical motion. *Metaphor and Symbol*, 18(2), 63–84. https://doi.org/10.1207/S15327868MS1802_1
- Karadöller, D. Z., Sümer, B., Ünal, E., & Özyürek, A. (2021). Spatial language use predicts spatial memory of children: Evidence from sign, speech, and speech-plus-gesture. In W. T. Fitch, C. Lamm, H. Leder, & K. Tessmar (Eds.), *Proceedings of the 43rd Annual Conference of the Cognitive Science Society* (pp. 672–678). Cognitive Science Society.
- Karadöller, D. Z., Sümer, B., Ünal, E., & Özyürek, A. (2022). Late sign language exposure does not modulate the relation between spatial language and spatial memory in deaf children and adults. *Memory & Cognition*, 51(3), 582–600. <https://doi.org/10.3758/s13421-022-01281-7>
- Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press.
- Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and gesture* (pp. 162–185). Cambridge University Press.
- Kita, S., Alibali, M. W., & Chu, M. (2017). How do gestures influence thinking and speaking? The gesture-for-conceptualization hypothesis. *Psychological Review*, 124(3), 245–266. <https://doi.org/10.1037/rev0000059>
- 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](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., & Meyer, A. S. (2014). Priming sentence planning. *Cognitive Psychology*, 73, 1–40. <https://doi.org/10.1016/j.cogpsych.2014.04.001>
- Krauss, R. M., Chen, Y., & Gottesman, R. F. (2000). Lexical gestures and lexical access: A process model. In D. McNeill (Ed.), *Language and gesture* (pp. 261–283). Cambridge University Press.
- Lakoff, G., & Johnson, M. (2008). *Metaphors we live by*. University of Chicago Press.

- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. MIT Press.
- Mamus, E., Speed, L. J., Özyürek, A., & Majid, A. (2022). The effect of input sensory modality on the multimodal encoding of motion events. *Language, Cognition and Neuroscience*, 38(5), 711–723. <https://doi.org/10.1080/23273798.2022.2141282>
- Mamus, E., Speed, L. J., Rissman, L., Majid, A., & Özyürek, A. (2023). Lack of visual experience affects multimodal language production: Evidence from congenitally blind and sighted people. *Cognitive Science*, 47(1), e13228. <https://doi.org/10.1111/cogs.13228>
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago Press.
- McNeill, D. (2005). *Gesture and thought*. University of Chicago Press.
- McNeill, D., & Duncan, S. D. (2000). Growth points in thinking for speaking. In D. McNeill (Ed.), *Language and gesture* (pp. 141–161). Cambridge University Press.
- Meyer, A. S., Sleiderink, A. M., & Levelt, W. J. M. (1998). Viewing and naming objects: Eye movements during noun phrase production. *Cognition*, 66(2), B25–B33. [https://doi.org/10.1016/S0010-0277\(98\)00009-2](https://doi.org/10.1016/S0010-0277(98)00009-2)
- Naigles, L. R., Eisenberg, A. R., Kako, E. T., Hightner, M., & McGraw, N. (1998). Speaking of motion: Verb use in English and Spanish. *Language and Cognitive Processes*, 13(5), 521–549. <https://doi.org/10.1080/016909698386429>
- Niu, C., Cienki, A., Ortega, G., & Coene, M. (2022). Learning to express causal events in Mandarin Chinese: A multimodal perspective. *Journal of Child Language*, 1–26. <https://doi.org/10.1017/S0305000922000447>
- Özçalışkan, Ş. (2003). Metaphorical motion in crosslinguistic perspective: A comparison of English and Turkish. *Metaphor and Symbol*, 18(3), 189–228. https://doi.org/10.1207/S15327868MS1803_05
- Özçalışkan, Ş. (2004). Typological variation in encoding the manner, path, and ground components of a metaphorical motion event. *Annual Review of Cognitive Linguistics*, 2(1), 73–102. <https://doi.org/10.1075/arcl.2.03ozc>
- Özçalışkan, Ş. (2005). Metaphor meets typology: Ways of moving metaphorically in English and Turkish. *Cognitive Linguistics*, 16(1), 207–246. <https://doi.org/10.1515/cogl.2005.16.1.207>
- Özçalışkan, Ş. (2007). Metaphors we move by: Children’s developing understanding of metaphorical motion in typologically distinct languages. *Metaphor and Symbol*, 22(2), 147–168. <https://doi.org/10.1080/10926480701235429>
- Özçalışkan, Ş. (2015). Ways of crossing a spatial boundary in typologically distinct languages. *Applied Psycholinguistics*, 36(2), 485–508. <https://doi.org/10.1017/S0142716413000325>
- Özçalışkan, Ş., Lucero, C., & Goldin-Meadow, S. (2016a). Does language shape silent gesture? *Cognition*, 148, 10–18. <https://doi.org/10.1016/j.cognition.2015.12.001>
- Özçalışkan, Ş., Lucero, C., & Goldin-Meadow, S. (2016b). Is seeing gesture necessary to gesture like a native speaker? *Psychological Science*, 27(5), 737–747. <https://doi.org/10.1177/0956797616629931>
- Özçalışkan, Ş., Lucero, C., & Goldin-Meadow, S. (2018). Blind speakers show language-specific patterns in co-speech gesture but not silent gesture. *Cognitive Science*, 42(3), 1001–1014. <https://doi.org/10.1111/cogs.12502>
- Özçalışkan, Ş., Lucero, C., & Goldin-Meadow, S. (2023). What the development of gesture with and without speech can tell us about the effect of language on thought. *Language and Cognition*, 1–22. <https://doi.org/10.1017/langcog.2023.34>
- Özçalışkan, Ş., & Slobin, D. I. (2003). Codability effects on the expression of manner of motion in English and Turkish. In A. S. Özsoy, M. Nakipoglu-Demiralp, E. Erguvanli-Taylan & A. Aksu-Koç (Eds.), *Studies in Turkish Linguistics* (pp. 259–270). Bogazici University Press.
- Özyürek, A. (2017). Function and processing of gesture in the context of language. In R. B. Church, M. W. Alibali, & S. D. Kelly (Eds.), *How the hands function in speaking, thinking and communicating* (pp. 39–58). John Benjamins Publishing.
- 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. <https://doi.org/10.1075/gest.5.1-2.15ozy>
- Papafrağou, A., Hulbert, J., & Trueswell, J. (2008). Does language guide event perception? Evidence from eye movements. *Cognition*, 108(1), 155–184. <https://doi.org/10.1016/j.cognition.2008.02.007>

- Papafragou, A., Massey, C., & Gleitman, L. (2002). Shake, rattle, n'roll: The representation of motion in language and cognition. *Cognition*, 84(2), 189–219. [https://doi.org/10.1016/S0010-0277\(02\)00046-X](https://doi.org/10.1016/S0010-0277(02)00046-X)
- Papafragou, A., Massey, C., & Gleitman, L. (2006). When English proposes what Greek presupposes: The cross-linguistic encoding of motion events. *Cognition*, 98(3), B75–B87. <https://doi.org/10.1016/j.cognition.2005.05.005>
- Perniss, P. (2018). Why we should study multimodal language. *Frontiers in Psychology*, 9, 1109. <https://doi.org/10.3389/fpsyg.2018.01109>
- Rojo López, A., & Cifuentes-Férez, P. (2015). On the reception of translations: Exploring the impact of typological differences on legal contexts. In I. Ibarretxe-Antuñano (Ed.), *Motion and space across languages: Theory and applications* (pp. 367–398). John Benjamins Publishing Company. <https://doi.org/10.13140/RG.2.1.3253.6486>
- Sakarias, M., & Flecken, M. (2019). Keeping the result in sight and mind: General cognitive principles and language-specific influences in the perception and memory of resultative events. *Cognitive Science*, 43(1), 12708. <https://doi.org/10.1111/cogs.12708>
- Skordos, D., Bunger, A., Richards, C., Selimis, S., Trueswell, J., & Papafragou, A. (2020). Motion verbs and memory for motion events. *Cognitive Neuropsychology*, 37(5–6), 254–270. <https://doi.org/10.1080/02643294.2019.1685480>
- Slobin, D. I. (1996). From “thought and language” to “thinking for speaking.” In J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 57–149). Cambridge University Press.
- Slobin, D. I. (2003). Language and thought online: Cognitive consequences of linguistic relativity. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind* (pp. 157–191). MIT Press.
- Slobin, D. I. (2004). The many ways to search for a frog: Linguistic typology and the expression of motion events. In S. Strömquist & L. Verhoeven (Eds.), *Relating events in narrative: Vol. 2. Typological and contextual perspectives* (Vol. 2, pp. 219–257). Lawrence Erlbaum Associates.
- Slobin, D. I. (2006). What makes manner of motion salient. In M. Hickmann & S. Robert (Eds.), *Space in languages* (pp. 59–81). John Benjamins Publishing Company. <https://doi.org/10.1075/tsl.66.05slo>
- Slobin, D. I. (2008). Relations between paths of motion and paths of vision: A crosslinguistic and developmental exploration. In V. C. M. Gathercole (Ed.), *Routes to language: Studies in honor of Melissa Bowerman* (pp. 197–221). Psychology Press.
- Slobin, D. I., Ibarretxe-Antuñano, I., Kopecka, A., & Majid, A. (2014). Manners of human gait: A cross-linguistic event-naming study. *Cognitive Linguistics*, 25(4), 701–741. <https://doi.org/10.1515/cog-2014-0061>
- Soroli, E., Sahraoui, H., & Sacchett, C. (2012). Linguistic encoding of motion events in English and French: Typological constraints on second language acquisition and agrammatic aphasia. *Language, Interaction and Acquisition*, 3(2), 261–287. <https://doi.org/10.1075/lia.3.2.05sor>
- Sümer, B., & Özyürek, A. (2022). Cross-modal investigation of event component omissions in language development: A comparison of signing and speaking children. *Language, Cognition and Neuroscience*, 37(8), 1023–1039. <https://doi.org/10.1080/23273798.2022.2042336>
- Talmy, L. (1975). Semantics and syntax of motion. In J. Kimball (Ed.), *Syntax and semantics* (Vol. 1–4, pp. 181–238). Academic Press. https://doi.org/10.1163/9789004368828_008
- Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Grammatical categories and the lexicon* (pp. 57–149). Cambridge University Press.
- Talmy, L. (2000). A typology of event intergration. In *Toward a cognitive semantics: Typology and process in concept structuring* (Vol. 2, pp. 213–288). MIT Press.
- Talmy, L. (2003). The representation of spatial structure in spoken and signed language. In K. Emmorey (Ed.), *Perspectives on classifier constructions in sign languages* (pp. 207–238). Lawrence Erlbaum Associates.
- ter Bekke, M., Özyürek, A., & Ünal, E. (2022). Speaking but not gesturing predicts event memory: A cross-linguistic comparison. *Language and Cognition*, 14(3), 362–384. <https://doi.org/10.1017/langcog.2022.3>
- Thinus-Blanc, C., & Gaunet, F. (1997). Representation of space in blind persons: Vision as a spatial sense? *Psychological Bulletin*, 121(1), 20–42. <https://doi.org/10.1037/0033-2909.121.1.20>
- Trueswell, J. C., & Papafragou, A. (2010). Perceiving and remembering events cross-linguistically: Evidence from dual-task paradigms. *Journal of Memory and Language*, 63(1), 64–82. <https://doi.org/10.1016/j.jml.2010.02.006>

- Ünal, E., Manhardt, F., & Özyürek, A. (2022). Speaking and gesturing guide event perception during message conceptualization: Evidence from eye movements. *Cognition*, 225, 105127. <https://doi.org/10.1016/j.cognition.2022.105127>
- Ünal, E., Richards, C., Trueswell, J. C., & Papafragou, A. (2021). Representing agents, patients, goals and instruments in causative events: A cross-linguistic investigation of early language and cognition. *Developmental Science*, 24(6), e13116. <https://doi.org/10.1111/desc.13116>
- van de Velde, M., Meyer, A. S., & Konopka, A. E. (2014). Message formulation and structural assembly: Describing “easy” and “hard” events with preferred and dispreferred syntactic structures. *Journal of Memory and Language*, 71(1), 124–144. <https://doi.org/10.1016/j.jml.2013.11.001>
- Zlatev, J., David, C., & Blomberg, J. (2010). Translocation, language and the categorization of experience. In V. Evans & P. Chilton (Eds.), *Language, cognition and space: The state of the art and new directions* (pp. 389–418). Equinox.

Cite this article: Ünal, E., Mamus, E., & Özyürek, A. (2024). Multimodal encoding of motion events in speech, gesture and cognition, *Language and Cognition* 16: 785–804. <https://doi.org/10.1017/langcog.2023.61>