Cognition 225 (2022) 105127

Contents lists available at ScienceDirect

Cognition

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

Speaking and gesturing guide event perception during message conceptualization: Evidence from eye movements

Ercenur Ünal [a,b,c,*], Francie Manhardt [b,c], Aslı Özyürek [b,c,d,**]

[a] Department of Psychology, Ozyegin Universiy, Nişantepe Mahallesi Orman Sokak, 34794, Çekmeköy, Istanbul, Turkey
[b] Centre for Language Studies, Radboud University, Erasmusplein 1, 6525, HT, Nijmegen, the Netherlands
[c] Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525, XD, Nijmegen, the Netherlands
[d] Donders Institute for Brain, Cognition and Behaviour, Heyendaalseweg, 135 6525, AJ, Nijmegen, the Netherlands

ARTICLE INFO

Keywords:
Motion events
Gesture
Multimodal language production
Thinking for speaking
Visual attention

ABSTRACT

Speakers' visual attention to events is guided by linguistic conceptualization of information in spoken language production and in language-specific ways. Does production of language-specific co-speech gestures further guide speakers' visual attention during message preparation? Here, we examine the link between visual attention and multimodal event descriptions in Turkish. Turkish is a verb-framed language where speakers' speech and gesture show language specificity with path of motion mostly expressed within the main verb accompanied by path gestures. Turkish-speaking adults viewed motion events while their eye movements were recorded during non-linguistic (viewing-only) and linguistic (viewing-before-describing) tasks. The relative attention allocated to path over manner was higher in the linguistic task compared to the non-linguistic task. Furthermore, the relative attention allocated to path over manner within the linguistic task was higher when speakers (a) encoded path in the main verb versus outside the verb and (b) used additional path gestures accompanying speech versus not. Results strongly suggest that speakers' visual attention is guided by language-specific event encoding not only in speech but also in gesture. This provides evidence consistent with models that propose integration of speech and gesture at the conceptualization level of language production and suggests that the links between the eye and the mouth may be extended to the eye and the hand.

1. Introduction

The idea that linguistic conceptualization of events builds on and guides apprehension of events is prominent in influential theories of language production. According to the thinking for speaking hypothesis (Slobin, 1996), speakers attend to the aspects of the world that they plan to speak about and in ways compatible with the lexical and syntactic constraints of their specific language. Similarly, in Levelt's (1989) language production model, speaking begins with a preverbal apprehension of the broad details of an event, including information about people, objects, places, time and the relations among them. This preverbal event representation is mapped onto a linguistic message taking into account the constraints on how entities, relations and spatiotemporal information are packaged into different lexical and syntactic structures, which ultimately culminates into an utterance. This model is supported by eye-tracking studies showing that while describing visual scenes, speakers allocate their attention to the features they plan to speak about (Gleitman, January, Nappa, & Trueswell, 2007; Griffin & Bock, 2000; Konopka & Meyer, 2014; Meyer, Sleiderink, & Levelt, 1998; van de Velde, Meyer, & Konopka, 2014) and in a way reflecting language-specific semantic and grammatical patterns (Norcliffe, Konopka, Brown, & Levinson, 2015; Sauppe, 2017; Sauppe, Norcliffe, Konopka, Van Valin, & Levinson, 2013; for an overview see Norcliffe, Harris, & Jaeger, 2015). Nevertheless, language is a multimodal phenomenon. Speakers frequently use gestures along with speech to convey information about events (McNeill, 2005). Furthermore, they do so in ways linked to language-specific encoding of events in speech (Kita & Özyürek, 2003). The purpose of the present study is to draw on evidence from the eye-tracking paradigm to test whether producing language-specific gestures along with speech further guides visual attention to events.

There are different views on how gesture production is linked to

* Corresponding author at: Ozyegin University, Nişantepe Mahallesi Orman Sokak, 34794, Çekmeköy, Istanbul, Turkey.
** Corresponding author at: Centre for Language Studies, Radboud University, Erasmusplein 1, 6525 HT, Nijmegen, the Netherlands.
E-mail addresses: ercenur.unal@ozyegin.edu.tr (E. Ünal), Francie.Manhardt@ru.nl (F. Manhardt), asli.ozyurek@ru.nl (A. Özyürek).

https://doi.org/10.1016/j.cognition.2022.105127
Received 10 June 2021; Received in revised form 24 March 2022; Accepted 11 April 2022
Available online 23 May 2022

(spoken) language production. One class of models propose that gestures are pre-linguistically generated from the visual imagery in visuo-spatial working memory and thus function as a direct window into thought (Krauss, Chen, & Chawla, 1996; Krauss, Chen, & Gottesman, 2000). A second class of models propose that gestures are generated from the speaker's communicative intent about what information they want to convey (de Ruiter, 2000, 2007; Melinger & Levelt, 2004). Part of this information is communicated via speech and part of it is conveyed via gesture. The information conveyed in speech and gesture may or may not overlap. Crucially, both classes of models propose that gesture production is not part of message preparation and therefore, the content and the form of gestures should not be influenced by language-specific constraints on how information is expressed in the accompanying speech. Therefore, these models cannot explain how gestures produced along with speech show language-specific patterns and how speech and gesture are linked systems.

Unlike these two classes of models, the Interface Model of multimodal language production (Kita & Özyürek, 2003) proposes that gesture is also planned during message preparation for language production. In this view, linguistic conceptualization interacts with the spatio-motoric imagery underlying gesture generation during online language production. Through these interactions, co-speech gestures represent information following language-specific constraints on information packaging in the speech that they accompany. Each co-speech gesture expresses semantic information encoded within one processing unit (i.e., verbal clause) in speech. The Interface Model uniquely predicts gesture production to require conceptualization as it is for speech production. If so, one might expect similar links between co-speech gesture production and event apprehension as found for speech production. One of the novelties of the present study is to test this prediction.

1.1. Multimodal linguistic encoding and conceptualization of motion events

Motion events provide an ideal domain for investigating how event apprehension during message preparation might be linked to speech and gesture production, and how these links might be indexed through visual attention. There is considerable cross-linguistic diversity in how languages map motion event components onto lexical or syntactic structures (Talmy, 1985). This diversity provides the grounds for looking for potential differences and language specificity in event conceptualization. Satellite-framed languages (e.g., English, Dutch) typically encode *manner* of motion in the main verb (e.g., "ran" sentence 1) and *path* of motion in elements outside of the verb (e.g., in pre-positional phrases, which are formed by adding a preposition "into" before a noun phrase as in "into the phone booth" in sentence 1). Verb-framed languages (e.g., Turkish, Greek), however, typically encode path of motion in the main verb (e.g., "girdi" "entered" in sentence 2), and manner of motion in subordinate verbs (e.g., "kosarak" "while running" in sentence 2). In verb-framed languages manner is optional and can be omitted.

(1)	The woman	**ran**	into	the phone booth
	Noun phrase	Verb	Preposition	Noun phrase
	Figure	Manner	Path	Ground

(2)	Kadın	(koş-arak)	telefon kulübesi-(n)e	**gir-di**
	woman	(run-CONN)	phone booth-DAT	enter-PST
	Noun phrase	(Verb)	Noun phrase	Verb
	Figure	(Manner)	Ground	Path
	'The woman entered the phone booth while running'			

It should be noted that these patterns are not the *only* ways in which motion events are encoded in these languages but rather indicate the most frequent and typical ways of encoding motion. Speakers of verb-framed languages occasionally encode path in elements outside of the verb, such as post-positional phrases either together with path verbs or as the sole expression of path information (Özçalışkan & Slobin, 2003). For example, in Turkish path of motion can be encoded occasionally without the main path verb but only through post-positional phrases, which are formed by adding a postposition (e.g., "içi-(n)e" "to-inside") after a noun phrase (e.g., "telefon kulübesi-nin içi-(n)e" "to inside the phone booth" in sentence 3).

(3)	Kadın	telefon kulübesi-nin	**içi-(n)e**	koş-tu
	woman	phone booth-GEN	inside-DAT	run-PST
	Noun phrase	Noun phrase	Postposition	Verb
	Figure	Ground	Path	Manner
	'The woman ran to inside the phone booth'.			

Furthermore, there are systematic cross-linguistic differences in co-speech gestures that depict path and manner of motion. Crucially, these cross-linguistic differences in gesture show striking parallels to the typological patterns in how path and manner information is encoded in speech in these languages. For example, English-speakers are likely to use one clause to express path and manner in speech (i.e., manner as main verb and path as pre-positional phrase; see sentence 1 above) and typically conflate path and manner components into a single co-speech gesture (Kita & Özyürek, 2003). In contrast, Turkish- and Japanese-speakers who are likely to use separate clauses to express path and manner in speech (i.e., path as main verb and manner as subordinate verb; see sentence 2 above), typically produce separate gestures for manner and path. They also tend to produce more path-only than manner-only gestures because path is encoded in the main verb (Kita & Özyürek, 2003, see also Kita et al., 2007; Özçalışkan, 2016; Özçalışkan, Lucero, & Goldin-Meadow, 2016a, 2016b; Özyürek, Kita, Allen, Furman, & Brown, 2005, Özyürek et al., 2008 and Gullberg, 2011 for converging evidence from the domain of placement events). Importantly, these cross-linguistic differences in speech and gesture surface in the descriptions of the very same events. Thus, despite the fact that the visual imagery for the events is the same, gesture patterns differ cross-linguistically, mirroring patterns found in speech for satellite- and verb-framed languages. These findings challenge the view that gestures are not part of message preparation (de Ruiter, 2000, 2007; Krauss et al., 1996, 2000). Instead, they can be taken as evidence for the Interface Model of speech and gesture production (Kita & Özyürek, 2003) according to which gesture production is constrained by the kind of semantic information that can be syntactically packaged in one processing unit (i.e., clause) in speech (Levelt, 1989).

Finally, there has been evidence for a tight link between event conceptualization and language-specific motion event encoding in speech from cross-linguistic eye-tracking studies. In one study, English- and Greek-speaking adults watched videos of motion events (e.g., a man skating to a snowman) while their eye movements were recorded and then described the events (Papafragou, Hulbert, & Trueswell, 2008). While viewing the events prior to speaking, both groups allocated more attention to the component that they were planning to encode in the main verb. Greek-speakers attended more to path of motion than English-speakers, and English-speakers attended more to manner of motion than Greek-speakers. Crucially, these cross-linguistic differences in attention allocation that emerged prior to speaking disappeared when participants freely inspected the events without preparing for linguistic encoding (see also Bunger, Skordos, Trueswell, & Papafragou, 2016, 2021; Bunger, Trueswell, & Papafragou, 2012; Flecken, von Stutterheim, & Carroll, 2014; Sakarias & Flecken, 2019; Trueswell & Papafragou, 2010). These findings strongly suggest tight links between visual attention and language-specific encoding of motion that emerge during speech planning.

Even though the current body of evidence on the tight link between speech and gesture production, and particularly language-specificity of gestures accompanying speech, is best explained by the Interface Model, there is one aspect of this model that remains to be tested empirically.

The Interface Model attributes language-specificity of co-speech gestures to interactions between event apprehension, linguistic conceptualization, and visual-spatial imagery during message preparation stage of multimodal language production. However, empirical evidence for this aspect of the model has been somewhat indirect because it comes from speech and gesture behavior and not from eye-gaze behavior during event apprehension. This is because all of the prior work on eye-gaze patterns during message preparation has focused on (spoken) language production in the auditory modality (with the exception of one study on sign language production; Manhardt et al., 2020). In addition to the modulation in eye-gaze patterns driven by speech production, there might be further modulation of visual attention driven by the additional language-specific encoding of event information in the gestural modality. Gestures produced with speech can encode information overlapping with speech as well as additional information not necessarily found in speech. For example, speakers may express path of motion in speech by saying "the woman entered the phone booth" and convey additional information about direction of motion by producing a co-speech gesture that directly maps onto the visual scene (e.g., moving index finger across space from left to right). Whether such language-specific encodings in co-speech gesture further guide visual attention remains to be seen.

1.2. Present study

In the present study, our primary goal is to seek empirical evidence from eye-gaze behavior for the integration of the speech and gesture during message conceptualization. As a secondary goal, we aim to replicate and extend prior evidence on the relation between visual attention and language-specific encoding of motion events in speech. To address these goals, we ask how speakers of Turkish (a verb-framed language) attend to path as opposed to manner of motion events while viewing the events in preparation for linguistic encoding in speech and gesture (i.e., linguistic task) versus freely inspecting events without preparing for any linguistic encoding (i.e., non-linguistic task). For linguistic encoding, we investigate how path and manner is encoded in speech and gesture and whether this is in line with previously shown typological patterns (Talmy, 1985, Slobin, 1996). For visual attention, we ask if attention allocated to path relative to manner varies (a) in relation to variations in linguistic encoding of path of motion in speech and (b) in relation to path gesture production accompanying path encodings in speech.

In language production, Turkish-speakers are expected to encode path in the main verb and manner outside of the main verb. In line with this typological encoding they might be expected to produce descriptions that encode only path of motion in speech because of optional encoding of manner (i.e., the element encoded outside of the main verb) in verb-framed languages. Alternatively, they may produce spoken descriptions that encode both path and manner of motion as seen in previous work with Turkish-speakers, when both path and manner are salient in an event (Özyürek et al., 2008). However, as more relevant for the aims of this study, the majority of path encodings in speech are expected to be within the main verb as opposed to outside of the verb – albeit with some variation.

In gesture, participants are expected to produce descriptions that encode only path of motion more frequently because semantic elements encoded in the main verb (i.e., path in this case) are more likely to determine the content of the gestures than elements encoded outside of the main verb – (Kita & Özyürek, 2003; Özyürek, 2018; Özyürek et al., 2005). Since Turkish has a verb-framed typology, speakers should be more likely to produce separate co-speech gestures for path and manner. Furthermore, as seen in previous work, Turkish-speakers should be more likely to leave out the gesture component corresponding to the element encoded outside of the main verb - in this case, manner of motion (Özçalışkan, 2016; Özçalışkan et al., 2016a, 2016b).

For eye movements, we expect the time course of visual attention to differ across linguistic and non-linguistic tasks, as found in speakers of other verb-framed languages that encode motion similarly to Turkish (e.g., Greek: Papafragou et al., 2008; Trueswell & Papafragou, 2010). Specifically, speakers should allocate more attention to path over manner of motion in the linguistic task than they do in the non-linguistic task. This is based on the fact that path is likely to be encoded in the main verb in such languages whereas manner is encoded outside of the main verb or omitted.

To be able to show more direct links between the specific choices in actual language use and the time course of visual attention within a single group of language users, we expect speakers to allocate more attention to path over manner of motion in the linguistic task when they encode path in speech compared to when they do not encode it in speech (i.e., mention only manner in their description). We also expect participants to allocate more attention to path over manner of motion when they encode path in the main verb as opposed to when they encode path outside of the verb. These predictions are based on the proposal that verbs are the main unit of sentence planning (Levelt, 1989, among others) as well as the thinking for speaking hypothesis (Slobin, 1996) which proposes that language-specific encodings in the verbs guide attention prior to speaking. Previous cross-linguistic eye-tracking studies have only compared speakers of satellite- and verb-framed languages at the group level without considering within-language variation in motion event encoding. Thus, the current study goes beyond previous studies in testing whether visual event apprehension varies in relation to different syntactic encodings of event components within a single language. Such evidence can illuminate what kind of linguistic encoding (i.e., path within the main verb versus outside of the verb) influences event apprehension in a more fine-grained way.

For eye movements regarding gesture production in the linguistic task, we expect even more attention allocated to path over manner of motion when path is encoded in both speech and gesture compared to when path is encoded only in speech. Unlike other models of gesture production (de Ruiter, 2000, 2007; Krauss et al., 1996, 2000), the Interface Model uniquely predicts gesture production to require similar conceptualization as speech production and to be constrained by the kind of semantic information can be packaged in one processing unit within the main verb (Kita & Özyürek, 2003). In the case of Turkish, as path is more likely to be encoded in the main verb, path gestures are expected to be produced with similar conceptualization of events as in speech production. Furthermore, path gestures might provide extra information about the direction of motion in the left-right axis not found in speech, which might then lead to more attention allocated to path of motion in the visual event.

2. Method

The methods reported in this study were approved by the Humanities Ethics Assessment Committee of the Radboud University.

2.1. Participants

Data were collected from adult native speakers of Turkish ($n = 36$, 10 males, mean age = 21.5 years, range = 19–24). All of the participants had learned Turkish from birth on and as their first language. Participants were students at Ozyegin University in Istanbul, Turkey and received course credit for their participation. Data from six additional participants were discarded due to trackloss ($n = 1$), the participant having amblyopia ($n = 1$), failing to follow the instructions in the linguistic task ($n = 3$) and equipment error ($n = 1$).

Sample size was determined based on previous eye-tracking work on the relation between event apprehension and utterance production. Cross-linguistic studies included 10 to 25 participants per task in each language group (Bunger et al., 2016; Flecken et al., 2014; Papafragou et al., 2008; Sakarias & Flecken, 2019; Trueswell & Papafragou, 2010) and single language studies used 36 participants (e.g., Gleitman et al.,

2007).

2.2. Stimuli

Stimuli consisted of short video clips depicting two types of events: motion events (target stimuli) featuring intransitive events of an agent moving in relation to a landmark, and transitive events of an agent performing actions on objects (fillers). One of two different female actors performed each event. All stimuli are available at https://osf.io/st5gb/.

2.2.1. Motion events

Fifty video clips that depicted a female actor moving with respect to a landmark object along a particular path with a particular manner served as the stimuli for motion events. Each video clip was 2500 ms long. Motion lasted throughout the entire 2500 ms.

The stimuli included five different spontaneous manners of motion, corresponding to: *walk, run, leap, skip,* and *hop* (*yürümek, koşmak, sıçramak, hoplamak, sekmek* in Turkish) and five different paths of motion, corresponding to: *to, past, into, from* and *out of* (*yaklaşmak, geçmek, girmek, uzaklaşmak, çıkmak* in Turkish). The complete set of stimuli included an equal number of path and manner variations. Manners of motion were filmed in a studio at Radboud University for the purpose of this study. Each actor performed each manner of motion against a green background. The video clips were edited in Adobe Premiere Pro CC 2015. First, each clip was cut to last 2500 ms. Then, the background of the video was made transparent using the ultra key feature of Adobe Premiere Pro. In order to create a scene, each manner of motion was combined with one of two different backgrounds (a white brick wall, or a light pink wall) and a gray asphalt-textured floor. Finally, motion paths were created by combining the moving figure with a landmark object (Fig. 1). For *to* and *into* paths, the landmark objects were placed near the final location of the actor's motion. For *from* and *out of* paths, the landmark objects were placed near the starting location. For *past* paths, the landmark objects were placed towards the final location of the motion, but such that the actor would pass the object during the video. Some landmark objects appeared twice, with a different token for each actor.

Previous eye-tracking work in the domain of motion events has revealed that speakers extract path of motion from similar events by predictively fixating on a goal object (Bunger et al., 2012, 2016, 2021; Papafragou et al., 2008; Trueswell & Papafragou, 2010) rather than tracing the trajectory of motion with their eyes since people rarely fixate on empty space (see also Kamide, Lindsay, Scheepers, & Kukona, 2016). Thus, the events that involved a goal directed path (*to, into,* or *past*) served as the target motion events. In order to ensure that participants did not only view goal directed motion, events that involved source paths (*from* or *out of*) were included as non-target motion events.

Each path-manner combination had two versions, performed by a different actor, creating a total of 50 motion events (see Appendix A for the complete list of motion events). These 50 events were divided into two lists, such that each manner-path combination appeared once in each list. The assignment of lists to tasks (non-linguistic, linguistic) was counterbalanced across participants. For each event list, an additional version was created by reversing the order of items. Thus, for each task (non-linguistic, linguistic) there were four presentation lists. Furthermore, the two actors, the two backgrounds (white, pink) and the direction of motion in the video (left-right, right-left) were counterbalanced across lists, and across each manner and each path within the lists.

2.2.2. Transitive filler events

Fifty additional video clips that depicted the same female actors performing every-day actions on objects (e.g., peeling a banana) served as the stimuli for transitive filler events (25 videos per actor). Transitive filler events were 2500 ms long (see Appendix B for a complete list of transitive filler events). Transitive filler events were also filmed at the same studio at Radboud University for the purpose of this study. Actors performed the actions on a gray table against a green background. Video clips were edited in Adobe Premiere Pro CC 2015. First, each clip was cut to last 2500 ms. Then, the green background was removed and replaced with one of the two backgrounds that were used for the motion events. The 50 transitive filler events were divided into two lists: one for the non-linguistic eye-tracking task, and one for the linguistic eye-tracking task. All participants saw the first set during the non-linguistic task and the second set during the linguistic task.

2.3. Procedure

Each participant was tested in a quiet room in their university campus in Turkish by a native speaker. Participants first signed a consent form. Then, they were seated approximately 60 cm away from a SMI RED 250 eye-tracker (SensoMotoric Instruments) attached to a DELL Precision M4800 laptop. Eye gaze was sampled (binocular) at a rate of 250 Hz (every 4 ms). Screen resolution was 1920 × 1080. The size of the stimulus videos was 1280 (width) mm X 720 (height) mm. The experiments were run through Presentation® software (Version 16.5, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com). All participants completed the three components of the experiment in the following order: (1) Non-linguistic eye-tracking task, (2) Filler task, (3) Linguistic eye-tracking task. At the end of the session, all participants completed a demographics and language background survey (Gullberg & Indefrey, 2003) and a post-experiment questionnaire. The non-linguistic task was presented first to avoid carry-over effects from the linguistic eye-tracking task onto the non-linguistic eye-tracking task. The filler task was included to distract the participant's attention from the previous task and lasted approximately 5 min. Each session lasted approximately 45 min.

2.3.1. Non-linguistic eye-tracking task

Participants watched 50 video clips of events (25 motion events and 25 transitive filler events) presented on the computer screen while their

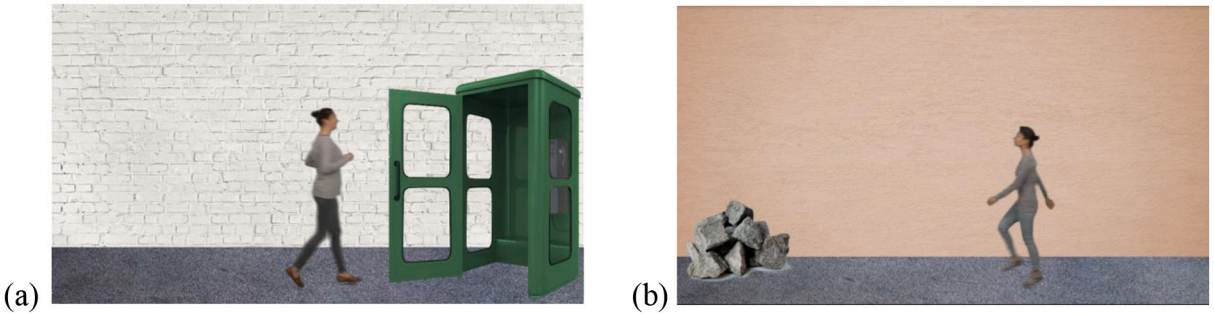

Fig. 1. Sample motion event stimuli: (a) "A woman running into a phone booth" (b) "A woman skipping to rocks".

eye movements were recorded. In each trial, participants first saw a fixation screen, containing a fixation cross in the center, for 1000 ms. Next, an event was shown for 2500 ms. Finally, a gray screen was presented until the participant clicked on a blue button on the mouse to advance to the next trial. In order to ensure attention to the screen, participants were given a secondary task (Flecken et al., 2014; Sakarias & Flecken, 2019). Participants were asked to press a button on the keyboard marked with a yellow sticker during the gray screen when a given event had been presented twice in a row. There were 5 repeating events in total. Crucially, all of these repeating events were transitive filler events.

Before the experimental trials started, participants completed 3 practice trials, followed by optional feedback and the opportunity to ask questions. After the practice trials, a 5-point calibration and validation procedure was completed. This part of the experiment lasted approximately 10 min.

In order to ensure that the setups used in the non-linguistic and linguistic tasks were similar, in both tasks the participants were seated across from a confederate whom they believed was another naïve participant. The confederate was included to make the linguistic task more communicative and to elicit more naturalistic descriptions from participants than, for example, speaking to a computer screen in an empty room. The confederate was instructed to just listen and not to direct any questions or comments to the participant. Crucially, in the non-linguistic task, the confederate was busy filling out a questionnaire on a laptop in front of her and did not engage with the participant.

2.3.2. Linguistic eye-tracking task

Participants watched 50 video clips of events (25 motion events and 25 transitive filler events) presented on the computer screen while their eye movements were recorded. In each trial, participants first saw a fixation screen, containing a fixation cross in the center, for 1000 ms. Next, an event was shown for 2500 ms. Finally, a gray screen was presented. Participants were asked to describe what had happened in the video to the confederate once the gray screen appeared. Participants were informed that their eye movements were not recorded during the gray screen, thus they were free to move, look at the other participant and use their hands while speaking. Participants' descriptions (speech and co-speech gestures) during the linguistic task were recorded with a Canon video camera. After the participant had finished the description, the confederate clicked a button on the mouse marked with a blue sticker to initiate the next trial.

Before the experimental trials started, participants completed 2 practice trials, followed by optional feedback and the opportunity to ask questions. After the practice trials, a 5-point calibration and validation procedure was completed. Then, after calibration, the experimental trials started. The calibration procedure was repeated once in the middle of the task.

The confederate was the same research assistant as in the non-linguistic task whom the participant believed was another participant. The confederate was instructed to listen to the participant's descriptions and click the blue button when the other person finished describing. They were also told to listen the descriptions carefully because they would be asked to answer some questions afterwards. This part of the experiment lasted approximately 15 min.

2.4. Coding

Descriptions of target motion events were transcribed and coded for the presence of path and manner information in speech and gesture using ELAN software (Lausberg & Sloetjes, 2009) by a native speaker of Turkish.

2.4.1. Speech coding

First, event descriptions were segmented into clauses. Each clause consisted of a main verb and its subordinate verbs, if any. Clauses could be coordinated by conjunctions (*ve/and, ama/but, sonra/then*) or connective morphemes (*−erek, −e… e…, −ıp …ıp*). Each clause was coded for the presence of path and manner information in speech and gesture separately (following conventions in Allen et al., 2007). At the trial level participants' speech could either include one component of motion (path-only or manner-only) or both components (path + manner).

In speech, path information was coded as present if it was expressed with path verbs (e.g., *gir/enter, yaklaş/approach, geç/pass, git/go*) or outside of the verb in post-positional phrases (e.g., *içi-(n)e/to-inside*). All path mentions were further coded for how path information was encoded (i.e., within a path verb or outside of the verb). Manner information was coded as present if it was expressed as either a manner verb subordinated to a path verb with a connective (e.g., *koş-arak/ run*-CONN) or as main manner verbs (e.g., *koş/run, yürü/walk, zıpla/jump*).

In order to ensure reliability, 25% of the speech data were coded by a second coder who was also a native speaker of Turkish. The agreement between the coders for the presence of path and/or manner information in speech was 94.1% at the clause level ($\kappa = 0.921$). All disagreements were discussed to reach 100% agreement.

2.4.2. Gesture coding

First, we segmented gesture strokes (most meaningful part of the movement) that accompanied speech and represented path and or manner of motion. Each gesture was coded for the presence of path and manner information (following conventions in Özyürek et al., 2005, Özyürek et al., 2008). Path information was coded as present if speakers traced the figure's change of location across space. Speakers could trace the trajectory of motion either in the lateral axis (from left to right or from right to left) or in the sagittal axis (moving away from or towards the body). Pointing gestures to the location of the landmark object were not coded as path gestures because they do not trace the trajectory of motion and hence fail to convey path information. Manner information was coded as present if the speakers produced a gesture that depicted how the motion unfolds in a non-linear way with a body part chosen to represent the figure (e.g., inverted V-hand shape with wiggling fingers to indicate *walking*). Manner gestures could indicate the manner of motion from an observer's perspective (e.g., an index finger moving up and down to indicate *jumping*) or could be an enactment of the figure's posture during motion from an actor's perspective (e.g., moving the arms up and down to indicate *running*).

At the trial level, participants' gestures could either include one component of motion (path-only, Fig. 2a or manner-only, Fig. 2b) or both components (path + manner, Fig. 2c). When both components are gestured these gestures could either be a combination of separate path and manner gestures (e.g., a gesture like the one in Fig. 2a and another gesture like the one in Fig. 2b) or a single gesture that conflates path and manner, even though the latter pattern was quite rare (10%). See Fig. 2 for examples.

In order to ensure reliability, 25% of the gesture data were coded by the same second coder as for speech. The agreement between the coders for the presence of path and manner information in gesture were 87.9% at the clause level ($\kappa = 0.748$). All disagreements were discussed to reach 100% agreement.

2.5. Preprocessing of eye movement data

Two rectangular Areas of Interest (AoI) were defined for each of the target motion event stimuli using SMI BeGaze software. Path AoI was defined as the area surrounding the ground object based on previous eye-tracking work on spontaneous motion events (Bunger et al., 2012, 2016, 2021; Papafragou et al., 2008; Trueswell & Papafragou, 2010). Size and position of the Path AoI remained the same throughout the trial since the ground object remained static. Manner AoI was defined as the area surrounding the legs, torso and arms of the figure. Because the figure moved across the screen as the motion unfolded, the coordinates of the Manner AoI had to be updated throughout the trial. To do so,

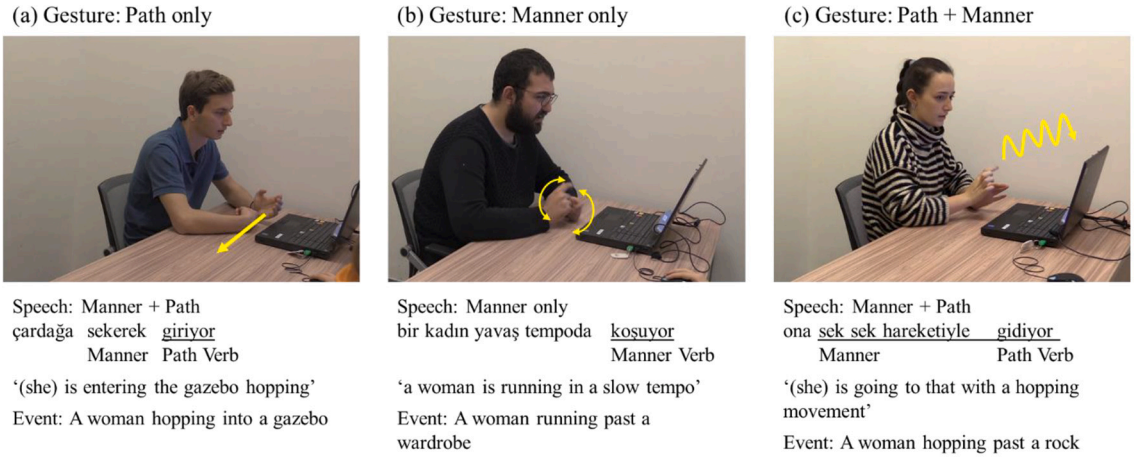

Fig. 2. Examples of path and manner encodings in gesture: (a) Path-only gesture (b) Manner-only gesture (c) Path + Manner gesture. In (a) and (c) path is encoded in the main verb and manner is encoded in a subordinate verb in speech. In (b) manner is encoded in the main verb in speech. Underlines indicate the parts of speech that the gesture overlaps with.

anchor points for the position of Manner AoI were created by repositioning the AoI at every 100 ms for the entire 2500 ms. Manner AoI size and shape remained the same across anchor points. Based on these anchor points BeGaze created a dynamic Manner AoI that moved along a connected path and the coordinates of the AoI were updated in a way that always included the legs, arms and torso of the figure. Fixations to the AoIs were computed by SMI BeGaze software. The onset and offset of stimuli for each trial were marked by a message sent from Presentation software to the eye-tracker. Using an R script (version 3.4.3) (R Core Team, 2018), we determined whether a fixation fell into one of the AoIs in successive 100 ms time bins for 2500 ms. Participants with more than 25% trackloss across all trials in a given task were excluded from the analysis for both tasks ($n = 1$ due to trackloss in the linguistic task). Additionally, we excluded trials in which trackloss was higher than 50% (non-linguistic task: 2.59%, linguistic task: 0.86%).

3. Results

3.1. Language production: Event descriptions in speech and gesture

Speech and gesture production data were analyzed using log-linear models with Poisson-distributed residuals. Models were fit using *glm* function with *stats* package in R (version 4.0.3; R Core Team, 2020). Significance levels for pairwise comparisons with corrections for multiple comparisons were obtained using *emmeans* (version 1.5.5–1; Lenth, 2021) and *multcomp* (version 1.4–16; Hothorn, Bretz, & Westfall, 2008) packages. Data and analysis code are available at https://osf.io/st5gb/. Table 1 shows the distribution of information about path and manner of motion across speech and gesture that we explore further below with statistical analyses for speech and gesture separately.

3.1.1. Speech

For speech analysis we tested to what extent event descriptions in speech conform to language-specific patterns of motion event encoding such that participants would be more likely to produce descriptions that encode path of motion only. A log-linear model tested the fixed effect of speech type (manner-only, path-only, path + manner) on counts of mention in speech (1 = mentioned, 0 = not mentioned) at the trial level as the dependent measure. Participants were more likely to use path + manner descriptions in speech compared to both manner-only ($\beta = 1.755$, $SE = 0.127$, $z = 13.84$, $p < .001$) and path-only ($\beta = 2.407$, $SE = 0.169$, $z = 14.21$, $p < .001$) descriptions (see Table 1). Participants were also more likely to use manner-only descriptions than path-only descriptions ($\beta = 0.653$, $SE = 0.200$, $z = 3.26$, $p = .003$). This indicated that, contrary to our initial expectation, most of the time participants produced descriptions that encoded both path and manner in speech.

Next, we tested the prediction that path of motion would be more likely to be encoded in the main verb as opposed to outside of the verb. For this analysis, we focused on a subset of the data (85%) in which participants encoded path of motion in speech either using a path-only description or a path + manner description (see Table 1). Thus, we excluded data from 15% of trials in which participants did not encode path information and encoded manner information only. When participants encoded path of motion in speech, as expected, the majority of the path mentions were in path verbs (63% of path mentions) and path mentions outside of the verb (i.e., in post-positional phrases only) were less frequent (37% of path mentions). A log-linear model tested the fixed effect of type of path encoding (post-positional phrases, numerically contrast coded as −1/2; verbs, numerically contrast coded as 1/2) on counts of mention in speech (1 = mentioned, 0 = not mentioned) at the trial level. There was a fixed effect of type of path encoding, indicating that, when participants encoded path of motion in speech, they were more likely to use path verbs than post-positional phrases only ($\beta = 0.542$, $SE = 0.097$, $z = 5.59$, $p < .001$).

Table 1
Proportion of Path and Manner mentions across speech and gesture types.

	Speech Type			
Gesture Type	Manner-only	Path-only	Path + Manner	Total
None	0.08 (0.04)	0.03 (0.02)	0.39 (0.07)	0.49 (0.08)
Manner-only	0.03 (0.02)	0.00 (0.00)	0.05 (0.02)	0.08 (0.03)
Path-only	0.02 (0.01)	0.03 (0.01)	0.22 (0.05)	0.26 (0.05)
Path + Manner	0.01 (0.01)	0.01 (0.01)	0.14 (0.03)	0.16 (0.04)
Total	0.14 (0.05)	0.07 (0.03)	0.79 (0.05)	1.00

Note. The values in the parentheses indicate standard error of participant means.

3.1.2. Gesture

Turning to motion event encodings in gesture, we first examined to what extent the gestures that speakers produced conform to language-specific patterns. Of interest was whether participants were more likely to produce gestures that only encode path of motion compared to path + manner or manner-only gestures due to language-specific encoding of path in the main verb in Turkish. To do so, we focused on the trials in which participants produced a gesture (51% of all trials) and assessed which of the three gestures types (path-only, manner-only, and path + manner) was most frequent. A log-linear model tested the fixed effect of gesture type (manner-only, path-only, path + manner) on counts of mention in gesture (1 = mentioned, 0 = not mentioned) at the

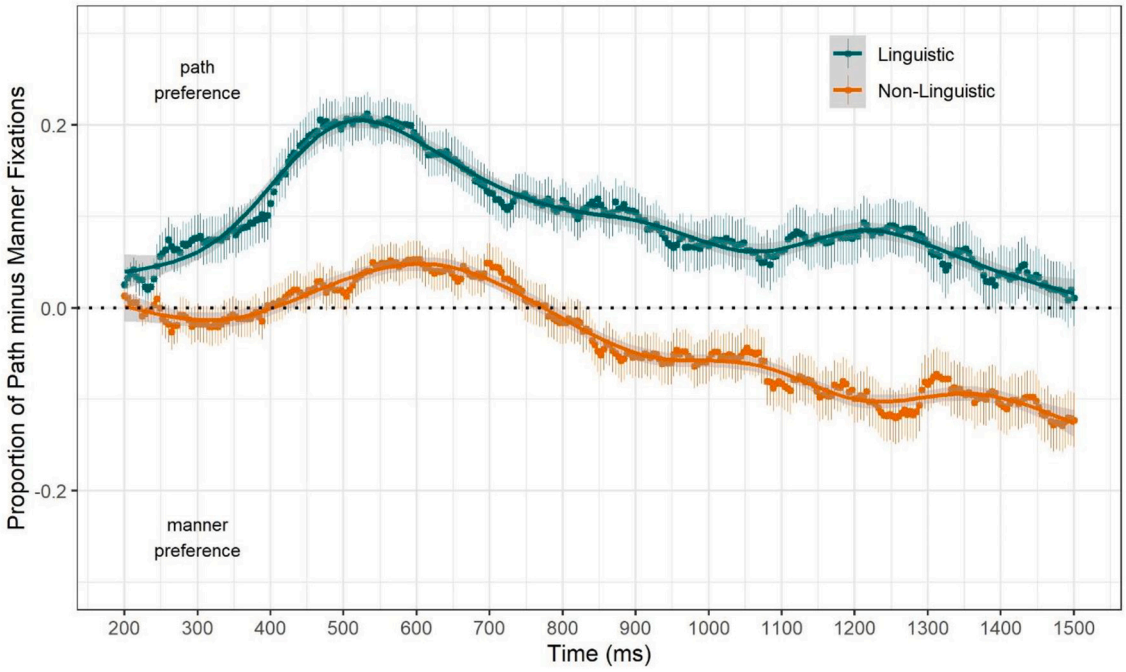

Fig. 3. Proportion of Path minus Manner fixations across linguistic and non-linguistic tasks over time (points). Error bars indicate standard error of participant means. Positive values indicate a preference to fixate on Path of motion; negative values indicate a preference to fixate on the Manner of motion.

trial level as the dependent measure. Participants were more likely to produce path-only gestures compared to both path + manner ($\beta = 0.530$, $SE = 0.138$, $z = 3.83$, $p < .001$) and manner-only ($\beta = 1.165$, $SE = 0.173$, $z = 6.75$, $p < .001$) gestures (Table 1). Furthermore, participants were more likely to produce path + manner gestures compared to manner-only gestures ($\beta = 0.635$, $SE = 0.187$, $z = 3.40$, $p = .002$). This confirms that path-only gestures were indeed produced most frequently by our participants.

Summarizing, language production data showed that in speech participants most frequently produced descriptions that encoded path and manner together. Furthermore, and most importantly for the purpose of our study, path was mostly encoded within the main verb in speech. In gesture, participants most frequently produced gestures that encoded only path of motion. These patterns largely conform to language-specific encoding of motion events in speech and gesture in Turkish. In the following section, we test the relation between visual attention and these language-specific encodings in speech and gesture.

3.2. Analysis of eye movements

We were interested in testing whether the time course of the relative attention allocated to path over manner during message preparation changed across tasks, types of path encoding in speech and types of path encoding in gesture. To test these hypotheses, we analyzed the time course of eye movements using Growth Curve Analysis (GCA; Mirman, 2014, Mirman, Dixon, & Magnuson, 2008).[1] GCA is a multilevel regression method designed for analyzing time course data. GCA uses polynomial functions to model time course and is able to capture changes in time course that follow any shape. Hence, this approach is suitable for modelling the change in eye movements over time in our data, which followed a non-linear shape (i.e., initial increase followed by a decrease, see Figs. 3–5). GCA is also able to quantify variation due to fixed effects (i.e., group-level effects; in our case: tasks, types of path encoding in speech and types of path encoding in gesture) as well as the random variation introduced by individual differences (i.e., participants or items). For our dependent variable, we followed prior eye tracking work in the motion event domain (Bunger et al., 2012, 2021; Papafragou et al., 2008; Trueswell & Papafragou, 2010) and used difference scores as a measure for preference to fixate on one event component over the other. Thus, our dependent variable was the difference between the proportion of fixations to the Path AoI (out of all fixations) minus the proportion of fixations to the Manner AoI (out of all fixations). For the analyses, data were aggregated into 100 ms time bins. All analyses were conducted on log transformed odds ratio of proportions of Path minus Manner fixations. We excluded 8.1% of the data due to participants not fixating on anywhere on the stimuli (either path or manner AoIs, or elsewhere on the scene) within a bin.

Since we were interested in examining the differences in eye movements tied to linguistic planning, our analyses focused on a subset of the time course of eye movements. Specifically, we focused on the window spanning 200 ms to 1500 ms after stimulus onset. We excluded the eye movements in the last 1000 ms of the trial (between 1500 ms and 2500 ms) from the analyses for two reasons. First, previous work has shown that event apprehension for utterance planning is rapid (Griffin & Bock, 2000) and eye movements in this earlier time window can reflect changes in visual event apprehension due to linguistic planning more accurately. Second, since the target motion events in our stimuli were all goal directed, the figure reached the landmark object at the end of the clip and therefore the path and manner AoIs overlapped in the last second. We also excluded the eye movements in the first 200 ms from the analyses since it takes about 200 ms for participants to plan and land a saccade (Matin, Shao, & Boff, 1993).

All models were fit with the *lme4* package (version 1.1.17; Bates, Mächler, Bolker, & Walker, 2015) in *R* (version 4.0.3; R Core Team, 2020). Polynomial growth functions were created using *psy811* package (version, 1.0; Mirman, 2015). *P* values for the *t*-tests on the parameter

[1] Recently, there have been concerns about the use of GCA based on data from the visual world paradigm in a language comprehension task (Huang & Snedeker, 2020). Following the approach recommended by Huang and Snedeker (2020), we modelled our data using binomial logistic regressions as well and replicated the findings from the GCA reported in the current article. Results and analysis code from both approaches can be can be found at: https://osf.io/st5gb/.

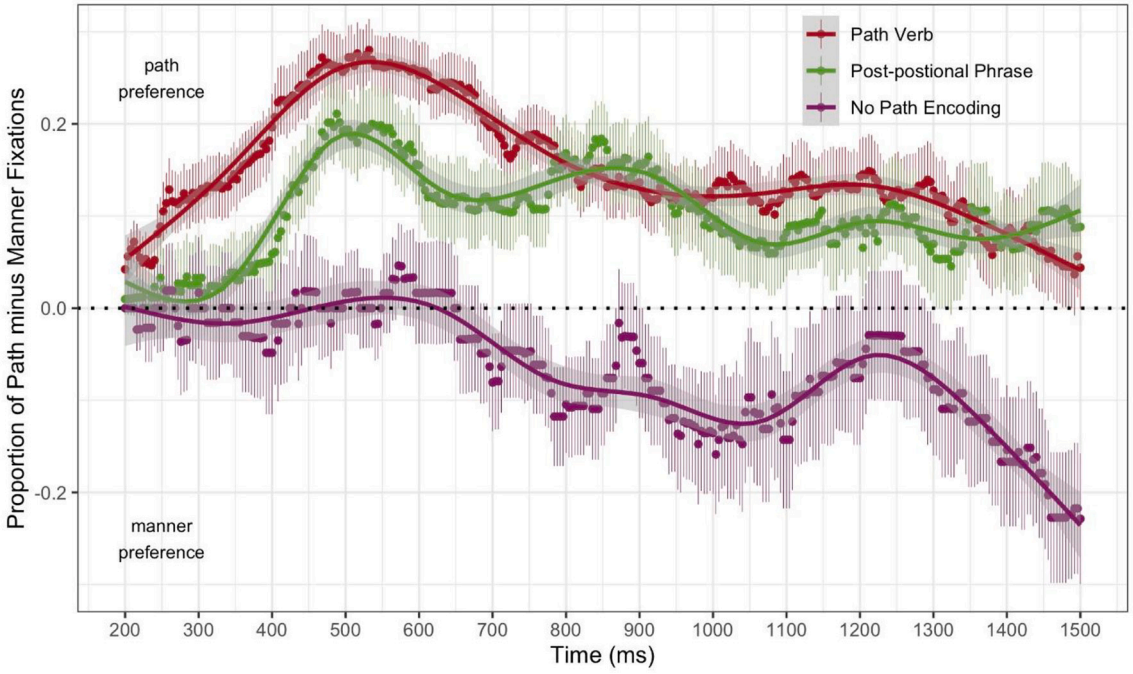

Fig. 4. Proportion of Path minus Manner fixations across different types of path encoding in speech over time (points). Error bars indicate standard error of participant means. Positive values indicate a preference to fixate on Path of motion; negative values indicate a preference to fixate on the Manner of motion.

estimates were obtained with *lmerTest* package (version 3.1–1, Kuznetsova, Brockhoff, & Christensen, 2017). Figures were produced using *ggplot2* package (version 3.2.1, Wickham, 2016). Details of model fitting are available in Supplementary Materials. Data and analysis code are available at https://osf.io/st5gb/.

3.2.1. *Eye movements in linguistic vs. non-linguistic tasks*

We first tested to what extent eye movements were guided by engaging in linguistic planning, such that participants would allocate more attention to path over manner of motion in the linguistic task compared to the non-linguistic task. Fig. 3 shows the proportion of

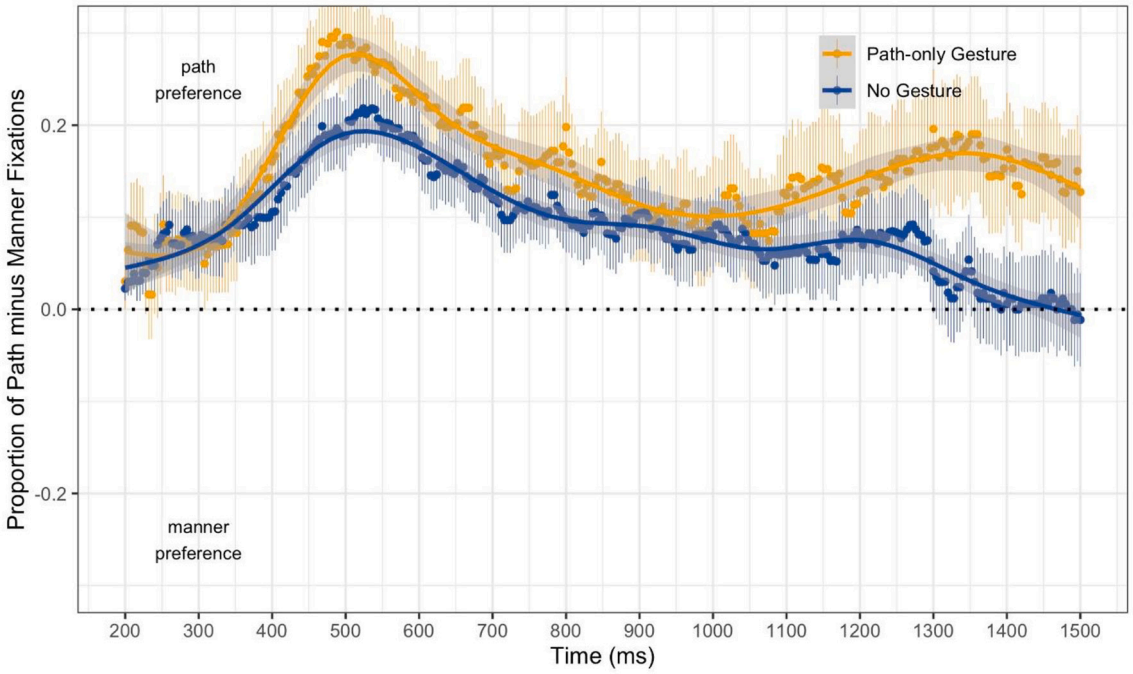

Fig. 5. Proportion of Path minus Manner fixations across Path-only Gesture and No Gesture trials when both path and manner is mentioned in speech over time (points). Error bars indicate standard error of participant means. Positive values indicate a preference to fixate on Path of motion; negative values indicate a preference to fixate on the Manner of motion.

Table 2
Parameter estimates of the fixed effects for the best-fitting model of proportion of Path minus Manner fixations across linguistic and non-linguistic tasks. Significant p-values that are critical to the analysis are in boldface.

Fixed Effect	*Estimate*	*S.E.*	*t*	*p-value*
(Intercept)	−0.209	0.044	−4.778	< 0.001
Task [N-Ling vs. Ling]	0.098	0.011	9.180	**< 0.001**
Linear Time	−0.397	0.020	−19.805	< 0.001
Quadratic Time	−0.217	0.020	−10.828	< 0.001
Cubic Time	−0.056	0.020	−2.786	0.005
Quartic Time	0.074	0.020	3.693	< 0.001
Task [N-Ling vs. Ling] x Linear Time	0.169	0.040	4.211	**< 0.001**
Task [N-Ling vs. Ling] x Quadratic Time	−0.011	0.040	−0.284	0.776
Task [N-Ling vs. Ling] x Cubic Time	−0.013	0.040	−0.327	0.744
Task [N-Ling vs. Ling] x Quartic Time	−0.015	0.040	−0.365	0.715

fixations to path minus manner over time across linguistic and non-linguistic tasks.

Polynomial growth functions were added stepwise to the model and the overall time course of eye movements were modelled with fourth-order orthogonal time terms in addition to the fixed effect of task (non-linguistic contrast coded as −1/2; linguistic contrast coded as 1/2). The model also included random intercepts for Subjects and Items (more complex models with random slopes did not converge). Parameter estimates from the model are presented in Table 2. Most importantly for present purposes, the model revealed a fixed effect of task: participants had higher preference to fixate on path over manner in the linguistic task compared to the non-linguistic task. Furthermore, there was an interaction between task and the linear time term, indicating that over time, the decrease in path preference was less steep in the linguistic task than the non-linguistic task. The model also revealed that the time course of the data was characterized by Quadratic, Cubic and Quartic terms for time; however, none of these time terms interacted with task, indicating that the curvature was similar across tasks. Overall, these findings indicate that the time course of eye movements varies across linguistic and non-linguistic task with more attention allocated to path over manner of motion in the linguistic task.

3.2.2. *Eye movements across different types of path encoding in speech*

Next, we tested whether and to what extent eye movements in the linguistic task varied across different types of path encoding in speech. Of interest was whether participants would allocate more attention to path over manner of motion when they encoded path in speech (with either a post-positional phrase only or a verb) compared to when they did not encode path in their speech. Also of interest was whether participants would allocate even more attention to path over manner of motion when they encoded in a path verb as opposed to when they encoded it outside of the verb in post-positional phrases only. Fig. 4 shows the proportion of fixations to path minus manner over time when participants did not encode path in speech at all (i.e., manner-only), when they encoded it as a post-positional phrase only and when they encoded it as a path verb.

Polynomial growth functions were added stepwise to the model and the overall time course of eye movements were modelled with fourth-order orthogonal time terms. The fixed effect of path encoding in speech (No Path, Post-positional Phrase, Path Verb) was tested with planned contrasts on only the linear and quadratic time terms. Adding the interactions between the fixed effect of path encoding in speech and the Cubic (χ^2 (2) = 0.178, p = .915), and Quartic (χ^2 (2) = 0.047, p = .977) time terms did not improve model fit (see Supplementary Materials for details). For the fixed effect of path encoding in speech, first we compared trials in which participants did not encode path in speech to any type of path encoding (no path encoding contrast coded as −2/3, post-positional phrase contrast coded as 1/3, and path verb contrast coded as 1/3). Then, we compared trials in which participants encoded path in post-positional phrases only to when they used path verbs (no path encoding contrast coded as 0, post-positional phrase contrast coded as −1/2, and path verb contrast coded as 1/2). The model also included random slopes for path encoding in speech by Subjects and Items (models with random slopes for time terms did not converge). Parameter estimates for the fixed effects from the best-fitting model are presented in Table 3.

As earlier, curvature was similar across different types of path encoding in speech: an initial increase in path preference was followed by a decrease and a second increase and decrease (quartic term). However, and most importantly for present purposes, both of the contrasts for path encoding in speech interacted with the linear time term. This indicated that the overall decrease in path preference was steeper when participants did not encode path in speech (i.e., encoded manner only) compared to when they encoded path in speech in any way (with either a post-positional phrase only or a path verb). Additionally, the overall decrease in path preference was steeper when participants encoded path in speech with a verb compared to when they encoded it with a post-positional phrase only. This reflects the fact that when participants encoded path in speech path preference was particularly high at the beginning of message preparation. However, by the end of the analysis window preference to fixate on path over manner was quite similar across path encodings in verbs and post-positional phrases only. Thus, time course of eye movements for path encodings in path verbs was characterized by a stronger negative slope. Overall, these findings indicate that the time course of eye movements in the linguistic task varies across different types of path encoding in speech with more attention allocated to path of motion compared to manner of motion when path was encoded in verbs.

3.2.3. *Eye movements in relation to path encoding in gesture*

Finally, we tested to what extent time course of eye movements in the linguistic task varied when they were accompanied by different types of

Table 3
Parameter estimates of the fixed effects for the best-fitting model of proportion of Path minus Manner fixations across different types of path encoding in speech. Significant p-values that are critical to the analysis are in boldface.

Fixed Effect	*Estimate*	*S.E.*	*t*	*p-value*
(Intercept)	−0.205	0.050	−4.075	< 0.001
Path Speech [No Path vs. Path]	0.187	0.071	2.616	**0.015**
Path Speech [PP vs. Verb]	0.022	0.051	0.424	0.674
Linear Time	−0.332	0.030	−11.125	< 0.001
Quadratic Time	−0.210	0.030	−7.042	< 0.001
Cubic Time	−0.069	0.026	−2.669	0.008
Quartic Time	0.058	0.026	2.265	0.024
Path Speech [Path vs. No Path] x Linear Time	0.273	0.074	3.704	**< 0.001**
Path Speech [PP vs. Verb] x Linear Time	−0.201	0.058	−3.434	**0.001**
Path Speech [Path vs. No Path] x Quadratic Time	−0.034	0.074	−0.461	0.645
Path Speech [PP vs. Verb] x Quadratic Time	−0.088	0.059	−1.501	0.133

Table 4
Parameter estimates from the best-fitting model on the proportion of Path minus Manner fixations across Path-only gesture and No Gesture trials when both path and manner is mentioned in speech. Significant p-values that are critical to the analysis are in boldface.

Fixed Effect	*Estimate*	*S.E.*	*t*	*p-value*
(Intercept)	−0.170	0.051	−3.352	< 0.001
Gesture Type [None vs. Path]	−0.045	0.079	−0.567	0.574
Linear Time	−0.243	0.032	−7.507	< 0.001
Quadratic Time	−0.176	0.032	−5.443	< 0.001
Cubic Time	−0.017	0.032	−0.541	0.589
Gesture Type [None vs. Path] x Linear Time	0.144	0.065	2.223	0.026
Gesture Type [None vs. Path] x Quadratic Time	0.152	0.065	2.339	0.019
Gesture Type [None vs. Path] x Cubic Time	0.199	0.065	3.077	**0.002**

path encoding in gesture. For this analysis, we focused on a subset of the eye-gaze data based on the linguistic encoding of event components in speech and gesture. As seen in Table 1, the most frequent encoding pattern was path + manner descriptions in speech and path-only gestures. In order to keep the semantic elements encoded in speech constant, we focused on the trials in which participants encoded both path and manner in speech. Then, we examined the time course of eye movements when participants did not use a gesture as opposed to when they used a path-only gesture. Thus, we excluded trials with less frequent speech (path-only and manner-only) and gesture (manner-only or path + manner) patterns from this analysis. This allowed us to test our prediction that path encoding in gesture in addition to what was already encoded in speech would be related to more attention to path over manner of motion in visual event apprehension. Fig. 5 shows the proportion of fixations to path minus manner over time when participants produced a path-only gesture and when they did not produce any gesture.

Polynomial growth functions were added stepwise to the model and the overall time course of eye movements were modelled with third-order orthogonal time terms in addition to the fixed effect of gesture type (no gesture contrast coded as −1/2; path-only gesture contrast coded as 1/2). The model also included random slopes for gesture type by Subjects and Items (models with random slopes for time terms did not converge). Parameter estimates from the model are presented in Table 4.

As Table 4 shows, there was no effect of gesture type, indicating no difference in overall path over manner preference when a path-only gesture was produced compared to when no gestures were produced. However, there was an interaction between gesture type and the cubic time term, indicating differences in curvature when participants produced a path-only gesture versus when they did not produce any gestures. This interaction is likely to be driven by two patterns in the data. First, the initial peak in path preference follows a deeper curve when participants produce a path-only gesture. Second, there is a second rise in path preference when participants produced a path-only gesture compared to when they did not produce any gestures. Overall, these data indicate that time course of relative attention allocated to path over manner of motion in the linguistic task was different across different types of path encoding in gesture, with more attention allocated to path of motion over manner of motion when path was additionally encoded in gesture.

3.2.4. Further exploration of the relation between eye movements and path encoding in gesture

In order to further evaluate how visual attention varies in relation to path encoding in gesture, we conducted two sets of exploratory analyses. In a first set of analyses, we checked whether the variation in the time course of fixations to path and manner of motion linked to gesture production was simply a byproduct of what was encoded in the accompanying speech. It is possible that the speech accompanying path-only gestures had more path encodings in verbs than with post-positional phrases only and thus the results reported above could be due to encoding differences in speech instead of having produced path-only gestures. To rule out this possibility, we examined if the encoding of path of motion in speech was similar (verbs versus post-positional phrases only) in cases when speech was accompanied with path-only gestures versus no gesture.[2] We found that, when participants produced a path-only gesture, 56% of these path-only gestures occurred with speech in which path was expressed with a verb and 44% of these path-only gestures occurred with speech in which path was encoded with a post-positional phrase only. When participants did not produce any gesture, in speech path was encoded with a verb in 61% of the time and with a post-positional phrase only in 39% of the time. In fact, a chi-square test revealed that the distribution of trials that path was encoded with a verb vs. a post-positional phrase only in speech was similar across trials with path-only gesture vs. no gesture (χ^2 (1) = 0.452, p = .502).

To ensure that the variation in the time course of eye movements in relation to path encoding in gesture remained significant after statistically controlling for type of path encoding in speech, we tested the best-fitting growth curve model on the time course of eye movements by adding type of path encoding in speech as a fixed factor (post-positional phrase contrast coded as −1/2, and path verb contrast coded as 1/2). The model structure for the remaining fixed and random effects was the same (see Supplementary Materials for details of model fitting and the complete list of parameter estimates). This model replicated the previously reported interaction between path encoding in speech and the linear time term (β = −0.290, SE = 0.066, t = −4.368, p < .001). Crucially, the interaction between gesture type and cubic time term remained statistically significant (β = 0.183, SE = 0.066, t = 2.769, p = .006) but did not further interact with path encoding in speech (β = 0.208, SE = 0.132, t = 1.571, p = .116). This indicates that the differences in curvature when participants produced a path-only gesture versus when they did not produce any gestures was observed both when participants encoded path in speech with a verb and when they encoded path in speech with a post-positional phrase only. These data confirm that the differences in the time course of eye movements linked to additional encoding of path of motion in gesture were sustained even after controlling for how path of motion was encoded in the accompanying speech.

In a second set of analyses, we explored whether there were any other systematic differences in the speech that accompanied path only gestures versus no gesture in terms of the ease of planning of the descriptions. It is commonly assumed that speakers produce gestures to compensate for difficulties in speech production. To eliminate the possibility that participants gestured about (and attended to) path of motion merely because they had difficulty speaking about it, we inspected the same subset of the data that was included in the analyses of eye

[2] For these trials, the distribution of the path-only vs. no gesture trials regarding how manner of motion was encoded was as follows. When participants produced a path-only gesture, 44% of these path-only gestures occurred with speech in which manner was expressed with a main verb and 56% of these path-only gestures occurred with speech in which manner was encoded in a subordinate verb or adverbial phrase. When participants did not produce any gesture, in speech manner was encoded with a main verb in speech 39% of the time with a subordinate verb or adverbial phrase 61% of the time.

movements in relation to path encoding in gesture. That is, we focused on the trials in which participants encoded both path and manner of motion in speech and produced either a path-only gesture or did not produce a gesture at all. Next, we coded for instances of disfluencies in speech. Disfluencies were defined as filled or unfilled pauses in speech or producing word fragments and self-corrections (following Graziano & Gullberg, 2018). Overall, participants were disfluent about path of motion only 5.9% of the time. When they produced path-only gestures, they were disfluent about path of motion 7.8% of the time, and they were not disfluent about path of motion 92.2% of the time. Similarly, when participants did not produce any gestures together with speech, they were disfluent about path of motion 4.8% of the time, and they were not disfluent about path of motion 95.2% of the time. A chi-square test confirmed that the distribution of the trials in which participants were versus were not disfluent about path of motion did not differ across path-only gesture versus no gesture trials (χ^2 (1) = 0.683, p = .409). These findings confirm that there were no systematic differences across path-only gesture versus no gesture trials in terms of difficulty in speaking about path of motion.

4. Discussion

Spoken language production guides visual event apprehension during message preparation (Gleitman et al., 2007; Griffin & Bock, 2000; Meyer et al., 1998) and in language-specific ways (Norcliffe, Konopka, et al., 2015; Sauppe, 2017; Sauppe et al., 2013). Our primary goal in the present study was to test if producing language-specific gestures along with speech further guides visual attention to events. Secondarily, as a novel contribution to previous work on cross-linguistic differences in event encoding in speech and visual attention, we tested whether eye gaze patterns vary in relation to variations in linguistic encoding within the typological framework of a specific language. Overall, our findings strongly suggest that language-specific encodings of path in the main verb (as opposed to outside of the verb) as well as producing path gestures along with speech guide visual attention allocated to path over manner of motion during message preparation.

4.1. Motion events in speech and gesture

In order to motivate our investigation of potential differences in visual attention linked to language-specific encoding of motion in speech and gesture, we began by exploring linguistic encoding of motion in speech and its links to gesture production in Turkish. This allowed us to re-establish that Turkish-speakers adhered to the patterns reported in previous typological and empirical work. We found that Turkish-speakers produced spoken descriptions that encoded both path and manner of motion more frequently than descriptions that encoded either only path or only manner. Even though this pattern was somewhat less expected based on typological patterns reported in prior work on motion (Slobin, 1996; Talmy, 1985), it is in line with similar reports from speakers of verb-framed languages especially when the manner of motion is salient, contrastive, and cannot be inferred from the context (Özyürek et al., 2008; Papafragou, Massey, & Gleitman, 2006; ter Bekke, Özyürek, & Ünal, 2022). Nevertheless, as expected, the majority of encodings included path of motion and these were expressed through path verbs. This is consistent with previous typological work on the encoding of motion events in verb-framed languages (Talmy, 1985; Turkish: Özyürek et al., 2008; Greek: Papafragou et al., 2008; Papafragou & Selimis, 2010).

In gesture, Turkish-speakers produced path-only gestures more frequently than gestures that encoded only manner or both path and manner. This path-only bias found for gesture supports the Interface Model of multimodal production by showing that the semantic elements that were more likely to be encoded in the main verb were also more likely to be encoded in gesture (Kita & Özyürek, 2003). The speech and gesture patterns in the present study conform to typological gesture patterns in verb-framed languages and contrast with data from speakers of satellite-framed languages where speakers use more manner gestures and express manner and path in a single gesture (Kita et al., 2007; Özçalışkan, 2016; Özçalışkan et al., 2016a, 2016b, Özyürek et al., 2005, Özyürek et al., 2008). The combination of the most frequent encoding patterns in speech and gesture observed in the present study also coheres with the findings of a recent study conducted on Farsi – a language that has a mixed verb-framed and satellite-framed typology (Akhavan, Nozari, & Göksun, 2017). In that study, speakers also encoded path and manner equally frequently in speech, using light verbs plus prepositions to encode path and adverbs to encode manner, and were more likely to produce gestures that encoded only path. Together, these data provide behavioral evidence for the idea that speech and gesture form a tightly integrated system where speech and gesture (at least those about motion) are integrated with what can be packaged in a verb. This idea is corroborated by an exploratory finding in our data: speech disfluencies about path of motion were equally likely to co-occur with or without a path gesture. This is in line with recent evidence that gesture production does not necessarily help speakers retrieve words spatial content (Kısa, Goldin-Meadow, & Casasanto, 2021 see also Graziano & Gullberg, 2018). Both sets of findings challenge the view that gestures are produced merely to compensate for difficulties in word retrieval. Finally, our speech and gesture production findings confirm that multimodal linguistic encoding of motion is a good test bed for investigating further links between visual attention and language-specific speech and gesture production.

4.2. Attention to motion events prior to speech and gesture production

Turning to eye movements, our eye-tracking data revealed that when Turkish-speakers linguistically encoded events, they allocated more attention to path over manner of motion compared to when they non-linguistically encoded events. These data offer further support for the idea that engaging in linguistic planning guides visual attention (Levelt, 1989). Our findings replicate findings of previous cross-linguistic eye-tracking studies on motion events (Bunger et al., 2012, 2016, 2021; Flecken et al., 2014; Sakarias & Flecken, 2019) including work with speakers of other verb-framed languages (Greek; Papafragou et al., 2008, Trueswell & Papafragou, 2010) and extend these findings to Turkish – a language that had not been studied in this respect before. This finding is also important in showing that path preference in visual attention observed in prior work with Greek-speakers is not merely a reflection of order of mention of event components. In Greek, path of motion is typically mentioned before manner of motion. On the other hand, Turkish is a verb-final language and path of motion is typically mentioned after manner of motion. Despite this variation in word order, speakers of both (verb-framed) languages allocate more attention to path compared to manner of motion during early event apprehension. This suggests that the semantic information encoded within the verb has an important role in guiding visual attention to events (Levelt, 1989).

Our findings also go beyond prior work by pinpointing which types of linguistic encoding in speech within the variations in a given typology are more likely to guide eye movements during language production. Specifically, we showed that Turkish-speakers allocated more attention to path over manner of motion when they encoded path in speech compared to when they did not. Furthermore, they allocated even more attention to path over manner when they encoded path within a verb compared to outside of the verb with post-positional phrases only (i.e., in line with the verb-framed typology). This is compatible with the thinking for speaking hypothesis (Slobin, 1996).

Only two prior studies thus far have examined between- and within-language variation in eye movements based on whether or not some motion event components were mentioned in speech (Bunger et al., 2016, 2021). This work demonstrated that when English- and Greek-speakers produced content-wise similar descriptions of caused motion events (e.g., mentioned both causative and resultative subevents) their

eye movements prior to speaking were similar. Our findings highlight the importance of looking beyond the content of the descriptions (i.e., whether or not an event component is mentioned) for showing subtle nuances in visual event apprehension tied to language-specific event encoding in speech. To our knowledge, our data offer the first empirical evidence that attention allocation prior to speaking not only varies cross-linguistically but also within speakers of a single language in ways linked to language-specific encoding of motion paths. Together, these data provide further evidence for the idea that verbs are the main processing units of speech planning (e.g., Bock, 1982; Griffin & Bock; Kita & Özyürek, 2003; Levelt, 1989; Norcliffe & Konopka, 2015, among others).

As a very novel contribution, we also showed that attention allocation prior to linguistic encoding was linked to language-specific encoding of motion event components in co-speech gestures. Turkish-speakers allocated even more attention path over manner of motion when their spoken descriptions that included both path and manner were accompanied by a path-only gesture compared to when they did not encode any motion event components in gesture. This pattern possibly emerged due to the fact that path gestures included additional information about the direction of the motion in the left-right axis, that was not necessarily conveyed in path speech. Crucially, the speech produced with path-only gestures was similar to the speech produced without any gestures in several respects, including the syntactic encoding of path of motion. Furthermore, the variation in visual attention linked to path gesture production persisted even after controlling for how path of motion was encoded in the accompanying speech. This indicates that differences in visual attention related to additional encoding of path/direction of motion in gesture emerged in addition to the differences found in relation to speech. In addition to complementing prior behavioral findings on speech and gesture, these findings suggest that prior evidence on the relation between visual event apprehension and spoken language production may be extended to multimodal language production. They also provide evidence consistent with the Interface Model of multimodal language production (Kita & Özyürek, 2003).

These patterns suggest that at the planning stage, there are interactions between visual event apprehension, linguistic constraints on how motion is encoded in a specific language and the spatio-motoric imagery underlying gesture production by showing that what kind of semantic information can be packaged in a processing unit within the main verb predicts not only gesture production but also attention allocation to event components. Even though the model by Kita and Özyürek (2003) has previously proposed this interface at the conceptualization stage of multimodal language production, this is the first empirical investigation that reveals eye-gaze patterns during message preparation that are compatible for this aspect of the model.

5. Conclusions

The present study offers novel evidence suggesting that visual event apprehension is guided by multimodal linguistic encoding of events and that the links between the eye and the mouth may be extended to the eye and the hand. These influences seem to be susceptible to language-specific constraints on event encoding in both speech and gesture. Together, these findings advance our understanding of language and its processing as a multisensory multimodal phenomenon. Finally, the approach reported in this study offers new possibilities for future work investigating previously hypothesized tight links between event representation and language production (Knott & Takac, 2021; Ünal, Ji, & Papafragou, 2021) by taking the multimodal nature of language into account.

Data availability

The data and analysis code for the present study are available from https://osf.io/st5gb/

CRediT authorship contribution statement

Ercenur Ünal: Conceptualization, Methodology, Software, Investigation, Visualization, Formal analysis, Writing – original draft, Writing – review & editing. **Francie Manhardt:** Methodology, Software, Formal analysis, Writing – review & editing. **Aslı Özyürek:** Conceptualization, Methodology, Writing – review & editing, Supervision, Funding acquisition.

Acknowledgements

This research was supported by a NWO-VICI grant awarded by The Dutch Research Council (grant number 277-70-013) to A.Ö. We thank Sura Ertaş and Kees Oerlemans for assisting stimuli preparation, Özge Baturlar, Nilüfer Akdoğan and Yağmur Keleş for assisting data collection, Melis Odabaş, Şevval Nur Yağlı and Şevval Cihankaya for assisting data transcription and coding, Kimberley Mulder and Jeroen Geerts for contributing to data processing and Susanne Brouwer for providing statistical advice.

Appendix A. List of motion events used in the linguistic and non-linguistic tasks

	Motion Events (Set A)		Motion Event Stimuli (Set B)
1	a woman walking to a trash can	1	a woman walking to a pine tree
2	a woman running to a ladder	2	a woman running to a lamp
3	a woman hopping to a fountain	3	a woman hopping to a ladder
4	a woman leaping to a tree	4	a woman leaping to a trash can
5	a woman skipping to a rocks	5	a woman skipping to a mirror
6	a woman walking into a gazebo	6	a woman walking into an orange market stand
7	a woman running into a phone booth	7	a woman running into a garage
8	a woman hopping into a green market stand	8	a woman hopping into a gazebo
9	a woman leaping into a white tent	9	a woman leaping into a phone booth
10	a woman skipping into a beach bar	10	a woman skipping into a circus tent
11	a woman walking past a plant	11	a woman walking past a sign
12	a woman running past a mirror	12	a woman running past a wardrobe
13	a woman hopping past a wardrobe	13	a woman hopping past a rock
14	a woman leaping past a sign	14	a woman leaping past a fountain
15	a woman skipping past a lamp	15	a woman skipping past a cactus
16	a woman walking from a bush	16	a woman walking from a red rain umbrella
17	a woman running from a table	17	a woman running from a fire hydrant
18	a woman hopping from a table	18	a woman hopping from an orange armchair

(*continued on next page*)

(*continued*)

	Motion Events (Set A)		Motion Event Stimuli (Set B)
19	a woman leaping from a yellow armchair	19	a woman leaping from a traffic cone
20	a woman skipping from a fire hydrant	20	a woman skipping from a drawers
21	a woman walking out of a greenhouse	21	a woman walking out of a arbor
22	a woman running out of a tent	22	a woman running out of a white sun umbrella
23	a woman hopping out of a umbrella	23	a woman skipping out of a bus stop
24	a woman leaping out of a arbor	24	a woman leaping out of a green house
25	a woman skipping out of a bus stop	25	a woman hopping out of a tent

Note: Assignment of event sets A and B to the linguistic and non-linguistic tasks was counterbalanced across participants.

Appendix B. List of transitive filler events used in the linguistic and non-linguistic tasks

	Filler Events (Set A)		Filler Events (Set B)
1	breaking a cookie in half	1	wearing a scarf
2	building blocks	2	putting on reading glasses
3	coloring a star	3	throwing a plastic ball
4	crushing paper	4	peeling a banana
5	cutting a piece of paper	5	cutting an apple
6	drawing a line with a ruler	6	tearing a piece of white paper
7	drinking water	7	reading a newspaper
8	dropping dices in a jar	8	putting on a hat
9	ironing a tablecloth	9	playing the flute
10	knocking down blocks	10	wearing a gray glove
11	making a phone call	11	eating a piece of cake
12	opening a bag of chips	12	biting an apple
13	opening a coke can	13	closing a box
14	pouring chips into a bowl	14	rolling dice
15	putting cream on hands	15	opening the cover of a book
16	putting sticky note on a paper	16	putting tape on a paper
17	putting on headphones	17	painting nails
18	squeezing toothpaste on toothbrush	18	combing hair
19	stapling papers	19	inflating a balloon
20	texting a message	20	tearing a piece of paper towel
21	unlocking a lock	21	wrapping yarn around a yarn ball
22	unzipping a pouch	22	putting cards on a table
23	wearing a coat	23	putting together pieces of a puzzle
24	wiping a table	24	putting paper clips on paper
25	writing a letter	25	blowing candles

Note: Participants always received event set A in the non-linguistic task and event set B in the linguistic task. Within each task participants saw motion events and filler events in a mixed order.

Appendix C. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2022.105127.

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

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

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

ter Bekke, M., Özyürek, A., & Ünal, E. (2022). Speaking but not gesturing predicts event memory: A cross-linguistic comparison. *Language and Cognition*, 1–23. https://doi.org/10.1017/langcog.2022.3

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

Bunger, A., Skordos, D., Trueswell, J. C., & Papafragou, A. (2016). How children and adults encode causative events cross-linguistically: Implications for language production and attention. *Language, Cognition and Neuroscience, 31*(8), 1015–1037. https://doi.org/10.1080/23273798.2016.1175649

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), 1–22. https://doi.org/10.5334/gjgl.1210, 28.

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

Flecken, M., von Stutterheim, C., & Carroll, M. (2014). Grammatical aspect influences motion event perception: Evidence from a cross-linguistic non-verbal recognition task. *Language and Cognition, 6*(1), 45–78. https://doi.org/10.1017/langcog.2013.2

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

Graziano, M., & Gullberg, M. (2018). When speech stops, gesture stops: Evidence from developmental and crosslinguistic comparisons. *Frontiers in Psychology, 9*. https://doi.org/10.3389/fpsyg.2018.00879. Article 879.

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

Gullberg, M. (2011). Language-specific encoding of placement events in gestures. In J. Bohnemeyer, & E. Pederson (Eds.), *Event representation in language and cognition* (pp. 166–188). Cambridge University Press. https://doi.org/10.1017/CBO9780511782039.008.

Gullberg, M., & Indefrey, P. (2003). *Language background questionnaire*. Nijmegen: Max Planck Institute for Psycholinguistics. https://doi.org/10.13140/RG.2.2.21793.63843

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal, 50*(3), 346–363. https://doi.org/10.1002/bimj.200810425

Huang, Y., & Snedeker, J. (2020). Evidence from the visual world paradigm raises questions about unaccusativity and growth curve analyses. *Cognition, 200*. https://doi.org/10.1016/j.cognition.2020.104251. Article 104251.

Kamide, Y., Lindsay, S., Scheepers, C., & Kukona, A. (2016). Event processing in the visual world: Projected motion paths during spoken sentence comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(5), 804–812. https://doi.org/10.1037/xlm0000199

Kısa, Y. D., Goldin-Meadow, S., & Casasanto, D. (2021). Do gestures really facilitate speech production? *Journal of Experimental Psychology. General*. https://doi.org/10.1037/xge0001135

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 & Cognitive Processes, 22*(8), 1212–1236. https://doi.org/10.1080/01690960701461426

Knott, A., & Takac, M. (2021). Roles for event representations in sensorimotor experience, memory formation, and language processing. *Topics in Cognitive Science, 13*(1), 187–205. https://doi.org/10.1111/tops.12497

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., & Chawla, P. (1996). Nonverbal behavior and nonverbal communication: What do conversational hand gestures tell us? In M. Zanna (Ed.), *Advances in experimental social psychology* (pp. 389–450). Academic Press. https://doi.org/10.1016/S0065-2601(08)60241-5.

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. https://doi.org/10.1017/CBO9780511620850.017.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

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

Lenth, R. V. (2021). emmeans: Estimated marginal means, aka least-squares means. In *R package version 1.5.5-1*. https://CRAN.R-project.org/package=emmeans.

Levelt, W. (1989). *Speaking*. MIT Press.

Manhardt, F., Özyürek, A., Sümer, B., Mulder, K., Karadöllller, D. Z., & Brouwer, S. (2020). Iconicity guides visual attention: A comparison between signers' and speakers' eye gaze during message preparation. *Journal for Experimental Psychology: Learning, Memory, and Cognition, 46*(9), 1735–1753. https://doi.org/10.1037/xlm0000843

Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics, 53*(4), 372–380. https://doi.org/10.3758/BF03206780

McNeill, D. (2005). *Gesture and thought*. University of Chicago Press. https://doi.org/10.7208/chicago/9780226514642.001.0001

Melinger, A., & Levelt, W. J. (2004). Gesture and the communicative intention of the speaker. *Gesture, 4*(2), 119–141. https://doi.org/10.1075/gest.4.2.02mel

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

Mirman, D. (2014). *Growth curve analysis and visualization using R*. Chapman and Hall / CRC.

Mirman, D. (2015). psy811: Tools for PSY811 multilevel regression. R package version 1.0. http://github.com/dmirman/psy811.

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language, 59*(4), 475–494. https://doi.org/10.1016/j.jml.2007.11.006

Norcliffe, E., Harris, A. C., & Jaeger, T. F. (2015). Cross-linguistic psycholinguistics and its critical role in theory development: Early beginnings and recent advances. *Language, Cognition and Neuroscience, 30*(9), 1009–1032. https://doi.org/10.1080/23273798.2015.1080373

Norcliffe, E., & Konopka, A. E. (2015). Vision and language in cross-linguistic research on sentence production. In R. Mishra, N. Srinivasan, & F. Huettig (Eds.), *Attention and vision in language processing* (pp. 77–96). Springer. https://doi.org/10.1007/978-81-322-2443-3_5.

Norcliffe, E., Konopka, A. E., Brown, P., & Levinson, S. C. (2015). Word order affects the time course of sentence formulation in Tzeltal. *Language, Cognition and Neuroscience, 30*(9), 1187–1208. https://doi.org/10.1080/23273798.2015.1006238

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

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., & Slobin, D. I. (2003). Codability effects on the expression of manner of motion in Turkish and English. In A. S. Özsoy, D. Akar, M. Nakipoğlu-Demiralp, E. Erguvanlı- Taylan, & A. Aksu-Koç (Eds.), *Studies in Turkish linguistics* (pp. 259–270). Boğaziçi University Press.

Özyürek, A. (2018). Role of gesture in language processing: Toward a unified account for production and comprehension. In S. A. Rueschemeyer, & M. G. Gaskell (Eds.), *Oxford handbook of psycholinguistics* (pp. 592–607). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198786825.013.25.

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.5.1.15ozy

Papafragou, 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. (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

Papafragou, A., & Selimis, S. (2010). Lexical and structural biases in the acquisition of motion verbs. *Language Learning and Development, 6*(2), 87–115. https://doi.org/10.1080/15475440903352781

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundations for Statistical Computing. https://www.Rproject.org/.

R Core Team. (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundations for Statistical Computing. https://www.Rproject.org/.

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

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), 1–30. https://doi.org/10.1111/cogs.12708

Sauppe, S. (2017). Word order and voice influence the timing of verb planning in German sentence production. *Frontiers in Psychology, 8*. https://doi.org/10.3389/fpsyg.2017.01648. Article 1648.

Sauppe, S., Norcliffe, E., Konopka, A. E., Van Valin, R. D., & Levinson, S. C. (2013). Dependencies first: Eye tracking evidence from sentence production in Tagalog. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Meeting of the Cognitive Science Society* (pp. 1265–1270). Cognitive Science Society.

Slobin, D. I. (1996). From "thought and language" to "thinking for speaking.". In J. Gumperz, & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70–96). Cambridge University Press.

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.

Trueswell, J. C., & Papafragou, A. (2010). Perceiving and remembering events crosslinguistically: 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., Ji, Y., & Papafragou, A. (2021). From event representation to linguistic meaning. *Topics in Cognitive Science, 13*(1), 224–242. https://doi.org/10.1111/tops.12475

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

Wickham, H. (2016). ggplot2: Elegant graphics for data analysis. *Springer-Verlag.*. https://doi.org/10.1007/978-3-319-24277-4