

How does linguistic framing of events influence co-speech gestures?

Insights from crosslinguistic variations and similarities

Asli Özyürek^{1,2}, Sotaro Kita³, Shanley Allen⁴,
Reyhan Furman⁵, and Amanda Brown^{2,4}

¹F.C. Donders Centre for Cognitive Neuroimaging, Radboud University /

²Max Planck Institute for Psycholinguistics, Nijmegen / ³University of

Bristol, UK / ⁴Boston University/ ⁵Bogaziçi University, Istanbul

What are the relations between linguistic encoding and gestural representations of events during online speaking? The few studies that have been conducted on this topic have yielded somewhat incompatible results with regard to whether and how gestural representations of events change with differences in the preferred semantic and syntactic encoding possibilities of languages. Here we provide large scale semantic, syntactic and temporal analyses of speech- gesture pairs that depict 10 different motion events from 20 Turkish and 20 English speakers. We find that the gestural representations of the same events differ across languages when they are encoded by different syntactic frames (i.e., verb-framed or satellite-framed). However, where there are similarities across languages, such as omission of a certain element of the event in the linguistic encoding, gestural representations also look similar and omit the same content. The results are discussed in terms of what gestures reveal about the influence of language specific encoding on online thinking patterns and the underlying interactions between speech and gesture during the speaking process.

Keywords: speech and gesture production, cross-linguistic comparison, motion events

Introduction

When we talk about events that we have observed, we segment our continuous perception into units that are verbalizable. All languages of the world (spoken or signed) have such properties, that is, lexical elements that segment aspects of an event and sequence them in a hierarchical fashion following certain linguistic constraints. For example, to talk about someone running across the street, one has to use separate lexical items to express the figure, manner and path of motion as well as the ground on which the motion takes place, as in the sentence “the man ran across the street”.

However, languages differ with regard to which semantic elements comprising the events are readily encoded and lexicalized. For example, while German and Dutch verbs of placing readily encode the position of the object placed in the semantics of the verb (e.g., Dutch: *leggen* for horizontal placement versus *zetten* for vertical placement) (Lemmens, 2002), English does not have separate verbs that encode such distinctions, even though they could be paraphrased in several ways when necessary (i.e., *put the book on the shelf in a vertical position*) (also see Levinson & Meira, 2003, for an overview of such differences across many languages).

Another variation in event descriptions across languages concerns the way semantic elements of an event are mapped onto syntactic structures (e.g., Slobin, 1987; Talmy, 1985). For example, different languages can be classified as either “satellite-framed” or “verb-framed” depending on how path and manner components of a motion event are typically lexicalized (Talmy, 1985). Speakers of satellite-framed languages such as English express the path in a so-called satellite, like *up*, and manner in the main verb, as shown in (1). However, in verb-framed languages such as Turkish, Spanish etc., path information is expressed in the main verb and manner information outside of it. In Romance languages like Spanish, manner is frequently expressed in the form of an adverbial gerund (2) (Slobin, 1996). In other verb-framed languages such as Turkish or Japanese, manner is expressed typically in the verb of the subordinate clause rather than as an adverbial gerund, as in the Turkish sentence in (3) (Kita & Özyürek, 2003; Özyürek & Kita, 1999).

(1) *The ball rolled down the hill*

(2) *Sale volando*

Exit fly-Gerund

He/she/it exits flying

- (3) *Top zıplay-arak tepe-den aşağı in-di.*
 ball jump-Connective hill-Ablative downness descend-Past
 ‘The ball descended the hill while jumping.’

What about co-speech gestures that we use as we talk about such elements of events? Co-speech gestures are spontaneous and frequent accompaniments to speech and the expressions in the two modalities have been found to be tightly integrated pragmatically, semantically, and temporally (Clark, 1996; Goldin-Meadow, 2004; Kendon, 2004; McNeill, 1985, 1992). A subset of co-speech gestures that are frequently used in event descriptions are called *iconic* gestures (McNeill, 1992) which convey meaning by their “iconic” resemblance to the different aspects of the events they depict (e.g., wiggling fingers crossing space to represent someone walking).

The question we address in the present study is whether iconic gestures are influenced by the semantic and syntactic encoding of aspects of events during online speaking and how such influence is realized. Iconic gestures are very appropriate to investigate such a question since in their representations they involve semantic elements of event components such as figure, ground, manner and path that are also encoded by languages in different ways. Thus they allow to investigate the question of whether co-speech gestures represent elements of the events as imagined (i.e., with no influence from linguistic representations of events) or are shaped by the way event components are represented by the syntactic and semantic structures of different languages. Recent research has provided evidence for the latter claim (Kita, 2000; Kita & Özyürek, 2003; McNeill, 2000; McNeill & Duncan, 2000; Müller, 1998) showing that gestures of the same events differ across speakers of typologically different languages. These findings are compatible with the view that in utterance generation there is interaction between linguistic and gestural/imagistic representations of events (e.g., Interface Hypothesis: Kita & Özyürek, 2003; Growth Point theory: McNeill & Duncan, 2000)

However, in the literature, the type of semantic coordination between gestural and linguistic representations have been depicted in different ways. For example, McNeill and Duncan compared English and Spanish speakers’ depictions of motion events in narrations of a Sylvester and Tweety cartoon, “Canary Row” (McNeill, 2000; McNeill & Duncan, 2000). They found that Spanish speakers are more likely to omit manner in their spoken utterances about a motion event that includes manner — possibly due to the fact that manner is not encoded in the main verb and that manner verbs are not very rich in Spanish (Slobin, 1996). In such cases, Spanish speakers have been found to

express manner in a *compensatory* way in their gesture and distribute the gesture over many path phrases in speech (i.e., “manner fog”), as illustrated by individual examples. In contrast, English speakers are more likely to express manner together with path in their speech. Their gestures have been found to either express both manner and path or downplay the manner and show only path in certain discourse contexts- i.e., when manner is not the new or focus of information. Thus there is evidence that gestural representations of events compensate for those expressed in speech and provide extra information in languages where speakers are more likely to omit certain elements than other languages (i.e., Spanish versus English).

However, Kita and Özyürek (Kita & Özyürek, 2003; Özyürek & Kita, 1999) have shown other types of semantic coordination between gestural and linguistic representations of events. They compared how Japanese, Turkish, and English speakers verbally and gesturally express motion events with data taken from the narrations of the above-mentioned animated cartoon. In their studies two analyses were carried out. The first analysis concerned an event in which a protagonist swung on a rope like Tarzan from one building to another. In Turkish and Japanese, there is no readily accessible expression that semantically encodes agentive change of location of the protagonist with an arc trajectory. Whereas in English, this aspect of the event can be easily encoded with the verb *to swing*. Thus it was found that all English speakers used the verb *swing*, which encodes the arc shape of the trajectory. On the other hand, Japanese and Turkish speakers used verbs which do not encode the arc trajectory (e.g., Turkish: *gidiyor* (goes)). Paralleling these distinctions in semantic encoding, it was found that Japanese and Turkish speakers were more likely to produce a straight gesture, which does not encode the arc trajectory, than English speakers and English speakers produced only gestures with an arc trajectory.

Their second analysis concerned how speech and gesture express the manner and path of an event in which the protagonist rolled down a hill. It was found that verbal descriptions differed cross-linguistically in terms of how manner and path information is lexicalized along the lines discussed by Talmy (1985). English speakers typically used a manner verb and a path particle or preposition (e.g., *he rolled down the hill*) to express the two pieces information within one clause. In contrast, Japanese and Turkish speakers separated manner and path expressions over two clauses with path in the main clause and manner in the subordinated clause (e.g., *he descended as he rolled*). Given the assumption that a clause approximates a unit of processing in speech production (e.g., Garrett, 1982; Levelt, 1989), presumably English speakers were likely

to process both manner and path within a single processing unit, whereas Japanese and Turkish speakers were likely to need two processing units. Consequently, Japanese and Turkish speakers should be more likely than English speakers to separate the components of manner and path in preparation for speaking so that the two pieces of information could be dealt with in turn, as compared to English speakers. The gesture data confirmed this prediction. In depicting how an animated figure rolled down a hill after having swallowed a bowling ball, Japanese and Turkish speakers were more likely to use separate gestures, one for manner and one for path, and English speakers were indeed more likely to use just one gesture to express both. It was concluded from these two sets of findings that gestures encode those aspects that fit conceptualization of the event for the purposes of speaking in a particular language, rather than compensate for aspects of representations that are hard to encode linguistically or omitted for discourse purposes.¹

However, the few studies conducted on this topic have certain limitations. Some were based on individual examples of speech and gesture pairs without providing quantitative distribution from a variety of speakers of each language (McNeill & Duncan, 2000). Others were conducted on only a few motion event types and without taking the tight temporal synchrony between speech and gestures into account (Kita & Özyürek, 2003; Özyürek & Kita, 1999). In addition, the possibility could not be ruled out that the gestural differences across languages could be due to other factors (e.g., cultural differences in patterns of movement, preference for certain perspectives in depicting events, etc.) than differences in the syntactic or lexical semantic encoding of the events *per se*.

In the present study, we try to overcome the limitations of previous studies by investigating speech and gesture relations in 10 different motion event descriptions from 20 English and 20 Turkish speakers. In our analyses we compare different types of linguistic framing of events and the co-occurring gestures. First, we investigate the nature of gestural representations when speakers encode one element of the event and omit the other in their speech to see whether gestures co-occurring with these utterances compensate for the omitted information or not. Secondly, we compare gestural representations when speakers express both manner and path in their speech but differ in the way this information is encoded syntactically, that is, in one verbal clause (satellite-frame, in English) or two verbal clauses (verb-frame, in Turkish) across languages. Finally, we also analyze to what extent the expressions of manner and/or path information in gestures synchronize in time with expressions of

manner and path information in speech and whether languages differ in this respect or not.

According to Interface Hypothesis of speech and gesture production (Kita & Özyürek, 2003), there is interaction between linguistic and gestural/imagistic representations of events during speaking. That is, gestures do not merely encode imagistic representations of the event but those aspects that fit conceptualization of the event for the purposes of speaking. Gestures are shaped by the linguistic encoding choices and representations during speaking. Thus according to this hypothesis, when speakers omit one of the event components in their speech, then we expect them to omit that component in their gestures as well, regardless of the language spoken. However, in cases where speakers of languages express both event components but use different syntactic frames to do so, then we expect their gestures to also differ in order to fit the conceptualization of the event for speaking. If these predictions hold, then we argue gestures to be revealing *thinking-for-speaking* (Slobin, 1987, 1996) patterns, that is, language specific conceptualization and thinking patterns tuned and adopted for linguistic encoding choices rather than merely the spatial imagery of the event.

Method

Participants

Participants in the study were 20 monolingual Turkish and 20 monolingual English speakers. All were adults ranging in age from about 18 to 40 and were university students in either Istanbul (Turkish) or Boston (English).

Materials

Data were collected by elicitation, using a set of ten video clips depicting motion events involving simultaneous manner and path (Özyürek, Kita, & Allen, 2001). Five manners and three paths were depicted, yielding the following combinations: JUMP+ASCEND, JUMP+DESCEND, JUMP+GO.AROUND, ROLL+ASCEND, ROLL+DESCEND, ROTATE+ASCEND, ROTATE+DESCEND, SPIN+ASCEND, SPIN+DESCEND, and TUMBLE+DESCEND. The manner JUMP involves an object moving vertically up and down (moving along a flat or inclined surface), ROLL involves an object turning on its horizontal axis (moving along an inclined surface), ROTATE and TUMBLE both involve an object turning on its horizontal

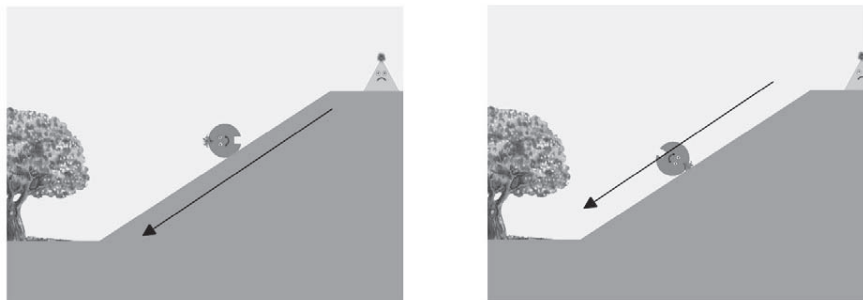


Figure 1. Selected stills of target event ROLL+DESCEND

axis (moving vertically through the air), and *SPIN* involves an object turning on its vertical axis (moving along an inclined surface).

Each video clip was between 6 and 15 seconds in duration, and had three salient components: an entry event, a target motion event, and a closing event. All clips involved a round red smiling character and a triangular-shaped green frowning character, moving in a simple landscape. We refer to them here as Tomato Man and Green Man. As an example, the *ROLL+DESCEND* clip goes as follows. The initial landscape on the screen is a large hill ending in a tree; Tomato Man is located at the top of the hill. Green Man enters the scene from the right and bumps into Tomato Man [entry event], then Tomato Man rolls down the hill [target motion event], and finally hits the tree [closing event], as illustrated in Figure 1.

Procedure

Participants were tested individually in a quiet space at their university. All interactions were videotaped for later coding and analysis. During the warm-up phase, the experimenter showed participants a typical scene from a clip and introduced them to the characters and the landscape. She then gave them two practice rounds with clips involving motion events like those in the test clips and asked them to tell what happened in the clip to a listener who purportedly had not seen it.

In the testing phase, the experimenter presented the ten test clips for the participant to narrate, following the same format as in the warm-up phase. If participants did not mention the target event in their narration, either the experimenter or the listener encouraged them to do so with a question that did not focus explicitly on either manner or path.

Transcription and Coding

Speech Transcription

All speech relevant to the target motion events was transcribed by native speakers of the relevant language into MediaTagger, a video-based computer program (Brugman & Kita, 1995). Note that we did not transcribe any speech that described exclusively the entry event, the closing event, or the setting of the scene. The relevant speech for each participant was segmented into “sentences,” which we define here as a matrix clause plus its subordinates, if any. Examples are shown in (4 and 5), with sentence segmentations indicated by square brackets.

- (4) English:
 a. [He rolled up the hill.]
 b. [And he is spinning as he goes down the hill.]
 Turkish:
 c. [Domates adam yuvarlan-arak tepe-yi çıkı-yo]
 tomato man roll-Connective hill-Accusative ascend-Present
 ‘tomato man ascends the hill while rolling’

Two matrix clauses separated by a coordinating conjunction (i.e., *and*, *but*, and *or* for English, and *ve*, *sonra* for Turkish) were as considered as two sentences. Many participants used more than one sentence to describe a given target motion event. We refer to the full set of sentences that describe a particular target motion event as a “target-event description.”

The description for the ROTATE+DESCEND clip given in (5) illustrates this process. The speaker uttered everything in (5) as his description of the clip. Only the portion describing the target motion event of the Tomato Man rotating down into the water (i.e., the target-event description) was transcribed into MediaTagger, as indicated in (5) by curly brackets. The target-event description was then divided into three sentences, as indicated by square brackets.

- (5) *There’s a ledge on the right and Triangle Man is floating in the water on the left. Tomato Man slides off sort of Wile E. Coyote style, where he doesn’t just fall straight off, but goes about halfway in the air {[and then falls down]. [So he spins down,] [spins down]} and lands next to Triangle Man.*

In order to establish reliability of the identification and segmentation of sentences, twenty percent of the data were independently processed by a second coder who was either a fluent or native speaker of the relevant language. For each clip, the second coder identified the stretch of discourse describing the

target event and segmented it into sentences. The percentages of the original coder's sentences with which the second coder agreed in terms of identification and segmentation of target event sentences were as follows: for English, 92% and for Turkish, 88%.

Speech Coding

Each sentence was coded by native speakers of the relevant language as one of four categories according to the structural patterns of information packaging in speech relating to manner and path. In this coding, manner refers to the secondary movement (rotation along different axes, or jumping) of the figure that co-occurs with the translational movement in the target events. Path refers to the directionality or trajectory specifications for the translational movement.

Some sentences included only one of the motion elements such as manner or path. The first category, "Manner-only," denotes use of only a manner element in the sentence (i.e., no path). Sentences coded as Manner-only in English include simple manner verbs (6a), manner verbs with or without some further description of the manner, and phrases which describe the manner without a manner verb (6b).

- (6) a. *And then tumbles head over heels.*
 b. *And does a little couple of rounds.*

Turkish Manner-only sentences include constructions similar to the English ones shown in (6a), but nothing like that in (6b).

The next category, "Path-only," indicates use of only a path element (i.e. no manner) in the sentence. In English, sentences coded as Path-only include the light path verb *go* followed by directional path particles or adpositional phrases (7a), or other path verbs optionally followed by directional path particles or adpositional phrases (7b).

- (7) a. *He goes up a hill.*
 b. *It fell.*

In Turkish, sentences coded as Path-only include light path verbs (*come* and *go*) as in (8a) and other path verbs as in (8b), both with optional postpositional phrases that include spatial nouns specifying the source or the goal of the path.

- (8) a. *Aşağı-ya gel-iyor.*
downness-Dative come-Present
'(He/she/it) comes down.'
b. *Sonra yukarı çık-tı.*
then upness ascend-Past
'Then (he/she/it) ascended (to) the upness.'

For sentences in which both manner and path were mentioned, two coding categories were distinguished. The category “Tight” denotes a tight packaging of both manner and path in one sentence, that is, a unit involving one verb and one closely associated non-verbal phrase. Sentences coded as Tight differ somewhat across languages. English Tight sentences include manner verbs followed by directional path particles or prepositional phrases (9a), and phrases describing manner followed by a directional path particle or prepositional phrase (9b).

- (9) a. *He rolled up the hill.*
b. *And he did his little two-step down the hill.*

Tight sentences also occur in Turkish, although they were rarely used. A typical example of this includes a manner verb with a postpositional directional path phrase, but crucially no path verb (10).

- (10) *Domates adam aşağı yuvarlan-ıyor tepe-den.*
tomato man downness roll-Present hill-Ablative
'Tomato Man rolls down the hill.'

The second category of sentences in which both manner and path were mentioned is labeled “Semi-tight.” This code denotes a semi-tight packaging of manner and path in one sentence, with each of these expressed by a separate verbal element, one subordinated to the other. In English, the subordinated form can be either a fully tensed verb (11a) or a progressive participle functioning as an adverbial (11b).

- (11) a. *He spins in circles while he's going down.*
b. *Triangle Man ascends the hill twirling.*

In Semi-tight constructions in Turkish, the manner verb is subordinated to the main path verb with the use of a connective — mostly *-arak* as in (12a), and very rarely *-ip*. Another possibility is to use a reduplicated manner verb functioning as an adverbial and subordinated to the main path verb, as in (12b).

- (12) a. *Domates adam yuvarlan-arak yokuş-u in-di.*
 tomato man roll-Connective hill-Accusative descend-Past
 ‘Tomato man descended the hill while rolling.’
- b. *Üçgen dönne-dönne çık-tı.*
 triangle turning-turning ascend-Past
 ‘Triangle ascended turning turning.’

Finally, sentences which included more than one type of packaging of manner and path were coded for each relevant type. For example, the sentence in (13) was coded as both Path-only and Tight.

- (13) *When he went down, he was spinning down.*

In order to establish reliability of the coding, twenty percent of the data were independently processed by a second coder who was either a fluent or native speaker of the relevant language. The second coder judged the category type (i.e., Manner-only, Path-only, Tight, Semi-tight) for each sentence that had been segmented and transcribed by the original coder. The agreement between coders for this judgment was as follows: for English: 93%, and for Turkish: 91%.

Gesture transcription

We transcribed gestures that occurred concurrent with sentences in the target event descriptions that contained manner and/or path. The stroke phase of gestures (Kendon, 1980; McNeill, 1992) was isolated by frame by frame video analysis, according to the procedure detailed in Kita, van Gijn, and van der Hulst (1998). We excluded gestures that overlapped with more than one utterance type (e.g., Tight and Path only) or non target-event utterances (see speech transcription above).

Gesture coding

Gestures that encoded the manner and/or path of the target event were called *target-event gestures*. They were classified into five types: *Manner*, *Path*, *Conflated*, *Combined* or *Unclear*. Manner gestures encoded manner of motion (e.g., a repetitive up and down movement of the hand to represent jumping) without encoding path. Path gestures expressed change of location without encoding manner. Conflated gestures expressed both manner and path at the same time all throughout the stroke (e.g., repetitive up and down movements superimposed on diagonal downward change of location of the hand, representing jumping down the slope). Combined gestures were two-handed gestures, in

which each hand was a different type (e.g., one hand was Manner, the other hand was Conflated). Finally some gestures were coded as Unclear either because they were hard to segment, hard to categorize for any of the two categories above or were unclear with regard to whether they were representational gestures or just self-adaptors.

For purposes of clarity we excluded gestures that were Combined or Unclear from the analysis. We also excluded the few gestures that included the use of body to represent change of location or manner (e.g., trying to show rotation using mainly head, shoulders and arms) since these representations bias towards Manner only and thus do not allow comparison for the frequency of the use of Conflated gestures.

In order to establish reliability of the identification and segmentation of target-event gestures, twenty percent of the data was independently processed by a second coder. For each clip, the second coder identified target-event gestures in the discourse, and segmented the stroke phase of the gestures. 81% of the original coder's gesture strokes ($N = 108$) had an overlap with a gesture stroke identified and segmented by the second coder. Among these gestures, the discrepancy between the two coders was on average 1.72 frames ($SD = 2.02$) (1 frame = 33.3ms) at the beginning of the stroke and 2.54 frames ($SD = 4.74$) at the end. Among the gesture strokes that were identified by both coders, 90% of the original coder's strokes overlapped with a stroke coded by the second coder coded with the discrepancy of 5 video frames (167ms) or less at the beginning and the end of the stroke.

Furthermore, in order to establish reliability of the gesture type classification, the second coder judged the gesture type (i.e., Manner, Path, Conflated, etc.) for each target-event gesture stroke that had been identified and segmented by the original coder. The agreement between coders was 89% for the gesture type.

Results

Speech

Since our main investigation is to see whether gestural representations of the same events differ with different syntactic constructions across languages and how, first we analyzed the differences in speech between the languages. First, we investigated to what extent speakers of different languages were more likely to mention only one of the elements (i.e., either path or manner) in their

Table 1. Average proportion of sentences that express different types of manner and path information across languages

	Both Manner and Path	Manner Only	Path Only	Total N of sentences
English	0.62 (0.16)	0.12 (0.09)	0.26 (0.12)	315
Tight	0.52 (0.16)			
Semi-tight	0.07 (0.07)			
Turkish	0.65 (0.18)	0.07 (0.08)	0.28 (0.14)	294
Tight	0.02 (0.01)			
Semi-tight	0.63 (0.19)			

Note: Parentheses indicate standard deviations

sentences about the target events. Secondly, we focused on only those sentences in which both manner and path were expressed and compared the type of syntactic packaging (i.e., Tight or Semi-tight) that each group of participants used in their descriptions.

Inclusion of manner and/or path across languages: The first analysis showed that the proportion of sentences that included both types of information (i.e., path and manner) was similar across languages; English, $M = 0.62$, $SD = 0.16$, and Turkish, $M = 0.65$, $SD = 0.19$. In the rest of the sentences either Path or Manner was omitted. To compare whether languages differed in terms of using Path-only or Manner-only sentences, a 2×2 repeated measures ANOVA with language type (Turkish and English) as the between subjects variable and information type (Manner-only or Path-only) as the within subjects variable was conducted. There was a main effect of information type ($F(1,38) = 50.991$; $p < 0.001$), but no interaction with language. This showed that both Turkish and English speakers used a similar amount of Path-only or Manner-only sentences (see Table 1). However, all speakers preferred Path-only sentences to Manner-only sentences, regardless of the language they were speaking, as evidenced by the main effect.

Differences in syntactic packaging between the languages: In the next analysis the proportion of sentences in which both manner and path are expressed were compared in terms of the type of syntactic packaging preferred in each of the two languages. A 2×2 repeated measures ANOVA with language (Turkish versus English) as the between subjects variable and syntactic packaging type (Tight versus Semi-tight) as the within subjects variable revealed an interaction between language and syntactic type ($F(1,38) = 269.69$, $p < 0.001$). As expected from the typological differences, further Bonferroni adjusted t tests revealed that English speakers used more Tight syntactic packaging than Turkish

speakers ($t(38) = -13.4, p < 0.001$) and Turkish speakers used more Semi-tight packaging than English speakers ($t(38) = 11.76, p < 0.001$) (see Table 1).

To summarize, the speech analysis showed that the speakers' choices of focusing on one element of the event was similar across the languages. However, when they expressed both components, English and Turkish speakers used different syntactic framing for expressing event components. Next we investigated how gestural representations co-occurred with these different linguistic choices.

Gesture

In the gesture analysis, we first looked at how gestures represented information when either the path or the manner element was omitted from the sentence. Secondly, we analyzed whether the differences in syntactic packaging of information influenced the information packaging of information in gesture when both manner and path were expressed in speech.

All the analyses were conducted on the proportions of gesture types out of all the gestures that co-occurred temporally with different sentence types (e.g., Path-only, Manner-only, Tight, Semi-tight etc.) for each subject.

Relations between the informational content of gestures and speech: One crucial question is when speakers express only one piece of information in their sentences (i.e., in Manner-only and Path-only sentences), what does the content of their gestures look like? Do they compensate for the omission of information in speech or do they also omit that same information in gestures? To investigate this question, we analyzed the type of gestures that accompanied Manner-only and Path-only sentences.

Gestures in the context of Path-only sentences: In this analysis, only gestures that accompanied Path-only sentences were included. Each subject had to contribute with at least 5 gestures for his/her data to be included in this analysis to have enough variance for the statistical comparisons. A total of 11 English and 7 Turkish speakers were included in the analysis. The proportions of each type of gesture out of all the gestures that occurred during Path-only sentences were calculated (see Figure 2). A 2*3 repeated measures ANOVA with language as the between subjects variable and gesture type (Manner, Path or Conflated) as the within subjects variable revealed only an effect of gesture type ($F(2,32) = 217.07; p < 0.001$). That is, speakers of English and Turkish did not differ in terms of the type of gesture they preferred when they expressed only path information in speech. Further Bonferroni adjusted t tests among the gesture

types revealed that speakers of both languages preferred Path over Manner ($t(17) = -18.7, p < 0.001$) or Conflated gestures ($t(17) = -14.07, p < 0.001$) when they expressed only path information in their speech.

Gestures in the context of Manner-only sentences: In this analysis, only gestures that accompanied Manner-only sentences were included. Due to the fact that there were many fewer Manner-only sentences, each subject had to contribute at least 3 gestures to be included in this analysis. A total of 8 English speakers and 4 Turkish speakers were included in the analysis. A 2×3 repeated measures ANOVA with language as the between subjects variable and gesture type (Manner, Path or Conflated) as the within subject variable revealed only an effect of gesture type ($F(2,20) = 4.58; p < 0.05$). That is, speakers of English and Turkish did not differ in terms of the type of gesture they preferred when they expressed only manner information in speech. Further Bonferroni adjusted t tests among the gestures types revealed that speakers of both languages used more Manner ($t(11) = 3.40; p < 0.01$) and Conflated ($t(11) = 2.79; p < 0.05$) gestures than Path gestures when they expressed only manner information in their sentences (see Figure 3).

In sum, the analysis showed that the information expressed both in gesture and speech showed strong parallels regardless of the language spoken. That is, when speakers of both languages expressed only path in their speech they were more likely to use Path gestures. Likewise when they expressed only manner in

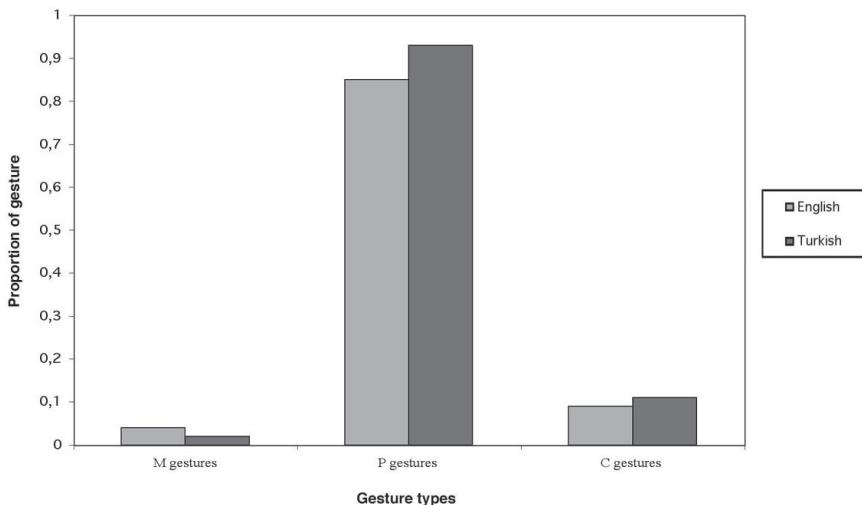


Figure 2. Proportions of gesture types (Manner, Path, Conflated) accompanying Path-only sentences across languages

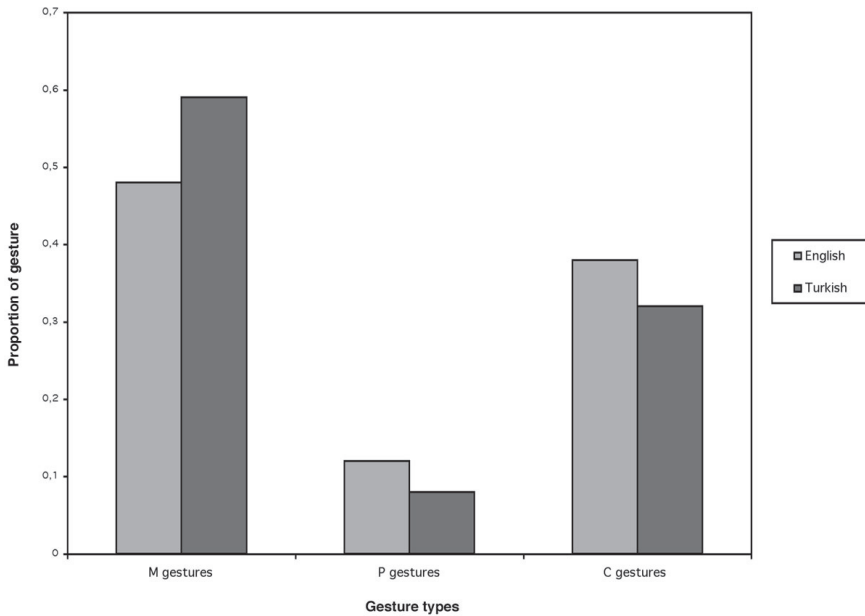


Figure 3. Proportions of gesture types (Manner, Path, Conflated) accompanying Manner-only sentences across languages

their speech, they included gestures that contained manner (both Manner and Conflated gestures), but crucially not Path gestures that would mismatch, or compensate the informational content of the utterance.

Gestures accompanying different syntactic packaging of manner and path information: In this analysis, only gestures that accompanied Tight sentences in the English sample and Semi-tight sentences in the Turkish sample were included. Each subject had to contribute at least 5 gestures to be included in this analysis. A total of 18 English speakers and 20 Turkish speakers were included in the analysis. A 2*3 repeated measures ANOVA with language (Turkish and English) as the between subjects variable and the gesture type (Manner, Path or Conflated) as the within subject variable revealed interaction between the two factors ($F(2,72) = 19.33; p < 0.001$). Bonferroni adjusted t tests showed that English speakers used more Conflated gestures than Turkish speakers ($t(36) = 5.55; p < 0.001$). On the other hand Turkish speakers used more Manner ($t(36) = -3.14; p < 0.05$) and Path ($t(36) = -3.4, p < 0.05$) gestures than English speakers (see Figure 4).

This analysis shows that in addition to the informational content, syntactic packaging of information also influences the type of gestural representations.

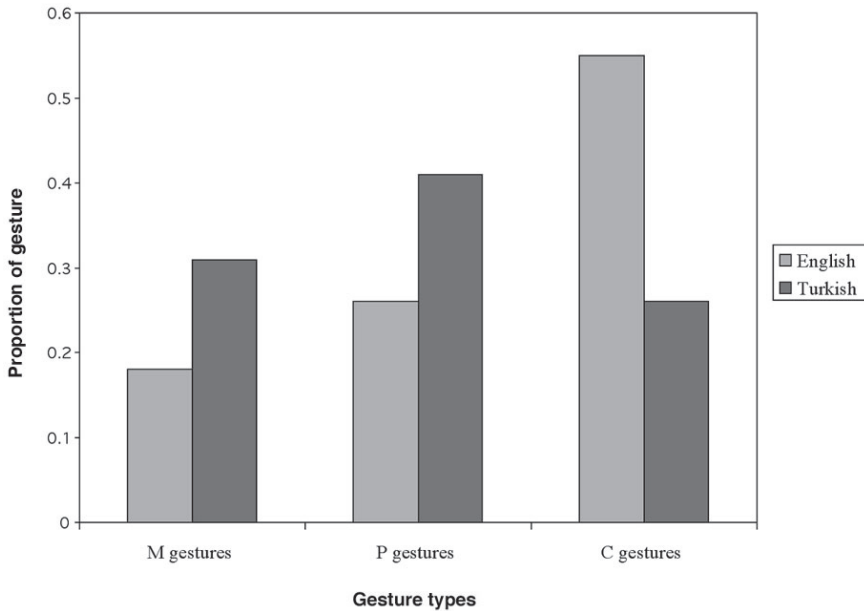


Figure 4. Proportions of gesture types (Manner, Path, Conflated) accompanying Tight and Semi-tight sentences across languages

Namely, gestures reveal similar representations to the linguistic encoding of events.

Temporal synchrony of information between gestures and speech: In the final analysis we investigated the tight temporal synchrony between the types of gestures and element of speech they co-occurred with in the sentence. If gestures are influenced by the online conceptualization of events during online syntactic and semantic encoding, we expect them also to be tightly synchronized with the speech segment they co-occur with. We tested this hypothesis with Manner and Path gestures since Conflated gestures could overlap with speech that could express both manner and path. We calculated the proportion of Path gestures that co-occurred with path speech (i.e., with path verbs, path particles, and prepositional phrases). Note that path gestures were not included in the “path matching” category if they co-occurred exclusively with manner speech, expression of figure (e.g., tomato man) or discourse markers, thus without overlapping with any path speech. Here, if any part of the stroke phase of the Path gesture overlapped with the relevant path speech, we considered it as synchronous. Likewise, the proportion of manner gestures that overlapped with manner speech (i.e., manner verbs, manner elaborations) were calculated.

The percentage of Path gestures that co-occurred with path speech was 82% for Turkish and 75% for English. For Manner gestures, the co-occurrence rates with manner speech were 86% for Turkish and 65% for English. That is, for both languages, a majority of Path and Manner gestures overlapped with immediately relevant speech. Chi square tests conducted on the number of manner gestures ($X^2 = 2,00$ $df = 1$, $p = 0.15$) and path gestures ($X^2 = 2,00$ $df = 1$, $p = 0.16$) that overlapped with relevant speech did not reveal significant differences between the languages in this respect. That is, in both languages, gestures overlapped with the speech that they semantically coordinate with to a similar extent.

Conclusion and Discussion

The aim of the present study was to investigate whether and how the linguistic framing of events influences gestural representations of the same events during online speaking. We attempted to go beyond the few studies that have investigated this question by providing data from a large sample, comparing two typologically different languages, English and Turkish, which use different linguistic framing for motion events (i.e., satellite-framed versus verb-framed). We also conducted tight semantic as well as temporal analyses for a large sample of gesture and speech pairs depicting various motion events from speakers of the two languages.

The results showed that in cases where there are differences in the semantic and syntactic encoding of motion event elements, gestural representations vary in ways that fit the language specific encoding differences. This was illustrated by the differences in gesture types between Turkish and English speakers' gestures that overlap with Tight versus Semi-tight speech. Specifically when speakers used one verbal clause they preferred to use one gesture that expressed both elements and when they used two verbal clauses, they were more likely to use separate gestures for manner and path. These results support previous findings (Kita & Özyürek, 2003; Özyürek & Kita, 1999), extending their generalizability to various motion event types and also providing further evidence from temporal synchrony analysis.

However, in cases where there were no language specific differences, that is, in terms of the use of Manner-only and the Path-only sentences, the gesture patterns of the speakers of the two languages looked alike. Here gestural information was found to fit the semantic encoding of the event rather than

compensate or convey meaning not expressed by speech. This pattern was found to be the same in the two languages. Note that these results are unlike found for Spanish in the McNeill and Duncan (2000) study where manner gestures were found to compensate for the omission of manner content in speech. It would be revealing if a similar large scale study were conducted between Spanish and English to investigate the information coordination between speech and gesture.

The differences in the distributions of gesture types in different sentence contexts (i.e., Figures 2, 3, and 4) reveal that the differences in gesture types across languages found both in the current study as well as in Kita & Özyürek (2003) can not be explained merely due to cultural factors. Namely, it is not the case that Turkish speakers have a general preference for Manner and Path gestures but not for Conflated gestures compared to English speakers, regardless of the content of the concurrent speech. Rather, they show that gestural differences between English and Turkish speakers (Figure 4) could be directly attributed to the online choice of different semantic and syntactic encoding patterns, since gestural differences wash out when both Turkish and English speakers choose to express either only path or only manner in their speech.

Finally, the temporal synchrony analysis is in line with the rest of the findings in the sense that the information coordination between the two modalities is also reflected in the temporal synchrony of semantic information in the two channels regardless of the language spoken (McNeill, 1992). However, it is important to note that in both languages, non-typical alignments were also observed in around 25% of the cases. Further research is necessary to investigate the nature of such combinations, what they reveal, and whether they show different distributions across languages.

Further analysis is also needed to investigate how the discourse level encoding of information interacts with the analyses we have provided here. It is possible that the gestural coordination of information might be further sensitive to whether the manner and/or path information was new or old in discourse as shown in the McNeill and Duncan's (2000) analysis.

Overall, the findings of this study show that even though there are differences in the way gestures encode same events across languages, this can be explained by one and similar process underlying speech and gesture production in speakers of different languages. That is, during online speaking gestural and linguistic representations interact in such a way that gestures reflect the underlying online conceptualization that fits appropriate semantic and syntactic encoding of events. In this paper we attempted to unpack the nature of

this multi-modal semantic information coordination at the sentence level and found it to be similar across speakers of different languages, at least for the two typologically different languages we have studied at length.

Acknowledgments

This research was supported by grant BCS-0002117 from the National Science Foundation to the first three authors, and by a Turkish Academy of Sciences (TUBA) Young Scientist Award to Özyürek. Substantial logistical support was provided by the Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands, and by Koç University, Istanbul, Turkey. Audiences at the Max Planck Institute for Psycholinguistics offered helpful comments and discussion on previous versions of this paper. Tilbe Göksun helped in establishing reliability.

Note

1. This is not to say that gestures will not depict any representations not expressed in speech. The point here is that their representations will be influenced by the linguistic encoding possibilities in speech.

References

- Brugman, Hennie & Sotaro Kita (1995). Impact of digital video technology on transcription: a case of spontaneous gesture transcription. *KODIKAS/CODE: Ars Semeiotica, An international journal of semiotics*, 18, 95–112.
- Clark, Herbert (1996). *Using language*. Cambridge: Cambridge University Press.
- Garrett, Merrill (1982). Production of speech: Observations from normal and pathological language use. In Willem Ellis (Ed.), *Normality and pathology in cognitive functions* (pp. 19–76). London: Academic Press.
- Goldin-Meadow, Susan (2004). *Hearing gesture: How our hands help us think*. Cambridge: Harvard University Press.
- Kendon, Adam (1980). Gesticulation and speech: Two aspects of the process of utterance. In Mary R. Key (Ed.), *The relation between verbal and nonverbal communication* (pp. 207–227). The Hague: Mouton.
- Kendon, Adam (2004). *Gesture. Visible action as utterance*. Cambridge: Cambridge University Press.
- Kita, Sotaro (2000). How representational gestures help speaking. In David McNeill (Ed.), *Language and gesture* (pp. 162–185). Cambridge: Cambridge University Press.
- Kita, Sotaro, Ingeborg van Gijn, & Harry van der Hulst (1998). Movement Phases in signs and co-speech gestures, and their transcription by human coders. In Ipke Wachsmuth

- & Martin Fröhlich (Eds.), *Gesture and sign language in human-computer interaction, International Gesture Workshop Bielefeld, Germany, September 17–19, 1997, Proceedings*. Lecture Notes in Artificial Intelligence, Volume 1371 (pp. 23–35). Berlin: Springer-Verlag.
- Kita, Sotaro & Asli Özyürek (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, 16–32.
- Lemmens, Maarten (2002). The semantic network of Dutch posture verbs. In John Newman (Ed.), *The linguistics of sitting, standing, and lying* (pp. 103–139). Amsterdam: Benjamins.
- Levinson, Stephen & Sergio Meira (2003). ‘Natural concepts’ in the spatial topological domain — adpositional meanings in cross-linguistic perspective: An exercise in semantic typology. *Language*, 79(3), 485–516.
- Levelt, Pim (1989). *Speaking*. Cambridge: MIT Press.
- McNeill, David (1985). So you think gestures are nonverbal? *Psychological Review*, 92, 350–371.
- McNeill, David (1992). *Hand and mind*. Chicago: University of Chicago Press.
- McNeill, David (2000). Analogic/analytic representations and cross-linguistic differences in thinking for speaking. *Cognitive Linguistics*, 11 (1/2), 43–60.
- McNeill, David & Susan Duncan (2000). Growth points in thinking-for-speaking. In David McNeill (Ed.), *Language and gesture* (pp. 141–161). Cambridge: Cambridge University Press.
- Müller, Cornelia (1998). *Redebegleitende Gesten: Kulturgeschichte-Theorie-Sprachvergleich*. Berlin: Berlin Verlag Arno Spitz.
- Özyürek, Asli & Sotaro Kita (1999). Expressing manner and path in English and Turkish: Differences in speech, gesture, and conceptualization. In Martin Hahn & Scott Stoness (Eds.), *Proceedings of the twenty first annual conference of the Cognitive Science Society* (pp. 507–512). Mahwah, NJ: Lawrence Erlbaum.
- Özyürek, Asli, Sotaro Kita, & Shanley Allen (2001). *Tomato man movies: Stimulus kit designed to elicit manner, path and causal constructions in motion events with regard to speech and gestures*. Nijmegen, The Netherlands: Max Planck Institute for Psycholinguistics, Language and Cognition group.
- Slobin, Dan (1987). Thinking for speaking. In Jon Aske, Natasha Beery, Laura Michaelis, & Hana Filip (Eds.), *Proceedings of the 13th annual meeting of the Berkeley Linguistic Society* (pp. 435–445). Berkeley: Berkeley Linguistic Society.
- Slobin, Dan (1996). From “thought and language” to “thinking for speaking”. In John Gumperz & Stephen Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70–96). Cambridge: Cambridge University Press.
- Talmy, Len (1985). Semantics and syntax of motion. In Timothy Shopen (Ed.), *Language typology and syntactic description, Vol.3, Grammatical categories and the lexicon* (pp. 57–149). Cambridge: Cambridge University Press.

Authors' address

Asli Özyürek
Max Planck Institute for Psycholinguistics
Wundtlaan 1, 6525 XD
Nijmegen
The Netherlands
Email: asliozu@mpi.nl

About the authors

Asli Özyürek, Ph.D., is a research scientist at the FC Donders Center for Cognitive Neuroimaging and Max Planck Institute for Psycholinguistics, Nijmegen. She does research on relations between speech and gesture in production and comprehension as well as on sign languages and gesture systems of “homesigner” children. She is also interested in the relations between language and conceptualization and what gestures, sign languages and home-sign systems reveal about this relation.

Sotaro Kita, Ph.D., is a Senior Lecturer at the Department of Experimental Psychology at the University Bristol. His main research interests are cognitive psychological, interactional, and ethnographic studies of the relationship between speech and spontaneous co-speech gestures. In addition his research interests include child language acquisition, semantics and pragmatics of spatial expressions, and cross-linguistic studies of spatial conceptualization.

Shanley Allen, Ph.D., is an Associate Professor in the Program in Applied Linguistics and the School of Education at Boston University. Her research explores the first language acquisition of morphology and syntax, with a focus on comparing acquisition patterns across languages, as well as with children learning Inuktitut (Eskimo) in northern Canada. Her other interests include bilingual acquisition, specific language impairment, and the acquisition of co-speech gesture.

Reyhan Furman, MA is a doctoral student at the Linguistics Department, Bogazici University, Istanbul. Her research focuses on the event structure representations of monolingual and bilingual children and adults, in language and co-speech gestures. She is also interested in children's acquisition of verb argument structure and the acquisition of complex constructions.

Amanda Brown, MA, is a doctoral student in the Program in Applied Linguistics at Boston University and the Max Planck Institute for Psycholinguistics, Nijmegen. Her dissertation research investigates bilateral interactions between established and emerging language systems using analyses of speech and co-speech gestures in Japanese speakers of English.