

# Development of Cross-Linguistic Variation in Speech and Gesture: Motion Events in English and Turkish

Aslı Özyürek

Radboud University Nijmegen and Max Planck  
Institute for Psycholinguistics

Sotaro Kita

University of Birmingham

Shanley Allen

Boston University

Amanda Brown

Syracuse University

Reyhan Furman

Radboud University Nijmegen and  
Boğaziçi University

Tomoko Ishizuka

University of California at Los Angeles

The way adults express manner and path components of a motion event varies across typologically different languages both in speech and cospeech gestures, showing that language specificity in event encoding influences gesture. The authors tracked when and how this multimodal cross-linguistic variation develops in children learning Turkish and English, 2 typologically distinct languages. They found that children learn to speak in language-specific ways from age 3 onward (i.e., English speakers used 1 clause and Turkish speakers used 2 clauses to express manner and path). In contrast, English- and Turkish-speaking children's gestures looked similar at ages 3 and 5 (i.e., separate gestures for manner and path), differing from each other only at age 9 and in adulthood (i.e., English speakers used 1 gesture, but Turkish speakers used separate gestures for manner and path). The authors argue that this pattern of the development of cospeech gestures reflects a gradual shift to language-specific representations during speaking and shows that looking at speech alone may not be sufficient to understand the full process of language acquisition.

**Keywords:** cospeech gestures, motion events, cross-linguistic, Turkish, English

Different languages have different ways of distributing features of the same spatial information across linguistic units (e.g., Slobin,

1987; Talmy, 1985). One of the challenges for the field of language development has been to explain how children born to different languages learn language-specific ways of encoding spatial information. Previous research has shown that children's early spatial expressions largely follow the language-specific distinctions of their adult counterparts (e.g., Allen et al., 2007; Choi & Bowerman, 1991; Özçalışkan & Slobin, 1999), demonstrating that children are tuned to the language-specific semantic distinctions of their languages early on. In addition, some universal preferences for linguistic encoding of spatial information have been posited (Allen et al., 2007; Bowerman, 1982; Johnston & Slobin, 1979), suggesting that language learning can also be guided by cognitive prerequisites or linguistic defaults.

These studies have focused mostly on children's speech patterns. However, the ability to represent and communicate about space is not limited to verbal expressions. Speakers often gesture as they speak, especially when they talk about space (Rauscher, Krauss, & Chen, 1996). Recent research has shown that speech and cospeech gesture form an integrated system (Bernardis & Gentilucci, 2006; Clark, 1996; Kendon, 2004; McNeill, 1992, 2005) and that they develop in close relation to each other during childhood (Bates, 1976; Capirci, Iverson, Pizzuto, & Volterra, 1996; McNeill, 1992, 2005; Nicoladis, Mayberry, & Genesee, 1999; Özçalışkan & Goldin-Meadow, 2005). Cospeech gestures have also been found to reflect children's representations not necessarily expressed in speech at certain stages of development (Alibali &

---

Aslı Özyürek, Center for Language Studies, Department of Linguistics, Radboud University Nijmegen, Nijmegen, The Netherlands, and Max Planck Institute for Psycholinguistics, Nijmegen; Sotaro Kita, School of Psychology, University of Birmingham, Birmingham, United Kingdom; Shanley Allen, Department of Literacy and Language, Counseling and Development, Boston University; Amanda Brown, Department of Languages, Literatures and Linguistics, Syracuse University; Reyhan Furman, Center for Language Studies, Department of Linguistics, Radboud University Nijmegen, and Department of Linguistics, Boğaziçi University, Istanbul, Turkey; and Tomoko Ishizuka, Department of Linguistics, University of California at Los Angeles.

This research was financially supported by National Science Foundation Grant BCS-0002117, awarded to Aslı Özyürek, Sotaro Kita, and Shanley Allen; the Max Planck Institute for Psycholinguistics; and the Turkish Academy of Sciences. We thank Koç University and Boston University and schools and preschools in Istanbul, Turkey, and Boston for their cooperation, as well as child and adult participants for their help. We are also grateful to audiences at the Nijmegen Gesture Centre in the Netherlands; the International Congress for the Study of Child Language, held in Madison, Wisconsin; the University of Connecticut; and the University of Chicago for helpful discussion of the ideas presented here.

Correspondence concerning this article should be addressed to Aslı Özyürek, Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD, Nijmegen, The Netherlands. E-mail: asliozu@mpi.nl

Goldin-Meadow, 1993; Ehrlich, Levine, & Goldin-Meadow, 2006; Pine, Nicola, & Messer, 2004). Moreover, gestures vary across adult speakers of different languages, reflecting language-specific encoding of space in typologically different languages (Kita & Özyürek, 2003; McNeill & Duncan, 2000; Müller, 1998; Özyürek, Kita, Allen, Furman & Brown, 2005). Yet we know little about how cross-linguistic variation in gesture develops in relation to that in speech. The way children use cospeech gestures may offer a secondary window into how they learn the language-specific distribution of spatial information into linguistic units and may provide additional cues to understanding the overall process of language acquisition.

We investigate how children use their speech and gestures to express motion events when they learn typologically different languages, here Turkish and English. Previous research has shown that adult speakers of these languages represent components of motion in different ways both in their speech and gesture (Kita & Özyürek, 2003; Özyürek et al., 2005). The specific question we pose in this article is whether the speech and gesture of English- and Turkish-speaking children (ages 3, 5, and 9) pattern in adult-like ways from the earliest age or whether language specificity develops over time.

### Development of Linguistic Encoding of Motion Events Cross-Linguistically

Languages differ in the way that elements of a motion event are packaged into different syntactic units (Talmy, 1985). Speakers of so-called satellite-framed languages such as English tend to conflate the semantic elements motion and manner in the main verb (e.g., *run*), and express the path in a nonverbal element, namely a path particle or “satellite” (e.g., *down*), with all three elements expressed together in one verbal clause (e.g., *he rolled down*). In contrast, speakers of verb-framed languages such as Turkish tend to conflate motion with path in the main verb (e.g., *in-* [descend]) and express manner in a subordinate verb (e.g., *koş* [run]), using two verbal clauses to express the three semantic elements (e.g., *adam koşarak tepeden indi* [the man descended the hill while running]; Allen et al., 2007).

Children are able to extract individual elements of manner and path components of an event in nonlinguistic tasks very early on (e.g., between 7 and 15 months in Pulverman, Stootsman, Golinkoff, & Hirsh-Pasek, 2003). Furthermore, they tune their linguistic patterns to the constructions in their target languages quite early. Özçalışkan and Slobin (1999), using narrations of a wordless picture storybook, have shown that as early as 3 years of age, children learning the verb-framed languages Turkish and Spanish use more path verbs in their speech (e.g., *exit*), whereas children learning the satellite-framed language English use more manner verbs (e.g., *fly*), reflecting the differences between adult speakers of these languages. Similar patterns have been found experimentally for 3-year-old children learning English versus Korean (Oh, 2003), for slightly older child (age 4–12) and adult speakers of English and Greek (Papafragou, Massey, & Gleitman, 2002), and in spontaneous speech data for very young speakers of English and Korean (Choi & Bowerman, 1991).

The above-mentioned studies have mostly focused on speakers’ preferences for manner versus path verbs as a consequence of differences in lexicalization patterns across languages. However, a

few studies have also looked systematically at how adults and children package manner and path syntactically when both elements are salient and need to be expressed (Allen et al., 2007; Oh, 2003). Allen et al. (2007), using narrations of short animated clips, have shown that 3-year-old English-speaking children package both manner and path in one clause more often than same-age Turkish-speaking children, whereas Turkish-speaking children use two clauses more often than English-speaking children, each reflecting cross-linguistic differences in adult patterns. However, Turkish-speaking children, unlike their adult counterparts, use one-clause constructions to talk about both manner and path 20 percent of the time, suggesting that early children’s speech shows universal as well as language-specific preferences for clausal packaging of information.

In sum, previous cross-linguistic research on speech has established that children mainly exhibit language-specific encoding of motion from an early age, although some similarities across child speakers of different languages may also exist. However, very little is known about the way children born to different languages encode motion in gesture in relation to speech at different ages.

### Gestural Expressions of Motion Events Cross-Linguistically

Cospeech gestures also represent elements of motion events such as figure, ground, manner, and path. These are evident in a subset of gestures called “iconic gestures,” which convey meaning by their iconic resemblance to certain aspects of the events and objects they depict (e.g., wiggling fingers across space to represent someone walking; McNeill, 1992). Recent research has revealed cross-linguistic differences in iconic gestures depicting motion events. In a study of adult English, Turkish, and Japanese speakers narrating a cartoon, Kita and Özyürek (2003) demonstrated that gestures representing the manner and path components of a single motion event parallel the way information is packaged at the clause level in each language. Hence, Japanese and Turkish speakers were more likely to use separate gestures for each element, whereas English speakers were more likely to use just one gesture that conflated both elements. These differences were replicated with Turkish- and English-speaking adults in descriptions of 10 different motion events (Özyürek et al., 2005).

These findings are generally in line with the view that in utterance generation there is dynamic and on-line interaction between linguistic, gestural, and spatial representations of events (e.g., Kita & Özyürek, 2003; McNeill and Duncan, 2000). More specifically, they are compatible with the interface hypothesis of speech and gesture production proposed by Kita and Özyürek (2003), which claims that the way information is presented in gesture is influenced partially by the way it is linguistically packaged in a unit of speech production at the moment of speaking and partially by the spatial and motoric representation of the events. With regard to the language effect, the interface hypothesis proposes that representation in gesture is shaped by how much language can be packaged in one production unit. Given the assumption that a clause approximates one unit of speech production in adults (Bock & Cutting, 1992; Garrett, 1982; Levelt, 1989), the interface hypothesis predicts that English speakers predominantly use one gesture that conflates both manner and path, as both can be expressed in one unit of speech production, i.e., in one clause. In

contrast, as Turkish and Japanese speakers express each component in separate clauses in speech, they frequently use separate gestures for each component. This is because they have conceptualized the same information over more than one unit of production (i.e., in multiple clauses) during speaking.

Further evidence for the claim that what can be expressed in one clause influences gestural representation has been provided in another experimental study conducted only in English (Kita et al., 2007). In this study adult speakers used either a typologically congruent one-clause construction (e.g., “he rolls down the hill”) or a typologically incongruent two-clause construction (e.g., “he descends the hill while rolling”) to talk about simultaneously occurring manner and path. The congruent and incongruent uses were elicited by asking subjects to describe different event types (i.e., manner causing or not causing change of location of a figure) that elicited the two types of clausal packaging. The results showed that both the clause type and the event type had independent effects on shaping gestures. Consistent with the linguistic effect previously discussed for other studies, speakers in this study were more likely to use conflated gestures when they expressed manner and path in one clause (e.g., hand circling in the air while tracing a downward trajectory). However, they performed separated gestures like those of Turkish speakers when they used two clauses (e.g., just circling in the air, or just tracing a downward trajectory). Thus, when speakers’ online choices in speech go beyond the typologically congruent patterns, gestures also align with that pattern of information packaging, showing that the interactions between speech and gesture are online and dynamic during speaking.

Thus far very little is known about how children learning typologically different languages tune their gestures to the way they package both components at the clause level at the moment of speaking. However, previous research on how the relationship between speech and gesture develops in general can provide some relevant insight.

### Development of the Relations Between Speech and Gesture

Speech and gesture can develop in relation to each other in at least two possible ways. In one scenario, the two systems initially develop as relatively independent systems and become integrated later. In a second scenario, the systems emerge as one integrated system from early on.

Support for the first view comes from studies showing that the relations between speech and gesture in children do not always pattern in adult-like ways but rather change over time. For example, within the period between 9 months and around 2 years of age, children’s gestures precede and complement early spoken language at the one- and two-word stages (Acredolo & Goodwyn, 1988; Bates, 1976; Butcher & Goldin-Meadow, 2000; Capirci et al., 1996; Iverson, Capirci, & Caselli, 1994; Özcalışkan & Goldin-Meadow, 2005). Studies conducted with older children, for example between 5 and 8 years of age, have also reported both semantic and temporal asynchrony between the two systems (e.g., Alibali & Goldin-Meadow, 1993; Goldin-Meadow, 2003; Pine, Lufkin, Kirk, & Messer, 2007). One series of studies shows that children on the cusp of solving Piagetian conservation problems are more likely to express complementary information in their speech and

gesture (i.e., mismatches; Alibali & Goldin-Meadow, 1993). Pine et al. (2007) have further shown that in such cases of transition, gestures might even temporally precede the relevant speech segment in non-adult-like ways. Such gestures have been interpreted as indexing transitional periods in cognitive development (Alibali & Goldin-Meadow, 1993) or as revealing analogue representations of concepts that are not yet verbalized in speech (Pine et al., 2007). Overall, the underlying assumption in the above studies has been that speech and gesture might start as relatively independent systems that then become integrated later.

However, a few studies that have looked at the development of the relations between speech and gesture cross-linguistically have argued for another view, namely that the two systems are integrated early on, around 3 years of age. Nicoladis and colleagues have examined the development of gestures in a longitudinal study with French–English bilingual children between 2 and 3.5 years of age (Nicoladis et al., 1999), and in a cross-sectional study with older children whose average age was 4 years 3 months (Nicoladis, 2002). They found that each child produced more gestures in the language in which they were more proficient, where proficiency was measured by mean length of utterance in each language. Mayberry and Nicoladis (2000) have argued that, even at early ages, the development of gesture is linked to the development of language in general and to the specific languages children learn.

McNeill (2005) has also suggested that the two systems interact and that learning language in general shapes children’s iconic gestures as early as 3 years of age. McNeill investigated how Spanish-, Mandarin Chinese-, and English-speaking children age 3 to 11, as well as adults, represent simultaneously occurring manner and path in gesture in descriptions of the same motion event analyzed in Kita and Özyürek (2003). He found that all children in all age groups used fewer conflated gestures and more manner only and path only gestures than their adult counterparts, who all predominantly used conflated gestures.<sup>1</sup> McNeill (2005) proposed that this dominant and possibly universal pattern found in gestures of all children could be due to the effect of learning language in general on gestural representations. For example, learning that event components can be segmented and expressed by different words (e.g., one word for manner, one word for path, as in *roll up*) shapes how events are represented in gestures as well (i.e., one gesture for manner and one gesture for path).

Note that both Nicoladis et al. (1999) and McNeill (2005) share the same view that speech and gesture are integrated in general around 3 years of age, even though they have contradictory findings with respect to whether the language-specificity of gestures is visible in early stages of development. Their view contrasts with other claims that the two systems start out as relatively independent of each other and integrate later (e.g., Pine et al., 2007). None of these studies, however, has examined the development of ges-

<sup>1</sup> The fact that Spanish (verb-framed language) speakers used mainly conflated gestures seems to contradict the findings presented in Kita and Özyürek (2003) and what would be predicted by the interface hypothesis. However, McNeill (2005) reports no descriptive or quantitative analysis of the alignments between gesture and accompanying speech in terms of clause types. It is possible that Spanish adult speakers in McNeill’s data predominantly used one-clause (typologically incongruent) rather than two-clause constructions (typologically congruent), which might have influenced their gesture types as in Kita et al. (2007).

tural representations in relation to the type of language-specific encoding of information in the accompanying speech. Yet they would make different predictions with regard to whether linguistic specificity in these two modalities develops in parallel or separate ways, as is outlined below.

### Present Study

Here we investigate when and how cross-linguistic variation in speech and gesture depicting motion events develops in children ages 3, 5, and 9 learning two typologically different languages, namely Turkish and English. We asked children and adults to produce short narratives as they talked about 10 animated short movies (used in Allen et al., 2007; Kita et al., 2007; and Özyürek et al., 2005). The movies contain motion events in which simultaneously occurring manner and path are both salient. In analyzing the narratives of the participants, we focus on how simultaneously occurring manner and path components are packaged at the clausal level (one or multiple clauses) and how language-specificity of gestures accompanying these expressions develops. We tested two main hypotheses: (a) that children's gestures reflect adult-like differences from early on, and (b) that the language-specific differences in gestures develop later and are preceded by universal patterns in younger children. Going beyond previous research (e.g., McNeill, 2005), we conducted our investigation in such a way as to take into account three pieces of information: the type of clausal packaging of manner and path components used in speech, the type of representations in the accompanying gesture at the moment of speaking, and the relationship between clausal packaging and gestural representation.

### Predictions

Because we used the same data for English- and Turkish-speaking adults and 3-year-olds reported in Allen et al. (2007), Kita et al. (2007), and Özyürek et al. (2005), we already knew some of the speech and gesture patterns. English adult speakers use both the typologically congruent and incongruent clausal packaging patterns, the former being used more frequently than the latter, whereas Turkish adults use mainly typologically congruent patterns. Three-year-old children's speech is largely language specific, even though some seemingly universal patterns are also present at that point (e.g., a few Turkish-speaking children use typologically incongruent one-clause constructions like English-speaking children). For English-speaking 5- and 9-year-olds we did not expect further changes, and for Turkish speakers we expected children to reduce their use of one-clause constructions by either 5 or 9 years of age.

With regard to adult gesture patterns, Özyürek et al. (2005) have shown differences between speakers of the two languages when their gestures accompany typologically congruent clausal packaging in each language. Furthermore, Kita et al. (2007) have shown that English speakers prefer separated gestures when they use two clauses (typologically incongruent) and conflated gestures when they use one-clause constructions (typologically congruent) to package both manner and path.

With regard to patterns of gesture development, we can make different predictions with regard to our two main hypotheses based on previous research. One hypothesis is that children's gestures

reflect adult-like patterns in both languages from the earliest age. This would be predicted by Nicoladis et al. (1999) and also by the interface hypothesis (Kita & Özyürek, 2003). If children encode linguistic distinctions at the clause level in an adult-like way, then their gestures should also show adult-like cross-linguistic differences. The interface hypothesis assumes that what can be processed in one unit of production will shape gestures. Previous research has provided evidence that this unit is a clause for adults. If children's speech patterns show adult-like patterns and if they also have one clause as a processing unit, then their gestures would be expected to show adult-like differences.

The alternative hypothesis is that children's gestures show more similarity to those of other children across languages than to those of their adult counterparts within languages. That is, speech might become language specific earlier than gesture, and thus gesture would exhibit adult-like differences only later in development. In this scenario, one potential pattern is that both Turkish- and English-speaking children's gestures might represent simultaneously occurring manner and path in an analogue fashion, that is, with manner and path predominantly expressed simultaneously in a single conflated gesture no matter what type of clausal packaging is preferred. This would support the view that children's gestures show analogue representations of events independent of the linguistic encoding in speech (Pine et al., 2007). Another way children's gestures would look more similar to each other than to those of their adult counterparts is that both Turkish- and English-speaking children's gestures might represent manner and path separately, as found in McNeill (2005), possibly due to learning that language segments event components into separate units, such as words.

### Method

#### *Participants*

Participants in the study were 80 native speakers of Turkish and 80 native speakers of English. Twenty participants in the adult groups were university students in either Istanbul (Turkish) or Boston (English). The remaining 60 participants in each group were children, with 20 per group at each of three ages (3, 5, and 9). The child groups had similar mean ages and age ranges as well as gender distribution across languages (see Table 1). Child and adult participants in both countries came from middle to high socioeconomic status groups.

#### *Materials and Procedure*

Data were collected by elicitation, using a main set of 10 video clips depicting motion events involving simultaneous manner and path and two practice clips that resembled the main clips (Özyürek, Kita, & Allen, 2001). In the main set, 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* involved an object moving vertically up and down (always moving along a flat or inclined surface); *roll* involved an object turning on its horizontal axis (always moving along an inclined surface); *rotate* and *tumble* both involved an object turning on its horizontal axis



Table 1  
*Distribution of Speakers in Each Group*

Language	Age group	<i>n</i>	<i>M</i>	Range	Gender
Turkish	3-year-olds	20	3 years 8 months	3 years 6 months to 4 years	10 boys, 10 girls
	5-year-olds	20	5 years 7 months	5 years 6 months to 5 years 11 months	10 boys, 10 girls
	9-year-olds	20	9 years 4 months	8 years 9 months to 10 years 1 month	11 boys, 9 girls
	Adults	20	22.5 years	20–25 years	10 men, 10 women
English	3-year-olds	20	3 years 8 months	3 years 3 months to 4 years 3 months	8 boys, 12 girls
	5-year-olds	20	5 years 6 months	5 years 3 months to 6 years 1 month	4 boys, 16 girls
	9-year-olds	20	9 years 4 months	8 years 10 months to 10 years	7 boys, 13 girls
	Adults	20	27.6 years	18–40 years	10 men, 10 women

(always moving vertically through the air); and *spin* involved an object turning on its vertical axis (always moving along an inclined surface). Each video clip was between 6 and 15 s in duration. All clips involved a round red smiling character and a triangular-shaped green frowning character, moving in a simple landscape. Participants chose their own names for the characters; we refer to them here as “Tomato Man” and “Green Man.”

All clips had three salient components: an entry event, a target motion event, and a closing event. As an example, the roll + descend clip goes as follows. The initial landscape on the screen is a large hill with a downward slope, at the end of which is a tree. Tomato Man and Green Man enter the scene from the right. Green Man hits Tomato Man (entry event), then Tomato Man rolls down the hill (target motion event), and finally hits the tree at the end of slope (closing event). Figure 1 gives a sequence of the roll + descend clip with the target event shown in the middle.

Practice events were quite similar in structure to the experimental events. In one of the practice events, Tomato Man slides up a hill after being pushed by Green Man. In the other, Green Man spins around a tree and then Tomato Man jumps twice in place.

Participants were tested individually in a quiet space at their university (adults) or preschool, school, or after-school program (children). The procedure had two parts. During the warm-up phase, the experimenter showed participants two practice clips and asked the participant to recount what happened in the clip to a listener (one of our research assistants) who purportedly had not seen it. Following the practice clips, the experimenter showed the experimental clips. All practice and experimental clips were presented on a laptop computer. After each clip a black screen appeared, reducing the likelihood of points to the screen. If the participants did not mention the target event in their narration, either the experimenter or the listener encouraged them to do so

with a question grounded in either the entry event (e.g., “What happened after Green Man bumped into Tomato Man?”) or the closing event (e.g., “What happened before Tomato Man fell off the cliff?”). Crucially, the experimenter did not provide any biased information about either the manner or the path. Half of the subjects saw the clips in one order, and the other half saw them in the reverse order. All interactions were videotaped for later coding and analysis.

### *Speech Coding*

All speech describing the target motion events was transcribed by native speakers of the relevant language into MediaTagger, a video-based computer program (Brugman & Kita, 1995). As is evident from the English examples below, many participants used more than one utterance to describe a given target event. We refer to the full set of utterances used by one participant to describe a particular target motion event as a “target-event description.” Each of the three examples below constitutes a target-event description. The attribution for each example indicates the subject group (e.g., EA = English-speaking adult, T3 = Turkish-speaking 3-year-old), the subject number (e.g., EA-18 = Subject 18 in the EA group), and the name of the clip that was the stimulus for the sentence (e.g., spin + ascend).

1. “Tomato rolled down the hill.” (E3-38, roll + descend)
2. “The Green Guy goes up the cliff. He’s, like, spinning around while he’s going up.” (EA-18, spin + ascend)
3. “The Red Guy twirled down.” (E3-19, rotate + descend)

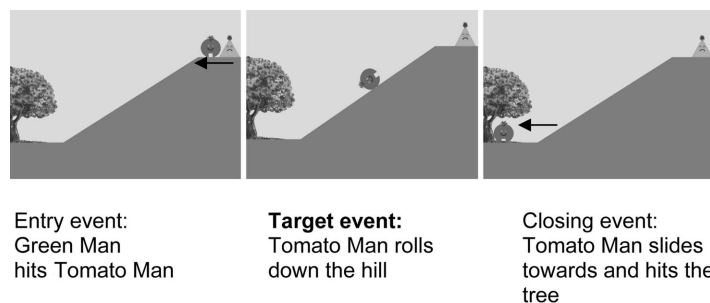


Figure 1. Selected stills from the roll + descend motion event

Several types of utterances were excluded from analysis: those that were not fully intelligible, those that were interrupted before completion, those that resulted from experimental error, and those that were about the target event but contained no reference to the specific manner or path in the clip being described.

Each target-event description consisted of one or more utterances. Depending on the syntactic packaging of manner and path in these utterances described below, each target-event description was then coded and classified into one of three categories: (a) including only *one-clause expressions* of manner and path, (b) including only *multiclauses expressions* of manner and path, or (c) including both types of expressions.

In all target-event descriptions, manner refers to the secondary movement (rotation along different axes, or jumping) of the figure that co-occurs with the translocational movement in the target events. Path refers to the directionality or trajectory specifications for the translational movement.

*One-clause expressions.* A target-event description was coded as including one-clause expressions if both manner and path were syntactically expressed in one clause, that is, a unit involving one verb and one closely associated nonverbal phrase. English one-clause expressions included manner verbs followed by directional path particles or prepositional phrases, as in Example 1 above. Target-event descriptions with one-clause expressions also occurred in Turkish, although they were rare. A typical example of this includes a manner verb with a postpositional directional path phrase, but crucially no path verb, as in Example 4.

4. *Domates adam aşağı yuvarlan-ıyor tepe-den.*  
tomato man downness roll-Present hill-Ablative

"Tomato Man rolls down the hill."

(TA-14, roll + descend)

*Multiclauses expressions.* In target-event descriptions that were coded as including multiclauses expressions, manner and path were distributed over separate clauses as path-only or manner-only clauses. These two clauses were conjoined in two different ways in the target-event descriptions as described next. In the first type, one motion element (either manner or path) was expressed in the main clause, whereas the other was expressed in a subordinate clause. In English, the subordinated form can be either a fully tensed verb (Example 5a) or a progressive participle functioning as an adverbial (Example 5b).

- 5a. "He spins in circles while he's going down."

(EA-18, spin + descend)

- 5b. "Triangle Man ascends the hill twirling."

(EA-20, spin + ascend)

In similar constructions in Turkish, the manner verb is subordinated to the main path verb with the use of a connective, mostly *-arak*, as in Example 6.

6. *Domates adam yuvarlan-arak yokuş-u in-di.*  
tomato man roll-Connective hill-Accusative descend-Past

"Tomato man descended the hill while rolling."

(TA-02, roll + descend)

In the second type of multiclauses expression, manner and path are each expressed in independent main clauses. These are sometimes conjoined by discourse markers such as *and*, *but*, and *or* in English, as in Example 3 above, and *ve* [and] and *sonra* [then] in Turkish, as in Example 7.

7. *Sonra yuvarlandı. Sonra aşağı indi.*  
then roll-Past then downness descend-Past

"Then (it) rolled. Then (it) descended down."

(T3-01, roll + descend)

The path-only and manner-only clauses were defined as follows in the two languages. In path-only clauses, there is only a path element (i.e., no manner). In English, clauses coded as path-only include the light path verb *go* followed by directional path particles or prepositional phrases (Example 8a), or other path verbs optionally followed by directional path particles or prepositional phrases (Example 8b).

- 8a. "He goes up a hill." (EA-27, roll + ascend)

- 8b. "It fell." (E3-11, spin + ascend)

In Turkish, clauses coded as path only include light path verbs (*come* and *go*), as in Example 9a, and other path verbs as in Example 9b, both with optional postpositional phrases that include spatial nouns specifying the source or the goal of the path.

- 9a. *Aşağı -ya gel-iyor.*  
downness-Dative come-Present

"(He/she/it) comes down." (TA-12, roll + ascend)

- 9b. *Sonra yukarı çık-tı.*  
then upness ascend-Past

"Then (he/she/it) ascended (to) the top."

(T3-02, jump + ascend)

In manner-only clauses there is only a manner element in the clause (i.e., no path) both in English and Turkish. An English example is in Example 10.

10. "And the Red Guy twirled." (E3-06, roll + ascend)

*Both types of expressions.* Finally, a few target-event descriptions contained both one-clause and multiclauses expressions of manner and path, as in Example 11.

11. "Grinch goes around the tree, but he bounces around the tree." (EA-01, jump + go.around)

The first sentence expresses only path in the main clause. The second sentence expresses manner and path together in one clause. This type of target-event description overall was then coded including both a one-clause and a multiclauses expression.

*Reliability.* To establish reliability of the coding, a second coder independently processed 20 percent of the data. The second coder judged the category type of the expressions (i.e., one-clause, multiclauses) in the event descriptions that had been transcribed and segmented into sentences by the original coder. The agreement between coders for this judgment was 97% for English and 93% for Turkish. In cases of discrepancy, the judgment of the original coder was adopted.

### Gesture Coding

We transcribed gestures that depicted manner and/or path and were concurrent with one-clause and multiclausal expressions in the target-event descriptions that contained both of these elements. In deciding whether a gesture depicted manner and/or path, only the stroke (the meaningful phase) of the gesture was taken into consideration. The stroke (Kendon, 1980; McNeill, 1992) was isolated using frame-by-frame video analysis, according to the procedure detailed in Kita, van Gijn, and van der Hulst (1998).

Target-event gestures were classified into four types: *manner only*, *path only*, *conflated*, and *unclear*. Manner-only gestures encoded manner of motion (e.g., a repetitive up and down movement of the hand to represent jumping) without encoding path. Path-only gestures expressed change of location without encoding manner. Conflated gestures expressed both manner and path at the same time throughout the entire stroke (e.g., repetitive up and down movements superimposed on diagonal downward change of location of the hand, representing jumping down the slope). Finally, some gestures were coded as unclear because they were either hard to segment or hard to categorize for any of the categories above. For purposes of clarity, we excluded gestures that were unclear from the analysis. The gestures coded as unclear comprised 17% (3-year-olds), 15% (5-year-olds), 14% (9-year-olds), and 13% (adults) of all gestures.

We also excluded the few gestures that used mainly the body to represent change of location or manner (e.g., using head, shoulders, and torso but, crucially, not hands). These gestures comprised 4% (3-year-olds), 2% (5-year-olds), 2% (9-year-olds), and less than 1% (adults) of all gestures. Such gestures have been termed *character viewpoint* gestures in the literature (McNeill, 1992), because the speaker tries to enact the actions of the agent by mapping them directly onto his or her own body. The use of body in such gestures is biased toward only one representation, i.e., manner only or path only, as it is hard to represent change of location, for example, if a child is jumping in his seat to express manner.

Finally, each gesture included in the analysis was further coded for whether it overlapped with a one-clause or multiclausal expression in speech, because our investigation mainly focuses on which types of gestures (i.e., manner only, path only or conflated) are used in the context of certain types of clausal packaging (i.e., one-clause or multiclausal). In this coding each gesture was categorized as occurring in the context of either one-clause or multiclausal expression, as shown in Examples 12 and 13, below. Sometimes different types of gestures were used within the context of one type of clausal packaging. For example, both manner-only and path-only gestures were coded as occurring in the context of multiclausal expression in Example 12, but in the context of one-clause expression in Example 13. (The underlines in Examples 12 and 13 indicate where the strokes of the gestures are in relation to speech.)

12. He was spinning around while he was  
(manner only)

coming down the hill.  
(path only)

(multiclausal expression)  
(E9-13, rotate + descend)

13. Broccoli twisted all the way up.  
(manner only) (path only)

(one-clause expression)  
(E9-10, spin + descend)

In rare cases, if a participant used more than one type of clausal packaging (both one-clause and multiclausal) and multiple gestures in one target-event description, each gesture was categorized with regard to the type of clause it overlapped with. For example, if a participant used one path-only clause followed by a one-clause expression, as in Example 11 above, and used two gestures, the gesture overlapping with the path-only expression in speech would be categorized as overlapping with a multiclausal expression and the other gesture as overlapping with a one-clause expression.

Finally, in a few cases one gesture straddled two types of clauses. These gestures were excluded. We also excluded gestures that did not overlap with any clause that expressed the target event. Such excluded gestures comprised 1% (3-year-olds), less than 1% (5-year-olds), 2% (9-year-olds), and 3% (adults) of all gestures.

*Reliability.* In order to establish reliability of the gesture type classification, a second coder judged the gesture type (i.e., manner only, path only, conflated, unclear) for 20% of the target-event gesture strokes that had been identified and segmented by the original coder. The agreement between coders was 87% for the English data and 94% for the Turkish data. In cases of discrepancy, the judgment of the original coder was adopted.

## Results

### Speech

We first analyzed the speech of all the participants to determine whether the Turkish-English and adult-child patterns followed those predicted by the typological differences between the languages. Note that the data for the English- and Turkish-speaking adults and 3-year-olds come from the same data set reported in Allen et al. (2007). However, the criteria for using the data of individual participants, as well as the units of analysis used, are somewhat different across the two studies, with resulting slight differences in the findings reported. The study presented here also includes data from Turkish- and English-speaking 5- and 9-year-olds, which were not included in the Allen et al. (2007) article.

In the first speech analysis, we investigated to what extent each language and age group expressed both manner and path in their target-event descriptions. For each age and language group we calculated the proportion of events where both manner and path were mentioned, as shown in Table 2.

A  $4 \times 2$  analysis of variance (ANOVA) conducted on these proportions with age (3, 5, 9, adult) and language (English, Turkish) as factors revealed a main effect of age,  $F(3, 152) = 41.175$ ,  $p < .001$ ,  $\eta_p^2 = .44$ , but not language,  $F(1, 152) = 0.092$ ,  $p = .7$ ,  $\eta_p^2 = .00$ , or interaction between the two,  $F(3, 152) = 1.45$ ,  $p = .2$ ,  $\eta_p^2 = .02$ . Tukey's honestly significant difference (HSD) post hoc tests showed that all age groups differed significantly ( $p < .05$ ) from one another within each language. That is, younger groups expressed both manner and path together in their target-event descriptions less frequently than did older groups in both languages.

In the next analysis, out of all the event descriptions that included both manner and path (as in Table 2), we calculated the proportion of those that included one-clause versus multiclausal expressions in each age and language group and compared them across languages (see Figure 2). A  $2 \times 4$  ANOVA on the proportions of event descriptions that included one-clause expressions of manner and path showed a main effect of language,  $F(1, 151) = 646.324$ ,  $p < .001$ ,  $\eta_p^2 = .81$ , but not of age,  $F(3, 151) = 2.242$ ,  $p = .08$ ,  $\eta_p^2 = .04$ , and no interaction between the two,  $F(3, 151) = 1.88$ ,  $p = .1$ ,  $\eta_p^2 = .03$ . That is, from age 3 onward, English speakers used more one-clause expressions than did Turkish speakers (see Figure 2A). Note that Turkish speakers also used a few one-clause expressions, which are not typologically congruent. There was also a trend for 3- and 5-year-old Turkish children to use more of these one-clause expressions than their adult counterparts, but the difference did not reach significance.

In the next step, we conducted an ANOVA on the proportions of the same event descriptions that included multiclausal expressions. The results revealed again a main effect of language,  $F(1, 151) = 247.517$ ,  $p < .001$ ,  $\eta_p^2 = .62$ , but not of age,  $F(3, 151) = 0.770$ ,  $p = .5$ ,  $\eta_p^2 = .01$ , and there was no interaction between the two,  $F(3, 151) = 0.536$ ,  $p = .6$ ,  $\eta_p^2 = .01$ . These results again show that English- and Turkish-speaking children's expressions of manner and path differed from one another from 3 years of age onward, paralleling the adult differences, with English speakers producing fewer multiclausal expressions than Turkish speakers (see Figure 2B).

The results reported here essentially replicate those in Allen et al. (2007) in terms of the main language-specific differences found between English and Turkish speakers in adults and 3-year-olds. However, in the Allen et al. study, it was further found that Turkish 3-year-olds used significantly more one-clause expressions than their adult counterparts. In the current analysis presented in this article, a parallel but nonsignificant trend was found for Turkish 3- and 5-year-olds (see Figure 2A). This difference is likely due to the fact that in Allen et al., each subject had to contribute at least 3 events (out of 10) that included both manner and path to be included in the statistical comparisons. In the present study, all subjects' event descriptions that included both manner and path were included in the statistical analysis in order to be able to include as many verbal expressions as possible that could be used with gestures. This could have increased variability in the data and prevented the difference from reaching significance.

### Gesture

The speech analysis showed that English- and Turkish-speaking children as young as age 3 were already largely tuned to the

language-specific clausal packaging of manner and path expressions. Furthermore, in each language, speakers also used typologically incongruent expressions but to a lesser extent. In the subsequent analyses, we focused on gesture types produced in target-event descriptions that linguistically expressed both manner and path.

We conducted two separate analyses on the proportions of conflated versus separated gestures (manner only and/or path only): first, without considering the type of clausal packaging they accompanied, and second, taking into account the particular clause type they overlapped with. The purpose for conducting two analyses was to be able to see distinctly to what extent differences in gestures are robust across the groups versus specific to the clause types chosen by child and adult speakers at the moment of speaking.

*Development of conflated versus separated (manner only and/or path only) gestures in the context of all clause types.* In the first analysis, we investigated the use of conflated versus separated gestures in all types of clauses across languages and ages. For each subject the proportion of conflated gestures over all gesture types (conflated and separated [manner only and/or path only]) was calculated. Because not all participants used many gestures (especially at early ages) and because there was substantial variability in the number of gestures related to motion event descriptions in the child groups, only data from participants who contributed a total of six or more gestures were included in this analysis. In this way, we facilitated statistical analysis by maintaining enough variation in the proportions and avoiding excessively small denominators. As a result, the number of speakers in each group included in the analysis is as follows: TA = 20, T9 = 19, T5 = 16, T3 = 9, EA = 20, E9 = 18, E5 = 20, E3 = 16.

We compared this proportion of conflated gestures using a  $4 \times 2$  ANOVA with age (3, 5, 9, adult) and language (English, Turkish) as factors. There were main effects of both language,  $F(1, 130) = 26.263$ ,  $p < .01$ ,  $\eta_p^2 = .168$ , and age,  $F(3, 130) = 5.037$ ,  $p < .01$ ,  $\eta_p^2 = .104$ , and also a marginal interaction between the two,  $F(3, 130) = 2.289$ ,  $p = .08$ ,  $\eta_p^2 = .050$ . Figure 3 below shows that proportions of conflated gestures over all gesture types were higher in English than in Turkish and also higher in older age groups than in younger ones; the latter was especially true for English speakers, as indicated by the marginal interaction.

*Development of conflated versus separated (manner only and/or path only) gestures in the context of different clause types.* In the next analysis we investigated how conflated versus separated gestures develop depending on the different clause types they overlap with. We calculated the mean proportion of conflated gestures used by each participant, out of the total number of analyzable gestures for that given participant (i.e., conflated and separated [path only and/or manner only]), that overlapped with one-clause and multiclausal expressions in English and multiclausal expressions in Turkish (see Figure 4). We included only data from participants who used six or more gestures overlapping with these expressions for the statistical reasons explained above.

Note that gestures in the context of one-clause expressions (see Figure 2A for these proportions) were not included in the statistical comparisons, as the number of Turkish speakers who used at least one gesture with a one-clause expression was very small: T3 = 3, T5 = 9, T9 = 3, TA = 2. Furthermore, only a few of these

Table 2  
Events Where Both Manner and Path Are Expressed in Speech

Age group	Turkish		English	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Adults	0.90	0.12	0.84	0.12
9-year-olds	0.74	0.16	0.78	0.19
5-year-olds	0.62	0.26	0.58	0.14
3-year-olds	0.40	0.18	0.49	0.21



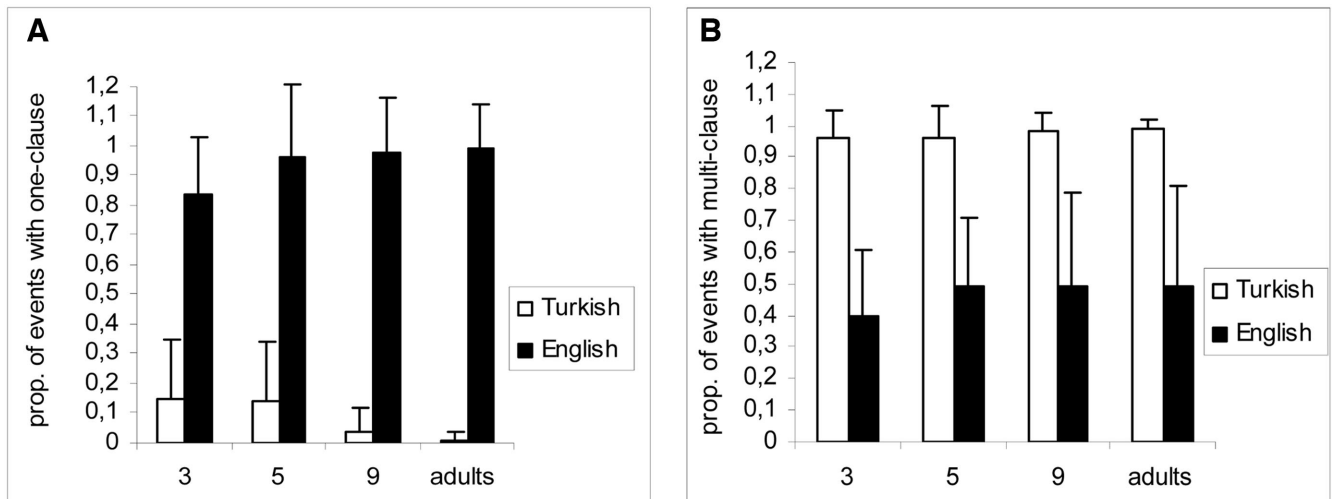


Figure 2. Proportion of event descriptions in which (A) at least one one-clause expression and (B) at least one multiclausal expression were used by Turkish and English speakers across ages. Prop. = proportion. Error bars represent standard deviations.

participants could contribute with six or more gestures to the statistical analyses.

The proportions in Figure 4 show that the distribution of conflated versus separated gestures differs according to the clause types they accompany. To test this, first we conducted a statistical comparison on the gestures that overlapped with typologically congruent expressions in each language, that is, one-clause expressions in English and multiclausal expressions in Turkish. According to previous research we would expect differences in adults' gestures to be most evident in the context of typologically congruent expressions in each language (Kita & Özyürek, 2003; Özyürek et al., 2005). The number of speakers in each group who contributed six or more gestures to the analysis is as follows:

TA = 20, T9 = 19, T5 = 15, T3 = 8, EA = 18, E9 = 14, E5 = 12, E3 = 8.

A  $4 \times 2$  ANOVA was conducted, with age (3, 5, 9, adult) and language (English, Turkish) as factors, on the mean proportions of analyzable gestures that overlapped with one-clause expressions in English and multiclausal expressions in Turkish (see Figure 4). The results revealed main effects of age,  $F(3, 106) = 5.31, p < .05, \eta_p^2 = .281$ , and language,  $F(1, 106) = 41.48, p < .001, \eta_p^2 = .131$ , as well as an interaction between the two,  $F(3, 106) = 3.40, p < .05, \eta_p^2 = .08$ . Tukey's HSD post hoc comparisons at each age revealed that for both the adult and 9-year-old groups, the English speakers produced a higher proportion of conflated gestures than did the Turkish speakers ( $p < .01$ ). However, for both

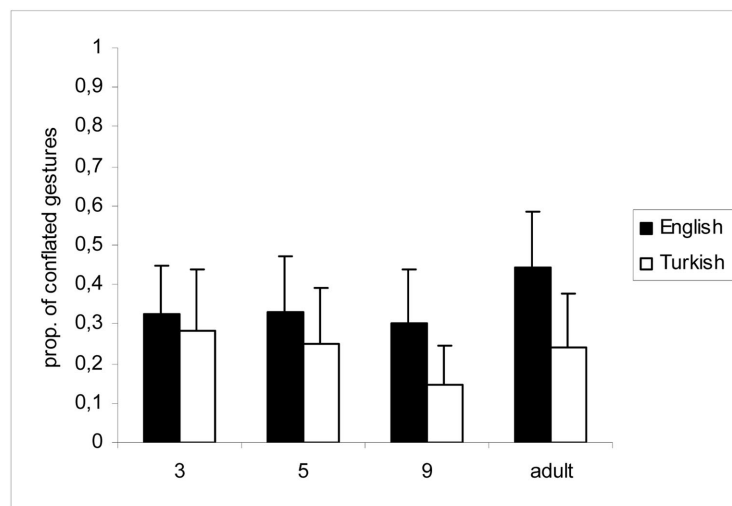


Figure 3. Proportion of conflated gestures in Turkish and English speakers across ages collapsing all clause types. Prop. = proportion. Error bars represent standard deviations.

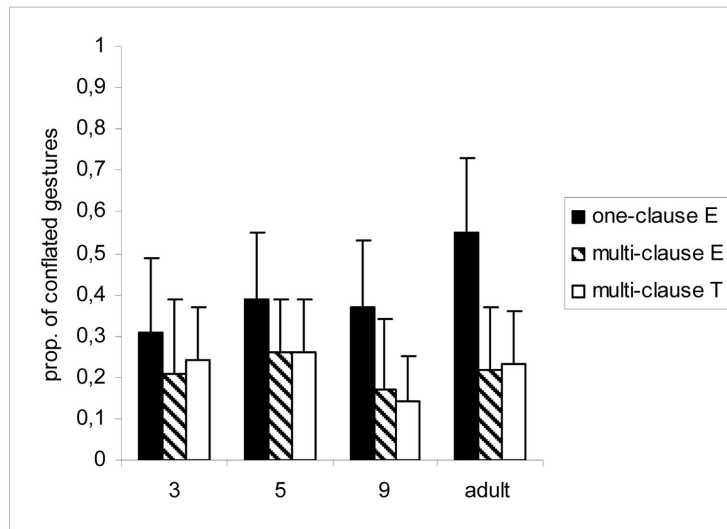


Figure 4. Proportion of conflated gestures used to accompany descriptions with multiclausal expressions in Turkish and one-clause and multiclausal expressions in English, by age. Prop. = proportion; E = English; T = Turkish. Error bars represent standard deviations.

the 5- and 3-year-old groups, there was no statistically significant difference between the English speakers and the Turkish speakers. Age comparisons among English speakers revealed that the proportion of conflated gestures for English-speaking adults was higher than that for English-speaking 3-year-olds ( $p < .01$ ), but no other differences between the other age groups were significant. For Turkish speakers, none of the age groups differed significantly from any of the others. Thus, the interaction was because the proportion of conflated gestures increased with age for English speakers but did not change with age for Turkish speakers.

As a second step, we conducted the same analysis but this time focused on the gestures that overlapped with the same type of clausal packaging in Turkish and English. We saw in the speech analysis (Figure 2B) that English speakers used a substantial number of multiclausal expressions, as did Turkish speakers, even though this is not the typologically congruent pattern in English. Therefore, in this analysis we compared the proportion of conflated gestures out of all analyzable gestures (i.e., conflated and separated [manner only and/or path only]) that overlapped with multiclausal expressions both in Turkish and English. As in the first analysis, only participants who used six or more gestures in this context were included in the analysis: TA = 20, T9 = 19, T5 = 15, T3 = 8, EA = 13, E9 = 12, E5 = 6, E3 = 3.

A  $4 \times 2$  ANOVA with age (3, 5, 9, adult) and language (English, Turkish) as factors was conducted on the proportions of conflated gestures that co-occurred with multiclausal expressions in all groups. Unlike the previous analysis, these results revealed no main effect of age,  $F(3, 88) = 1.120$ ,  $p = .07$ ,  $\eta_p^2 = .07$ , or language,  $F(1, 88) = 2.26$ ,  $p = .9$ ,  $\eta_p^2 = .00$ , and also no interaction,  $F(3, 88) = 1.56$ ,  $p = .9$ ,  $\eta_p^2 = .00$ . That is, when speakers of the two languages used the same clausal packaging of information, their gesture patterns also looked the same. Furthermore, no development was observed across ages in any of the languages (see Figure 4).

Overall, the two analyses above show that English- and Turkish-speaking children's gestures start out in similar ways, that is, by

using separated gestures for manner and path. English-speaking children shift over time from using separated to conflated gestures, whereas for Turkish no development is observed. Thus, English speakers deviate from Turkish speakers at age 9 and in adulthood. More importantly, the developmental and cross-linguistic effects are not robust across the groups. The differential development of gesture types in English speakers is restricted to gestures that overlap with one-clause expressions, not with multiclausal expressions.

*Children's preference for separated gestures across languages: Memory limitations or effect of learning language?* The early preference for children to represent manner and path in separated gestures found in the above analysis can be considered in line with what would be predicted by McNeill (2005). McNeill proposed that this early preference could be a universal effect of learning language; learning that there are separate words for different event components also induces gestures to be segmented. However, an alternative reason to expect an early bias to use separated gestures is potential working memory limitations during speech production. Working memory limitations of children have been suggested as possible explanations for limitations on cognitive processes such as language production (e.g., Case, Curland, & Goldberg, 1982; Chi, 1978). Thus, in the case of motion event expressions, production of both manner and path in speech may tax children's working memory to the extent that they can produce only one of these semantic elements in gesture while speaking.

In order to disentangle these two explanations, we differentiated two ways that separated gestures could be manifested. One possibility is for speakers to produce either a path-only or a manner-only gesture, omitting the other. The other possibility is for speakers to use both types in a segmented fashion (e.g., first manner only and then path only). For ease of referring back to these types, we label the first possibility *manner-only or path-only gestures* and the second *both manner-only and path-only gestures*.

The motivation for distinguishing these two potential manifestations of separated gestures is that they correspond with different

scenarios and predictions concerning children's working memory limitations versus the effect of learning language on gestures, as McNeill (2005) claims. The first scenario is as follows. Production of both manner and path in an utterance may tax children's working memory to the extent that they are able to coordinate production of only one of these semantic elements in gesture. If this were the case, young children would be more likely to produce manner-only or path-only gestures in the context of one-clause or multiclausal expressions than older children. That is, children would first focus on only one element in gesture, and this trend would decrease over time as their working memory increased and allowed them to produce both elements in gesture (i.e., either as conflated gestures or as both manner-only and path-only gestures). The second scenario is this: Children might initially be able to express both manner and path in gesture in the context of one-clause or multiclausal expressions in speech, but they might use separated gestures due to learning that event components are encoded by different linguistic units (McNeill, 2005). Under this scenario, children would begin with both manner-only and path-only gestures and would then gradually decrease use of these as the number of conflated gestures increases.

To test these possibilities, for each subject we calculated the proportions of one-clause expressions in English and multiclausal expressions in Turkish that were used with either (a) conflated gestures, (b) manner-only or path-only gestures, or (c) both manner-only and path-only gestures out of all one-clause and multiclausal expressions with analyzable gestures (Figure 5A and B). (Note that we did not have enough multiclausal expressions in English and one-clause expressions in Turkish that were accompanied by gestures to satisfy our statistical criteria described below.) Furthermore, within the multiclausal expressions in Turkish, we selected only those where manner and path expressions were linked within a sentential clause boundary (i.e., main and subordinated clauses) rather than distributed over separated sentences conjoined with a discourse boundary (e.g., "and then"), so that the speech unit selected in both English and Turkish would be similar,

i.e., a sentence. Finally, expressions with conflated gestures were also included in the analyses to see if we could replicate the cross-linguistic and developmental findings in the previous analysis with this different type of quantification. Note that the current analysis looks at the proportion of one-clause or multiclausal expressions that co-occurred with one of the above gesture types rather than looking at the proportion of gestures that were used within the context of a certain type of clause.

The analysis included only those subjects who contributed three or more one-clause expressions in English or multiclausal expressions in Turkish so that we could maximize the number of subjects included in each group while removing the variability due to excessively small denominators. The number of speakers in English and Turkish was as follows: EA = 20, E9 = 18, E5 = 16, E3 = 11, TA = 20, T9 = 20, T5 = 13, T3 = 6.

We conducted three different one-way ANOVAs, first in English and then in Turkish, with age as the only factor. The first three ANOVAs were conducted on the proportion of one-clause expressions in English that co-occurred with either conflated gestures, manner-only or path-only gestures, or both manner-only and path-only gestures.

The results showed that the proportion of one-clause expressions that co-occurred with conflated gestures increased with age,  $F(3, 61) = 3.95, p < .05, \eta_p^2 = .16$ , mirroring findings in the previous analysis. Tukey's HSD post hoc tests revealed significant differences between 3-year-olds and adults as well as between 9-year-olds and adults (all comparisons  $p < .05$ ; see Figure 5A). In contrast, the proportion of one-clause expressions that co-occurred with both manner-only and path-only gestures decreased significantly with age,  $F(3, 61) = 5.9, p < .001, \eta_p^2 = .22$ . The Tukey's HSD post hoc tests revealed significant differences between 3- and 9-year-olds as well as between 3-year-olds and adults (all comparisons  $p < .05$ ; see Figure 5A). However, the proportion of manner-only or path-only gestures did not change with age,  $F(3, 61) = 2.5, p = .07, \eta_p^2 = .11$ .

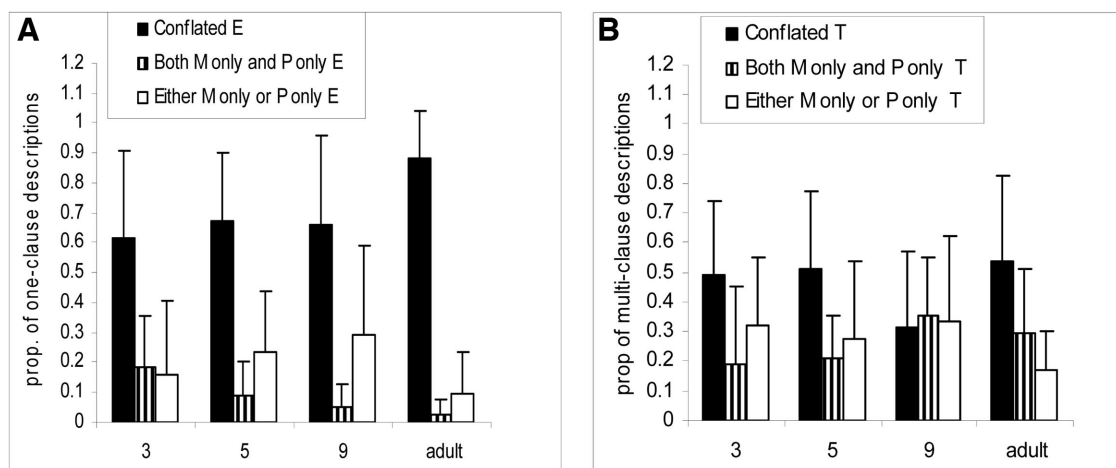


Figure 5. A: Proportion of one-clause expressions in English with conflated, both manner-only and path-only, and either manner-only or path-only gestures across ages. B: Proportion of multiclausal expressions in Turkish with conflated, both manner-only and path-only and either manner-only or path-only gestures across ages. Prop. = proportion; E = English; M = manner; P = path; T = Turkish. Error bars represent standard deviations.

The second set of ANOVAs was conducted in Turkish on the proportions of multiclausal expressions that co-occurred with either conflated gestures, manner-only or path-only gestures, or both manner-only and path-only gestures out of all such expressions with analyzable gestures. None of the comparisons revealed significant differences across ages in Turkish (all  $ps > .10$ ; see Figure 5B). The proportion of multiclausal expressions with either conflated or separated gestures did not change across the age groups.

The results of the analyses both in Turkish and English provide evidence against working-memory-based accounts of young children's preference for separated gestures. Contrary to expectations, one-clause expressions and multiclausal expressions that co-occurred with manner-only or path-only gestures stayed the same over time in both English and Turkish. Furthermore, in English one-clause expressions that co-occurred with both manner-only and path-only gestures decreased over time, suggesting that they were replaced with expressions that co-occurred with conflated gestures. This finding is not compatible with a working memory limitation account of the phenomena but rather is line with McNeill's (2005) idea that children's use of separated gestures reflects the breaking down of an event into its components, possibly due to learning that language segments events into units.

### Discussion

Previous research on how children learn language-specific ways of expressing spatial relations has mostly focused on speech and has shown that children are quite good at tuning their early semantic and syntactic patterns of encoding space to the patterns of their adult counterparts. However, whether and when children's cospeech gestures reflect language-specific ways of encoding spatial relations has not been studied thus far. In addressing this question, we tested two main hypotheses. One was that children would show language specificity in their gestures paralleling that in their speech from early on. The other hypothesis was that even though children's speech showed language specificity, their gestures would differ from the respective adult patterns and show more similarity to the gesture patterns of child speakers of other languages. We investigated the development of motion event expressions in speech and cospeech gestures of two typologically different languages: English and Turkish.

With regard to speech, we used the same data as in Allen et al. (2007). Even though here we used a different unit of analysis on the same data set, our results paralleled those of Allen et al. for 3-year-olds and adults. Turkish- and English-speaking children learned the language-specific ways of encoding manner and path (one-clause expressions in English and multiclausal expressions in Turkish) from 3 years of age onward. English-speaking children in all age groups also used typologically incongruent speech patterns as much as their adult counterparts did. Finally, Turkish-speaking 3- and 5-year-old children showed a slight trend to use typologically incongruent patterns more than their adult counterparts, but this difference did not reach significance.

On the other hand, we found that gestural representations take longer (i.e., around 9 years) to reflect adult-like differences across the languages. Our results show that English- and Turkish-speaking children's gestures start out in similar ways, that is, by using separated gestures for manner and path. English-speaking children over time shift from separated to conflated gestures and

deviate from Turkish speakers at ages 9 and adulthood, whereas for Turkish no development is observed. The differential development of gesture types in English speakers, however, is restricted to gestures that overlap with one-clause expressions but not with multiclausal expressions. Thus, the developmental and cross-linguistic effects are not robust across the groups but are tied to the type of clausal packaging they accompany.

In the next analysis we further tested whether children's preference for separated gestures could be explained by children's working memory limitations or rather supported McNeill's (2005) idea that children use separated gestures due to learning language. One could propose that production of both manner and path in speech might tax children's working memory to the extent that they would only be able to coordinate production of one of these semantic elements in gesture within an utterance. However, neither the English- nor the Turkish-speaking children used just one element in gesture (i.e., either manner only or path only) more frequently than their adult counterparts did. In fact, English-speaking children early on produced both elements in gesture but in a segmented fashion and only later shifted to using predominantly conflated gestures. This finding cannot be explained by working memory limitations but rather is more in line with what would be expected from McNeill's claim.

### *Implications of Findings for the Development of the Relations Between Speech and Gesture*

The finding that the gestures of children learning different languages start out in similar ways and do not reflect adult-like differences does not fit with what would be expected from earlier research on bilingual children showing that development of gestures is linked from an early age to the specific language being spoken (Nicoladis, 2002; Nicoladis et al., 1999). Similarly, our findings do not support the view that children's iconic gestures at early ages are generated purely from imagistic, mimetic, and analog representations of events, which are independent of speech (Pine et al., 2007).

Rather, our finding that both English- and Turkish-speaking children start out with separated gestures is in line with McNeill's (2005) previous findings. This pattern can then indeed reflect a universal early bias in children to segment event components into units as an influence of learning language, as claimed by McNeill.<sup>2</sup>

However, McNeill's (2005) account is suitable to explain only the early patterns we see in our data; it cannot explain the development of gestures linked to speech development. In particular, it does not explain why we find that English speakers, but not Turkish speakers, shift their gestures from separated to conflated and only in the context of one-clause, but not multiclausal, expressions. Here we offer one alternative hypothesis to explain our data, namely a processing explanation. We argue that this hypothesis is

<sup>2</sup> Children's tendency to segment components of events into linguistic units and introduce them into an emerging sign language has also been documented in Nicaragua. Senghas, Kita, and Özyürek (2004) investigated how simultaneous manner and path are expressed in event descriptions of the Sylvester and Tweety cartoon and found that new cohorts who entered the community as children used more signs that segmented manner and path into units than first cohort signers who entered the community as adults.



the most compatible one to explain our findings, even though it cannot be definitive given the limits of our study.

### *Development of Gestures Toward Language Specificity: A Processing Account*

In previous research, Kita and Özyürek (2003) and Kita et al. (2007) have claimed that adults' gestures are shaped by what can be packaged within one processing unit of language production, namely a clause. This model has been used to explain why English speakers use conflated gestures while producing one verbal clause in speech, but Turkish speakers use separated gestures while using two verbal clauses in speech to express manner and path. At first glance, our findings seem to contradict what we would expect from the interface hypothesis, because even though children's speech reflected typological distinctions, their gestures did not reflect the way each language packaged manner and path at the clause level.

Yet we propose that the interface hypothesis can still explain the results if we take into account possible differences between children and adults in terms of processing capacity, even though the surface structure of their speech is the same. Our claim is that the unit of processing shapes children's gestures, as it does adults' gestures, in accordance with the interface hypothesis. However, that unit may be smaller for children than for adults—perhaps a word or a phrase rather than a clause. If that were so, then children's gestures would represent only what can be processed within one phrase or word, thus resulting in separated gestures (e.g., one gesture for the manner verb and/or another for the prepositional path phrase in English). As the size of the processing unit increases, 9-year-old English-speaking children are able to process a full verbal clause at once and thus produce conflated gestures when they combine manner and path in one clause in speech. This proposal of a smaller processing unit for children than for adults can also explain why Turkish-speaking children's gestures appear adult-like from age 3, but English-speaking children's gestures do not. The manner and path representations of Turkish-speaking adults are still conceptualized in two units because they appear in separate clauses in speech. Thus, the difference in size between adult and child processing units does not affect this particular gesture pattern for Turkish speakers. This same processing account can also explain why we see a shift from separated to conflated gestures in English only in the context of one-clause expressions but not for multiclausal expressions.

The idea that children have a smaller processing capacity than adults during language production has already been suggested in the literature. English-speaking 2-year-old children are known to omit the subject noun phrase in places where adult speakers would not. The distribution of such non-adult-like omissions has been taken to support the idea that children have a smaller processing capacity than adults and can plan much less than adults at one time (L. Bloom, 1970; P. Bloom, 1990; Freudenthal, Pine, & Gobet, 2007; Valian, Hoeffner, & Aubry, 1996). For example, the subject is omitted more often in negative sentences than in affirmative sentences (L. Bloom, 1970). Another related finding is that the verbal phrase tends to be longer (in terms of number of morphemes) when the subject is omitted (P. Bloom, 1990; Freudenthal et al., 2007). That is, children may drop the subject to ease the burden of speech production processes when the utterance is long. This is compatible with the idea that children can process fewer

words than adults within the unit of speech production processes (see Hyams & Wexler, 1993, and references therein for alternative suggestions to explain children's omissions on the basis of children's non-adult-like grammatical settings). However, this previous research on children's processing limitations is conducted on children younger than age 4 and may not directly account for why processing limitations persist until 9 years of age at the clause level. Further research on children's processing capacity at older ages might shed light on the possibility that gestures index differences in children's language processing capacity at different ages.

Finally, we discuss one other factor, namely motoric limitations of children, which could be a counterexplanation to the interface hypothesis. One could claim that limitations in children's motoric ability prevent them from being able to perform conflated gestures at younger ages (i.e., in the case of English-speaking children). However, we find this explanation highly unlikely. If older children's and adults' shift from separated to conflated gestures were merely due to development of motoric ability, we would have seen a shift in English for all clause types, but we did not.

### *Conclusion*

Our findings overall show that language-specific encoding of motion develops earlier in speech than in gesture. It takes considerable time (i.e., until sometime after 9 years of age) for children's gestures to reflect adult-like and language-specific differences. Our analyses show that these developmental patterns cannot simply be explained by working memory or motoric limitations. The explanation most compatible with our data is that the patterns in gestures reflect differences in the way gestures are shaped by language over development (i.e., from a word or a phrase to a clause), as well as gradual shifts toward language-specific processing of event components during speaking. In languages where more than one event component needs to be expressed within one clause, as in English, it will take a longer time for children's gestures to take on an adult-like character. Our results are thus compatible with the view that there are interactions between language and gesture at an early age (Mayberry & Nicoladis, 2000; McNeill, 2005), but they go beyond previous research by suggesting that what changes over time is the unit of language production where interactions take place. Thus, gestures provide a window into gradual developmental shifts toward language-specific event representations generated for speaking, as well as changes in language production processes, which may not be evident by looking at speech only.

### *References*

- Acredolo, L., & Goodwyn, S. (1988). Symbolic gesturing in normal infants. *Child Development*, 59, 450–466.
- Alibali, M. W., & Goldin-Meadow, S. (1993). Gesture–speech mismatch and mechanisms of learning: What the hands reveal about a child's state of mind. *Cognitive Psychology*, 25, 468–523.
- Allen, S., Özyürek, A., Kita, S., Brown, A., Furman, R., Ishizuka, T., & Fujii, M. (2007). Language-specific and universal influences in children's packaging of manner and path: A comparison of English, Japanese, and Turkish. *Cognition*, 102, 16–48.
- Bates, E. (1976). *Language and context: The acquisition of pragmatics*. New York: Academic Press.

- Bernardis, P., & Gentilucci, M. (2006). Speech and gesture share the same communication system. *Neuropsychologia*, 44, 178–190.
- Bloom, L. (1970). *Language development: Form and function in emerging grammars*. Cambridge, MA: MIT Press.
- Bloom, P. (1990). Subjectless sentences in child language. *Linguistic Inquiry*, 21, 491–504.
- Bock, K., & Cutting, J. C. (1992). Regulating mental energy: Performance units in language production. *Journal of Memory and Language*, 31, 99–127.
- Bowerman, M. (1982). Starting to talk worse: Clues to language acquisition from children's later speech errors. In S. Strauss (Ed.), *U-shaped behavioral growth* (pp. 101–114). New York: Academic Press.
- Brugman, H., & Kita, S. (1995). Impact of digital video technology on transcription: A case of spontaneous gesture transcription. [Special issue]. *KODIKAS/Code: Ars Semeiotica*, 18, 95–112.
- Butcher, C., & Goldin-Meadow, S. (2000). Gesture and the transition from one- to two-word speech: When hand and mouth come together. In D. McNeill (Ed.), *Language and gesture* (pp. 235–257). New York: Cambridge University Press.
- Capirci, O., Iverson, J., Pizzuto, E., & Volterra, V. (1996). Gestures and words from the transition from one-word to two-word speech. *Journal of Child Language*, 23, 645–673.
- Case, R., Curland, M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory. *Journal of Experimental Child Psychology*, 33, 386–404.
- Chi, M. (1978). Knowledge structures and memory development. In R. S. Siegler (Ed.), *Children's thinking: What develops?* (pp. 73–96). Hillsdale, NJ: Erlbaum.
- Choi, S., & Bowerman, M. (1991). Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. *Cognition*, 41, 83–121.
- Clark, H. (1996). *Using language*. Cambridge, England: Cambridge University Press.
- Ehrlich, S., Levine, S., & Goldin-Meadow, S. (2006). The importance of gestures in children's spatial reasoning. *Developmental Psychology*, 42, 1259–1268.
- Freudenthal, D., Pine, J. M., & Gobet, F. (2007). Understanding the developmental dynamics of subject omission: The role of processing limitations in learning. *Journal of Child Language*, 34, 83–110.
- Garrett, M. (1982). Production of speech: Observations from normal and pathological language use. In W. Ellis (Ed.), *Normality and pathology in cognitive functions* (pp. 19–76). London: Academic Press.
- Goldin-Meadow, S. (2003). *Hearing gesture: How our hands help us think*. Cambridge, MA: Harvard University Press.
- Hyams, N., & Wexler, K. (1993). On the grammatical basis of null subjects in child language. *Linguistic Inquiry*, 24, 421–459.
- Iverson, J. M., Capirci, O., & Caselli, M. C. (1994). From communication to language in two modalities. *Cognitive Development*, 9, 23–43.
- Johnston, J. R., & Slobin, D. I. (1979). The development of locative expressions in English, Italian, Serbo-Croatian and Turkish. *Journal of Child Language*, 3, 529–545.
- Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In M. R. Kay (Ed.), *The relation between verbal and nonverbal communication* (pp. 207–227). The Hague, the Netherlands: Mouton.
- Kendon, A. (2004). *Gesture*. Cambridge, England: Cambridge University Press.
- 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, 16–32.
- 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. *Journal of Language and Cognitive Processes*, 22, 1212–1236.
- Kita, S., van Gijn, I., & van der Hulst, H. (1998). Movement phases in signs and co-speech gestures, and their transcription by human coders. In I. Wachsmuth & M. Fröhlich (Eds.), *International Gesture Workshop, Bielefeld, Germany, 1997: Proceedings: Gesture and sign language in human-computer interaction* (pp. 23–35). Berlin, Germany: Springer-Verlag.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Mayberry, R., & Nicoladis, E. (2000). Gesture reflects language development: Evidence from bilingual children. *Current Directions in Psychological Science*, 9, 192–196.
- McNeill, D. (1992). *Hand and mind*. Chicago: University of Chicago Press.
- McNeill, D. (2005). *Gesture and thought*. Chicago: University of Chicago Press.
- McNeill, D., & Duncan, S. (2000). Growth points in thinking-for-speaking. In D. McNeill (Ed.), *Language and gesture* (pp. 141–161). Cambridge, England: Cambridge University Press.
- Müller, C. (1998). *Redebegleitende gesticulation: Kulturgeschichte-theorie-sprachvergleich* [Cospeech gestures: Cultural history-theory-cross-linguistic comparison]. Berlin, Germany: Spitz.
- Nicoladis, E. (2002). Some gestures develop in conjunction with spoken language development and others don't: Evidence from bilingual preschoolers. *Journal of Nonverbal Behavior*, 26, 241–266.
- Nicoladis, E., Mayberry, R., & Genesee, F. (1999). Gesture and early bilingual development. *Developmental Psychology*, 35, 514–526.
- Oh, K.-J. (2003). Manner and path in motion event descriptions in English and Korean. In B. Beachley, A. Brown, & F. Conlin (Eds.), *Proceedings of the 27th Annual Boston University Conference on Language Development* (pp. 580–590). Boston, MA: Cascadilla Press.
- Özçalışkan, S., & Goldin-Meadow, S. (2005). Gesture is at the cutting edge of early language. *Cognition*, 96, 101–113.
- Özçalışkan, S., & Slobin, D. I. (1999). Learning how to search for the frog: Expression of manner of motion in English, Spanish, and Turkish. In A. Greenhill, H. Littlefield, & C. Tano (Eds.), *Proceedings of the 23rd Annual Boston University Conference on Language Development* (pp. 541–552). Boston, MA: Cascadilla Press.
- Özyürek, A., Kita, S., & Allen, S. (2001). *Tomato Man movies: Stimulus kit designed to elicit manner, path and causal constructions in motion events with regard to speech and gestures* [Videotape]. Nijmegen, the Netherlands: Max Planck Institute for Psycholinguistics, Language and Cognition Group.
- Özyürek, A., Kita, S., Allen, S., Furman, R., & Brown, A. (2005). How does linguistic framing of events influence co-speech gestures? Insights from cross-linguistic variations and similarities. *Gesture*, 5, 215–237.
- Papafragou, A., Massey, C., & Gleitman, L. (2002). Shake, rattle, 'n' roll: The representation of motion in language and cognition. *Cognition*, 84, 189–219.
- Pine, K., Lufkin, N., Kirk, E., & Messer, D. (2007). A microgenetic analysis of the relationship between speech and gesture in children: Evidence from semantic and temporal synchrony. *Language and Cognitive Processes*, 22, 243–246.
- Pine, K., Nicola, L., & Messer, D. (2004). More gestures than answers: Children learning about balance. *Developmental Psychology*, 40, 1059–1067.
- Pulverman, R., Stootsman, J., Golinkoff, R. M., & Hirsh-Pasek, K. (2003). The role of lexical knowledge in non-linguistic event processing: English-speaking infants' attention to path and manner. In B. Beachley, A. Brown, & F. Conlin (Eds.), *Proceedings of the 27th Annual Meeting of the Boston University Conference on Language Development* (pp. 662–673). Boston, MA: Cascadilla Press.
- Rauscher, F. H., Krauss, R. M., & Chen, Y. (1996). Gesture, speech, and

- lexical access: The role of lexical movements in speech production. *Psychological Science*, 7, 226–230.
- Senghas, A., Kita, S., & Özyürek, A. (2004, September 17). Children creating core properties of language: Evidence from an emerging sign language in Nicaragua. *Science*, 305, 1779–1782.
- Slobin, D. (1987). Thinking for speaking. In J. Aske, N. Beery, L. Michaelis, & H. Filip (Eds.), *Proceedings of the 13th Annual Meeting of the Berkeley Linguistic Society* (pp. 345–445).
- Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and syntactic description: Vol. III. Grammatical categories and the lexicon* (pp. 57–149). Cambridge, England: Cambridge University Press.
- Valian, V., Hoeffner, J., & Aubry, S. (1996). Young children's imitation of sentence subjects: Evidence from processing limitations. *Developmental Psychology*, 32, 153–164.

Received December 4, 2006

Revision received February 29, 2008

Accepted March 20, 2008 ■