What Do Speech-Gesture Mismatches Reveal about Speech and Gesture Integration? A Comparison of English and Turkish

ASLI ÖZYÜREK
Koç University

0. Introduction

The semantic and temporal synchrony between speech and spontaneous hand gestures has been taken as evidence that speech and gestures are part of the same system (McNeill 1992, Kendon 1997). This paper questions the validity of this assumption through a cross-linguistic comparison.

Work by Talmy (1985), Slobin (1996), and others has shown that languages lexicalize the semantic components of spatial relations in different ways. The general question investigated in this paper is whether and how gestures synchronize *semantically* and *temporally* with the accompanying speech in languages where the semantic elements of a motion event are lexicalized differently—namely, in Turkish and English. This question is investigated in two studies.

The first study questions the semantic synchrony assumption. Do gestures represent semantic elements of a motion event in the same way in different languages, or does the representation in gestures vary from one language to another as the lexical and syntactic encoding of semantic elements vary? If the semantic synchrony assumption is true, then it is expected that gestures of speakers of different languages will vary with differences in lexicalization patterns of semantic elements in these languages.

The second study investigates the temporal synchrony assumption. Does the information in the gesture content temporally synchronize with the information in the accompanying speech content in different languages? If the temporal synchrony assumption is true, then what is expressed in gesture will temporally overlap with what is expressed in the co-temporal speech in different languages.

1. Study 1: Semantic Synchrony

With regard to expressing motion events, such as describing a ball rolling down a hill, English and Turkish differ from each other in the way they lexicalize manner and path elements of a motion event, as seen in (1).

(1) English and Turkish expressions of a motion event

English: Speakers can express manner and path components within one verbal clause.

"rolls	down"
V	satellite
manner	path

Turkish: Speakers have to use two different verbal clauses to express manner and path.

"yuvarlan-arak	iniyor"
V-roll-Conn	V-descend
manner	path

In previous work (Özyürek and Kita 1999), we tested whether Turkish and English speakers' gestures vary paralleling these differences in the lexicalization patterns.

1.1. Subjects

15 American English and 17 Turkish speakers participated in this study. All subjects were monolingual speakers.

1.2. Method

Each subject was asked to see and talk about an animated cartoon "Canary Row" (8 minutes). In the cartoon Sylvester the Cat attempts to catch Tweety Bird in different ways, each including a series of motion events.

1.3. Coding

One scene from the cartoon was selected for detailed analysis of speech and gesture. In this scene Sylvester swallows a bowling ball that Tweety Bird throws into his mouth and with the force of this bowling ball he rolls down the street and ends up in a bowling alley. The linguistic expressions and gestures were coded as follows.

Linguistic expressions: Verbal descriptions of this scene were coded for whether each speaker used (a) a verb + satellite construction, or (b) separate verbs to describe the manner and path components of the cat's rolling down the hill.

Gestures: Speakers' gestures that accompanied verbal expressions of this scene were categorized into three types.

a) Manner-only gestures: Representing the manner of the motion event only (i.e. hand(s) or fingers rotate/wiggle without any trajectory component)

b) Path-only gestures: Representing the path of the motion event only (i.e. hand(s) move along a lateral or sagittal trajectory without any rotation/wiggling of the hands or fingers)

c) Manner-path conflated gestures: Representing both the path and the manner of motion simultaneously (i.e. hand(s) move along a lateral or sagittal trajectory while the hands or the fingers rotate/wiggle)

1.4. Results

1.4.1. Speech

English speakers used one verbal clause (e.g. *He rolls down*) to express both manner and path in the rolling event, whereas Turkish speakers used two verbal clauses (e.g. *He rolls and goes down the street*), as shown in Table 1. In the English sample, three speakers also used a manner-only verbal clause (e.g. *He is rolling*) and one speaker used a path-only clause (e.g. *He goes down*) in addition to their main verbal clause.

Table 1. Percentage number of subjects who expressed path and manner within one clause or in two clauses in the Turkish and English sample

Number of subjects	One clause	Two separate clauses
English (N=15)	100 %	0 %
Turkish (N=17)	0 %	100 %

1.4.2. Gestures

We also looked at whether Turkish and English speakers varied in terms of the way they used manner-only, path-only, and manner-path conflated gestures. First, more Turkish speakers than English speakers used manner-only gestures to describe the scene, as can be seen in Table 2.

Table 2. Percentage number of subjects who used manner-only gestures at least once in their repertoire of gestures

Number of subjects	Used at least once	Never used
English (N=15)	7 %	93 %
Turkish (N=17)	50 %	50 %

Second, more Turkish speakers than English speakers used path-only gestures to describe the scene, as seen in Table 3.

Table 3. Percentage number of subjects who used path-only gestures at least once in their repertoire of gestures

Number of subjects	Used at least once	Never used
English (N=15)	43 %	57 %
Turkish (N=17)	69 %	31 %

Lastly, there was no difference between the number of Turkish and English speakers who used manner-path conflated gestures, as shown in Table 4.

Table 4. Percentage number of subjects who used manner-path conflated gestures at least once in their repertoire of gestures

Number of subjects	Used at least once	Never used
English (N=15)	71%	29%
Turkish (N=17)	69%	31%

1.5. Summary and Conclusion

The way Turkish and English speakers used their gestures to represent the elements of a motion event paralleled the differences in their lexicalization of semantic elements. Turkish speakers used more manner-only and path-only gestures, which paralleled the fact that they used separate verbal clauses to describe both manner and path. In contrast, English speakers mostly used manner-path conflated gestures, which paralleled the fact that they can express both elements within one verbal clause; few of them used manner-only or path-only gestures.

These findings show that gestures also differ when the linguistic encoding possibilities vary in different languages. This provides evidence for the view that gestures and speech have semantic synchrony across different languages.

2. Study 2: Temporal Synchrony

In this section I investigate whether speakers of Turkish and English also temporally coordinate the content of their gestures (i.e. manner-only gestures) with the content of their speech (i.e. manner-only clauses). That is, does the content of gestures match the speech content they synchronise with temporally? There could be two possibilities:

a) Match: Gesture content matches the co-temporal speech content. In this case there is semantic overlap (partial or total) between what is expressed in gesture and the co-temporal speech (e.g. Speech: *He goes down the street*, Gesture: path-only gesture).

b) Mismatch: Gesture content does not overlap with the co-temporal speech content. In these cases, of course the gesture content and speech content are related to each other in the sense that they both express features of the same motion event. Therefore, one might say that "mismatch" is not the right term for their kind of semantic relationship. However, here I define it as "mismatch" in the sense that speech and gesture represent different aspects of the same motion event (e.g. Speech: *asagi indi* 'went down', Gesture: manner-only gesture).

2.1. Results

In order to investigate the differences between English and Turkish the content of gestures and the co-temporal speech were compared for the rolling down scene used in the previous study.

For this analysis, the content of each gesture (manner-only, path-only, or manner-path conflated) was taken and compared to the content of speech it temporally synchronized with (see Table 5).

Table 5. Percentage of matching and mismatching *speech-gesture combination units* (gesture phrase + accompanying speech content) in the two languages for the "rolling down" scene

Number of speech-gesture combinations	Match between speech and gesture	Mismatch between speech and gesture
English (N=23)	100 %	0 %
Turkish (N=39)	70 %	30 %

Table 5 shows that there are more mismatches between gestures and the co-temporal speech content in Turkish than in English. The following example shows how mismatches between speech and gesture occur in Turkish. Here a speaker describes the rolling down of the cat in one sentence using five gestures.

(2) Matches and mismatches between speech and gesture in Turkish

a. Speech: *Top bi sekilde*
ball somehow
Gesture: manner-only MISMATCH

b. Speech: *ziplaya ziplaya*
hopping hopping
Gesture: manner-path conflated MATCH

c. Speech: *yuvarlana yuvarlana*
rolling rolling
Gesture: path-only MISMATCH

d. Speech: *sokaktan*
on the street
Gesture: path-only MATCH

e. Speech: *gidiyo*
goes
Gesture: path-only MATCH

Translation: 'Ball somehow, hopping hopping, rolling rolling goes along the street'

These results show that the temporal synchrony assumption does not hold all the time for Turkish speakers, as we see in (a) and (c) above.

However, even though Turkish speakers do not frequently synchronize their gesture content with the exact temporal speech content (i.e. within one gesture-speech combination unit), they might be trying to synchronise the information content in their speech and gesture at the *sentence level*. In order to test this possibility, the information content in the whole sentence used to describe the motion event scene was compared with the information content of the co-temporal gestures. For example, in the Turkish case above the information content in the whole sentence was compared to the information content revealed in the five gestures that overlapped with the whole sentence. In this example, if we take the whole sentence into consideration, the information content in speech and gesture match. The analysis in Table 6 shows that the content of gestures and the co-temporal accompanying speech content match for both speakers of Turkish and English at the sentence level.

Table 6. Percentages of match and mismatch between speech and gesture at the sentence level

Number of sentences	Match between speech and gesture	Mismatch between speech and gesture
English (N=22)	100 %	0 %
Turkish (N=17)	100 %	0 %

2.2. Summary and Conclusion

Cross-linguistic comparison showed that in languages where lexical encoding of semantic elements is different, the temporal synchrony between speech and gesture does not always hold. That is, what is expressed in gesture and the exact co-temporal speech content do not always match in different languages. However, it is possible that in different languages the temporal synchrony between speech

and gesture holds at different levels (e.g. the sentence). For example, what is at stake for a Turkish speaker is whether the information in the gesture content overlaps with the information expressed in the speech in the whole sentence, whereas for an English speaker the match is between what is in the gesture and the exact co-temporal speech content.

3. General Conclusion and Implications

In this study, the assumption that speech and gesture have semantic and temporal synchrony was tested by comparing speech and gestures in two languages where the mapping between lexical and semantic elements is different—namely, in Turkish and English.

Study 1 showed that the semantic synchrony assumption holds across languages. Speakers of different languages use different gestures with different lexicalization patterns of semantic elements even though they describe the same motion event. Study 2, however, showed that the temporal synchrony assumption does not always hold for speakers of different languages. That is, what is expressed in gesture and the content of the exact co-temporal speech content do not always match.

The mismatches found in the Turkish sample show that the temporal synchrony assumption (McNeill 1992) between speech and gesture should be modified in ways that can cover cross-linguistic differences. It is possible that the temporal synchronization of the content in speech and the content in gesture is regulated differently during speech and gesture production in different languages. These findings also have implications for the claim that speakers of different languages plan their thinking for speaking in different ways as revealed by different temporal synchrony relationships between their speech and gestures (Slobin 1996, Kita 2000).

References

Kita, S. 2000. How Representational Gestures Help Speaking. In D. McNeill (ed.), *Language and Gesture*. Cambridge: Cambridge University Press, 162-185.

McNeill, D. 1992. *Hand and Mind: What Gestures Reveal about Thought*. Chicago: University of Chicago Press

Özyürek, A., and S. Kita. 1999. Expressing Manner and Path in English and Turkish: Differences in Speech, Gesture and Conceptualization. In M. Hahn and S. Stoness (eds.), *Proceedings of the Twenty-First Annual Conference of the Cognitive Science Society*. London: Lawrence Erlbaum Publishers, 507-512.

Slobin, D. 1996. Two Ways to Travel: Verbs of Motion in English and Spanish. In M. Shibatani and S. A. Thompson (eds.), *Essays in Semantics*. Oxford: Oxford University Press, 195-217.

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*. Cambridge: Cambridge University Press, 57-149.

Department of Psychology
Koç University
Rumeli Feneri yolu, 819100
Sariyer, Istanbul
Turkey

aozyurek@ku.edu.tr