

What do speech-gesture mismatches reveal about speech and gesture integration? A comparison between English and Turkish

A. Özyürek

Department of Psychology
Koc University, Istanbul

Introduction

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

The work by Talmy (1985), Slobin (1996) and others have shown that languages lexicalize the semantic components of spatial relations in different ways. Thus the general question investigated in this paper is whether and how gestures synchronise *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 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? The second study investigates the temporal synchrony question: Does the information in the gesture content temporally synchronise with the information in the accompanying speech content in different languages?

Study 1: Semantic synchrony question (Özyürek & Kita, 1999)

With regard to expressing motion events, such as describing a ball rolling down a hill, English and Turkish differs from each other in the way they lexicalize manner and path elements of a motion event: English speakers can express manner and path components within *one* verbal clause: "rolls down". Whereas Turkish speakers have to use *two* different verbal clauses to express manner and path "yuvarlan-arak iniyor" (rolling descends). In previous work, we tested whether Turkish and English speakers' gestures vary paralleling these differences in the lexicalization patterns.

Subjects

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

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.

Coding

— Speech

Verbal descriptions of one scene were coded. 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*. Rolling scene was coded for whether each speaker used 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 categorised into 3 types:

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

Results

— Speech

The results showed that English speakers mostly used one verbal clause (e.g. he rolls down) to express both the manner and path in the rolling event, whereas Turkish speakers used two verbal clauses (i.e., he rolls and goes down the street).

Table 1.- Percentage of subjects who expressed Path and Manner within one clause or in two clauses in the Turkish and English sample

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

— Gestures

We also looked at whether Turkish and English speakers varied in terms of the way they use Manner-only, Path-only, and Manner-Path conflated gestures. Results showed that more Turkish speakers than English speakers used *Manner-only gestures* to describe the scene. Results also showed that more Turkish speakers than English speakers used *Path only gestures* to describe the scene. Finally we found that there was no difference between the number of Turkish and English speakers who used Manner-Path conflated gestures.

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

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

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

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

Table 4.- Percentage of subjects who used Manner-Path conflated gestures at least once in their repertoire of gestures

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

The way Turkish and English speakers used their gestures to represent the elements of a motion paralleled the differences in their lexicalization of semantic elements. Turkish speakers used more Manner-only and Path-only gestures that paralleled the fact that they use separate verbal clauses to describe both manner and path. Whereas English speakers mostly used Manner-Path conflated gestures that paralleled the fact that they can express both elements within one verbal clause and few of them used Manner-only and Path-only gestures. This provides evidence for the view that gestures and speech have semantic synchrony across different languages.

Study 2: Synchrony question

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 clause)? There could be two possibilities:

- Match:** Gesture content matches the co-temporal speech content: There is 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);
- Mismatch:** There is no overlap of content between what is expressed in gesture and the co-temporal speech (e.g., Speech: *yucarlandi* 'rolled', Gesture: Path only gesture).

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.

Analysis 1

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 content it temporally synchronized with. Table 5 shows that there are *more* mismatches between gestures and the co-temporal speech content in Turkish than in English, contradicting the temporal synchrony assumption.

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 %

The following example shows how mismatches between speech and gestures occur in Turkish :

- Speech *Top bi sekilde* Gesture: Manner-only MISMATCH
ball somehow
- Speech *ziplaya ziplaya* Gesture: Manner-path conflated MATCH
hopping hopping
- Speech *yucarlana yucarlana* Gesture: Path-only MISMATCH
rolling rolling

- d. Speech *sokaktan*
on the street Gesture: Path-only MATCH
- e. Speech *gidiyo*
goes Gesture : Path-only MATCH

Analysis 2

Even though Turkish speakers do not frequently synchronise their gesture content and with the *exact* temporal speech content (i.e., within one gesture-speech combination unit as can be seen in the example above), 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 events 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 5 gestures that overlapped with the whole sentence. Table 6 shows that if we take the whole sentence into consideration the information content in speech and gesture *match*.

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 %

Conclusion

The 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. The synchrony of speech and gesture could be organized at different levels (phrase versus sentence) in different languages.

General conclusion

In this study the assumption that speech and gesture have semantic and temporal synchrony is tested by comparing speech and gestures in two languages where the mapping between lexical and semantic elements are different, namely in Turkish and English. Study one 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 two, however, showed that temporal synchrony assumption does not always hold for speakers of different languages. That is, what is

expressed in gesture and the content in the exact co-temporal speech content does not always match.

The mismatches found in the Turkish sample show the necessity that temporal synchrony assumption between speech and gesture should be modified in ways that can cover cross-linguistic differences. It is possible that the synchronisation of the content in speech and the content in gesture is coordinated differently during the speech and gesture production in different languages. 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 *whole sentence*. Whereas, for an English speaker the match is between what is in the gesture and the exact co-temporal speech segment.

References

- McNeill, D. 1992, *Hand and mind: What gestures reveal about thought*, Chicago: University of Chicago Press.
- Ozyurek, A. & Kita, S. 1999, "Expressing manner and path in English and Turkish: Differences in speech, gesture and conceptualization". In M. Hahn & S. Stoness (eds.), *Proceedings of the twenty first annual conference of the cognitive science society*, Lawrence Erlbaum Publishers: London, 507-512.
- Slobin, D. 1996, "Two ways to travel: Verbs of motion in English and Spanish". In M. Shibatani & S.A. Thompson (eds.), *Essays in Semantics*, Oxford: Oxford University Press, 195-217.
- Talmy, L. 1985, "Lexicalization patterns: Semantic structure in lexical forms", *Language typology and syntactic description. Vol. III. Grammatical categories and the lexicon*, in T. Shopen (ed), Cambridge: Cambridge University Press, 57-149.