Motion event expressions in language and gesture: Evidence from Persian

Niloofar Akhavan (nakhavan13@ku.edu.tr)
Department of Psychology, Koç University
Rumelifeneri Yolu Sariyer 34450
Istanbul - TURKEY

Nazbanou Bonnie Nozari (nozari@jhu.edu)
Department of Neurology, Department of Cognitive Science, Johns Hopkins University,
1629 Thames Street, Suite 350, Baltimore, MD 21231, USA

Tilbe Göksun (tgoksun@ku.edu.tr)
Department of Psychology, Koç University
Rumelifeneri Yolu Sariyer 34450
Istanbul - TURKEY

Abstract

How do people conceptualize motion events and talk about them? The current study examines how gestural representations of motion events arise from linguistic expressions in Persian, which has characteristics of both Talmy's satellite- and verb-framed languages. We examined native Persian speakers' speech and gestures in describing 20 motion events. We focused on two motion event components: path (trajectory of motion like *up*) and manner (how the action is performed like *jumping*). Results indicated that when expressing motion, Persian speakers produced path in both speech and gesture, whereas manner was conveyed only through speech (mostly as adverbs). Additionally, dynamic gestures tended to occur in the same order they were uttered. The difference between path and manner findings asks for further research to examine language-gesture interaction in detail among different languages. Results also suggest refinement in gesture theories that argue for one-to-one correspondence between speech and gesture.

Keywords: motion events, gesture, language and thought, Persian, Farsi

Introduction

The relation between language and thought has been a question for decades. Throughout the history of philosophy it has been implied that the limits of language are the limits of thinking and people of different languages have different thought processes (Wittgenstein, 1921; Whorf, 1956). More recently, Berman and Slobin (1994) stated, "the particular ways of filtering and packaging information is shaped by one's native language" (p. 613). This hypothesis, "thinking for speaking," argues that thinking is provoked by the requirements of a linguistic code. In particular, the information to be expressed has to be tailored to speaking and must be compatible with the lexical and constructional resources of a given language (Slobin, 1996). However, others argue that language underspecifies thought and cognitive organization is independent of language (e.g., Gleitman & Papafragou, 2005). In this paper, we investigate how Persian speakers conceptualize motion events in both speech and gesture.

Motion Events

Languages vary in how they encode motion elements. A motion event consists of four semantic components; figure, ground, path, and manner. *Figure* refers to a particular point in space with respect to another object. *Ground* refers to another physical object, which serves as a reference point with respect to which the figure is located. *Path* refers to the translational motion and *manner* refers to motor pattern of the movement of the figure (Slobin, 1996). Talmy (1985, 1991) categorizes most of the world's languages into two major types of *satellite-framed* (S-framed) and *verb-framed* (V-framed) languages based on the core elements of path and manner. S-framed languages such as English (Germanic), Mandarin (Sino-Tibetan), and Russian (Slavic), integrate motion with manner in the main verb and express path with a verb particle or a satellite thus leaving the verb free to encode manner (e.g., run down (the hill)). On the other hand, V-framed languages such as Spanish (Romance), Turkish (Turkic), and Hebrew (Semitic) incorporate motion with path in the main verb and express manner in the subordinated verb (e.g., in Turkish, *koşarak çıktı* 'go up runningly') thus, using two verbal clauses to express both path and manner.

After studying various languages, Slobin (1996) concludes that lexicalization patterns are presumed to strongly impact thinking and formation of visuo-spatial representations. However, language may not directly influence event apprehension. Individuals' attention for encoding motion events can be allocated to their language-specific components only when they need to speak about these events. For example, in a study comparing English and Greek speakers, Papafragou, Hulbert, and Trueswell (2008) found that language on cognition effects arise only when language is recruited to achieve a task, but not during event perception in general. Thus far, using various methodologies, many languages have been analyzed through Tamly's approach. To our knowledge, there is only one

study examining how Persian speakers encode motion events in narratives (Feiz, 2011).

Feiz (2011) claims that Persian exhibits a mixed typology, with characteristics of both S-framed and V-framed languages. The similarity to S-framed languages is apparent in cases where path information is expressed in path satellites. An example is: (az tappe) *baala davidan* 'to run up (the hill),' in which *baala* 'up' is a satellite, and *davidan* 'to run' is a verb containing manner information. There are cases, in which path information is coded in the verb (e.g. *charkhidan* 'to pirouette'), but these are not common, and most of them need an additional preposition to become a transitive verb (e.g., *dor -e- […] charkhidan* 'to circle around'). On the other hand, the number of verbs that contain manner information (e.g., *davidan* 'to run') is also not high in Persian (Feiz, 2011), leaving manner information to be expressed mostly in other parts of speech, such as adverbs, *davan davan* raftan 'to go runningly' (davan davan = runningly; raftan = to go). In this sense, Persian more closely resembles a V-framed language.

In general, many Persian verbs contain neither path nor manner information. This is due to the special structure of most verbs, which are a combination of a noun + a light verb (e.g., *harekat* 'motion' + *kardan* 'to do' = to move). The light verbs that appear in such compounds are limited in number and have different levels of fidelity to their original meaning, for example, *kardan* in *harekat kardan* preserves its original meaning 'to do (motion)' but the verb *zadan* 'to hit' means something very different when used as a light verb in *ghadam zadan* 'to stroll'. Thus, the core semantics of the light verbs are rarely interpreted literally, and the meaning of the verb relies heavily on its noun component.

These noun components also vary in how much semantic information they convey. Some, like *harekat* 'motion' are broad and underspecified, thus, *harekat kardan* can mean any type of motion. Some, like *ghadam* '(slow) step', have more specific semantics, thus, conveying a little more than just the basic action. But since many nouns do not carry detailed information, peripheral details like path and manner are usually left to other parts, such as prepositions and adverbs. The construct described above makes Persian a unique case for studying the relationship between language and gesture.

Gesture use in Motion Events

Spontaneous co-speech gestures are bodily motions that embody a meaning related to the accompanying speech. These gestures are commonly used for thinking and communicating information that are visuospatial in nature (Alibali, 2005; Kita & Özyürek, 2003), providing a great deal of information about the internal structure of the speech. They also reflect internal cognitive process and provide a window on the embodied nature of mind (Hostetter & Alibali, 2008). Co-speech gestures are closely linked, both in meaning and time, to the speech they accompany (McNeill, 2005).

Nevertheless, there has been an unresolved debate about whether speech and gesture form a tightly integrated communication system or whether they originate from the same representational system or two separate but interrelated systems (Alibali, Kita, & Young, 2000; Butterworth & Hadar, 1989; Goldin-Meadow, 2003; Kita & Özyürek, 2003; Krauss, Chen, & Gotfexnum, 2000; McNeill, 1992). Research by McNeill (1992, 2005) supports the view that speech and gesture originate from the same representational system. Along these lines, McNeill (1992) suggested that since gesture conveys information not explicitly encoded in speech, it provides a unique window to view underlying thought.

Other theories suggest that speech and gesture are generated by two separate but highly interrelated systems (Alibali et al., 2000; Kita, 2000; Kita & Özyürek, 2003; Krauss et al., 2000). For example, Kita (2000) proposed that gestures help to organize and package visuo-spatial information into units of language. Moreover, Kita & Özyürek (2003) proposed the Interface Model that also predicts priming between language and gestures. They emphasize the influence of language on gestures, but suggest independent systems for speech and gesture. According to this model, language-specific aspects can also be represented in the gestures people use.

Cross-linguistic studies suggest that speakers of different languages produce different gestures for the same concept, and these gestures follow the linguistic structure of the utterances in their language (e.g., Kita, 2000; Kita & Özyürek, 2003; McNeill, 2000; McNeill & Duncan, 2000). For example, English speakers express manner together with path in their speech and gesture. In contrast, Spanish speakers omit manner in their speech but express it in a compensatory way in their gesture, and their path gestures follow the verbs (McNeill & Duncan, 2000). Further studies with English, Turkish, and Japanese speakers have revealed that the gestural representations mainly corresponded to language-specific encodings of motion events (Kita & Özyürek, 2003; Kita et al., 2007; Özyürek et al., 2005). In particular, English speakers use one verbal clause to express both elements of path and manner with one manner + path conflated gesture (e.g., 'running up' is represented by a gesture of making index and middle fingers move upward direction while alternating fingers), whereas, Turkish speakers use two verbal clauses thus they more likely use two separate gestures for path and manner (e.g., 'going up runningly' is expressed by an upward motion for 'go up' and then alternating index and middle fingers without upward movement for 'run'). In the Turkish case path is in the main clause (go up) and manner is in the subordinate (adverbial) clause (running).

The close correspondence of linguistic structure to gesture, however, has not been universally supported. In a recent study comparing English and Turkish monolinguals with controlled stimuli (similar to the ones used in this study), Karaduman et al. (2015) found that English speakers produced more manner and path combinations in their

speech compared to Turkish participants, as expected. Interestingly, this difference was not apparent in their spontaneous gestures. In contrast to the previous findings, they found that speakers of both languages used predominantly path gestures in their gesture use, despite the differences in their utterances.

The Current Study

We reviewed evidence on the sensitivity of gestures to the structure of the language that they accompany. These results point to a close correspondence between the linguistic and gesture systems. The question is whether there are other factors that limit this one to one correspondence. Results of Karaduman et al. (2015), which show similar gesture production in spite of linguistic differences, point to a common component to gestures, one that may mirror universals of human cognition, rather than specifics of a language.

The current study aims to investigate how gestural representations of motion events stem from linguistic expressions in Persian, the unique characteristics of which we reviewed earlier. This is the first controlled study to examine Persian in terms of differences in spatial language characteristics and the way these differences are manifested in spontaneous co-speech gestures.

Due to the structure of verbs, discussed earlier in the paper, Persian speakers are expected to express path of motion with prepositions and manner of motion as verb or adverb together with using auxiliary verbs. Our critical prediction concerns gestures: if linguistic forms correspond very closely to gestures, as expected by the Interface Model (Kita & Özyürek, 2003), we predict that Persian speakers would use two types of gestures: (1) when the speech resembles English expressions conflating path and manner information, such as *baala davidan* 'to run up,' there would be one conflated gesture representing both path and manner of motion; (2) when the speech resembles Turkish as in the case of *davan davan bala raftan 'to go up runningly'* there would be two separate gestures; one referring to path and the other referring to manner of motion. If factors other than linguistic form influence the production of gestures in Persian speakers, we might instead see dissociation between gesture and speech. If this arises due to a cognitive universal, we may observe the same pattern reported by Karaduman et al. (2015), with predominance of path gestures.

Method

Participants

15 monolingual Persian speakers between the ages of 18 and 30 (7 females and 8 males) were tested in Iran.

Task and stimuli

Video clips of different motion events developed and standardized by Göksun et al. (under review) were used. Participants watched 20 dynamic movie clips, depicting different motion events with randomized combinations of 10 manners (hop, skip, walk, run, cartwheel, crawl, jump, twirl, march, step) and 9 paths (through, to, out of, under, over, in front of, around, across, into). The actions were performed by a woman in an outdoor area (see Figure 1 for sample stimuli).

Figure 1: Sample stimuli from the experimental task. The picture is a still frame from the movie clip of a motion event: jump over. The yellow arrows indicate the direction of the movement.

Procedure

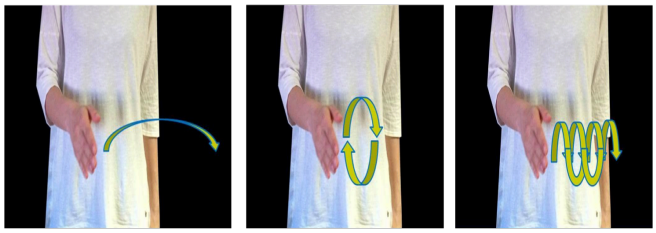
All participants were tested individually in their home environment. Before each task, two practice trials were given. Participants were then presented 20 trials in a randomized order. After watching each video, they were asked to describe the action in the clip. No instructions were given regarding gesture use. Participants' hands and torsos were videotaped.

Coding

Speech. All the speech was transcribed verbatim by the first author (a native Persian speaker). The transcribed speech was coded for the correct use of manner (how the action is performed) and path (the trajectory of action). First, for each trial, the coder assessed whether there was any manner and/or path information mentioned. Second, the pattern of speech responses in terms of path and manner information was categorized into groups of manner only, path only or manner + path together. Each trial containing a path or manner received a subcode as follows: For manner, it was coded as expressed in (1) a verb (*davidan* 'to run'), (2) an adverb (*Bodo bodo* [lit. 'run run'] 'in a running fashion'; *ley ley konan* 'hop hop doing'), and (3) the noun in a compound verb containing a light verb (*donbaal kardan* 'to chase). Path was also categorized into path as (1) a preposition (*kenare* 'side of'), (2) a verb (*charkhidan* 'to pirouette'), (3) a verb + a preposition (*dor charkhidan* 'to circle around'), (4) a light verb (*baala raftan* 'to go up'; *dar aamadan* 'to emerge'), (5) a light verb + a preposition (*az bein rad shodan* 'from pass do').

Gesture. Participants' spontaneous gestures were transcribed from the video. First, for each trial, the number

of gestures was coded. Second, the gestures were classified as static or dynamic. *Static gestures* referred to objects or locative properties of objects (e.g., pointing finger to refer to the preposition 'above'). *Dynamic gestures* involved the movements of hands that could represent the action of the person such as 'moving the index from left to right to display the direction of the motion'. Third, the dynamic gestures were classified into (1) manner only, (2) path only, and (3) path + manner together. Manner only gestures are those that enact the style of a motion without emphasizing the trajectory of the movement, the path (e.g., circular movement of the index finger in place to represent cartwheeling). Path only gestures show a direction without representing the manner (e.g., movement of the index finger in an arc pattern along the horizontal axis from right to left to represent 'across'). Path + manner gestures constitute both components simultaneously (e.g., circular movement of index finger along the horizontal axis from right to left to represent 'cartwheeling across'). Figure 2 represents these three types of gestures.



(a) (b) (c)

Figure 2: Sample gestures that represent (a) a path only motion (e.g., across), (b) a manner only motion (e.g., cartwheeling), and (c) a path + manner (cartwheeling across).

Results

Speech analyses

Participants expressed manner (M= 85.67%, SD = 8.42) and path (M=87.33%, SD =10.83) information similarly with no statistically significant difference between them, $t(14) = -.418$, $p = .682$. Next, we analyzed how participants encoded manner in speech. We found that people produced manner in adverbial form more frequently than in any other forms (M= 72.48%, SD =14.10), x^2 (2, N = 258) = 178.39, $p < .001$. For example, manner information was expressed as *'bepar bepar'* (in hop hop fashion) for hopping. We then analyzed path expressions and found that paths were mostly encoded with preposition + light verb ('*dor -e- derakht mire'* lit. = *around tree goes*, 'goes around the tree'; '*az khiyaban* rad shod' lit. = *from street cross did*, 'crossed the street'), x^2 (4, N = 268) = 380.32, p <.001 (see Table 1 for all numbers and corresponding percentages

Table 1: Number and percentages of manner and path expressions in speech

Manner	Number	Percentage
(1) Verb	40	15.0
(2) Adverb	187	72.5
(3) Light verb	31	12.3
Total	258	

Path	Number	Percentage
(1) Preposition	45	16.8
(2) Verb	2	.7
(3) Verb+Preposition	29	10.8
(4) Light verb	14	5.2
(5) Light verb+Preposition	178	66.4
Total	268	

Gesture analyses

Participants produced a total of 364 gestures in 237 out of 300 trials. On average, 72.5% of gestures were identified as dynamic, 9.3% of gestures were static, and 19.5% as beat gestures. In this paper, we only focused on dynamic gestures that referred to motions in the clips.

Next, we analyzed the overall pattern of dynamic gestures in terms of expressing manner and path information. Results showed that participants expressed significantly more path information in their gestures than manner information or path + manner information together (conflated), x^2 (2, N = 264) = 157.36, $p < .001$.

Last, we analyzed how participants used path and manner information in each trial. In these trial-based analyses, we coded whether participants used only path, only manner or both in each trial. For the trials where participants used both manner and path we also coded the order of their occurrence. The majority of dynamic gestures were identified as path only (M= 59.7%, SD =17.25) compared to manner only (M=11.2%, SD =16.98), manner + path conflations (M=8.7%, SD =10.41), or their combinations (M=20.4%, SD =12.56), x^2 (3, N = 206) = 138.58, p <.001. A closer look at the combined expressions indicated that people often used gestures for manner information before path information, the same order in which they were uttered, x^2 (1, N = 42) = 34.38 p<. 001. All numbers and corresponding percentages for the following analyses are provided in Table 2.

Table 2: Number and percentages of manner only, path only, and manner +path expressions in gesture

Dynamic Gesture	Number	Percentage
Manner Only	23	11.2
Path Only	123	59.7
Manner + Path (conflated)	18	8.7
Manner and Path	42	20.4
Total	206	

Speech – gesture relations

To further explore the information represented for motion event expressions, we analyzed whether path and manner were conveyed in both speech and gesture or in isolation. We found dissociation between the coexistence of the two gesture types and linguistic information. Participants tended to encode path information in both speech and gesture whereas manner was mostly produced within speech only, x^2 (3, N=474) = 58.91, $p < .001$ (see Table 3).

Table 3: Number and percentages of speech and gesture combinations

Combinations	Number	Percentage
Path Speech only	86	18.1
Manner Speech only	167	35.23
Path Speech-gesture	152	32
Manner Speech-gesture	69	14.5
Total	474	

Discussion

To our knowledge, this is the first study on motion event conceptualization in speech and gesture in Persian. We investigated how dynamic gestures contributed to motion expression in speech in a language that has characteristics of both Talmy's S- and V-framed languages.

As expected, Persian speakers frequently used adverbs, prepositions, and light verbs to describe both manner and path of the events. Interestingly, however, people's dynamic gestures mainly referred to path of motion, and not its manner. Manner + path conflated gestures made up only 8% of dynamic gestures. When looking at the overall and trial based gesture frequencies, Persian speakers tended to gesture for path information, whereas manner information was expressed in speech only.

The key question was whether variation in speech corresponded to the gestural expressions. The Interface Model suggests that there is an online interface between linguistic and gestural representations in utterance generation, in which spatial imagery is packaged into verbal units (Kita & Özyürek, 2003; Özyürek et al., 2005). Our results are only partially compatible with this account.

The majority of dynamic gestures described path of motion (60% of gesture) without including any manner information. This finding is in line with recent research by Karaduman et al. (2015), which indicated the predominant use of path gestures in contrast to manner gestures among both English and Turkish speakers. This supports a common and possibly a universal pattern to gesture production that may not be sensitive to linguistic structure. Why do English, Turkish and Persian speakers in our studies prefer path gestures to manner gestures? We cannot answer this question with certainty, but put forth possible reasons, leaving a more definite answer to future studies.

The easiness of manual movements for paths could be a factor. In particular, perhaps path is easier than manner to be displayed by hands due to its spatial configuration. Additionally, in this study we used naturalistic stimuli (as opposed to the cartoon events in the previous studies) and 20 sentences all containing different combinations of paths and manners. This imposes a high load on both language and gesture systems. While there might be close correspondence between representations in the two systems, the two may have different capacities and limits. For example, dual sequential representations might be harder to represent in the gesture system. If so, when faced with such demands, the system may drop the gesture that is manually more demanding.

Our results, however, provide support for Interface Model in 3 ways. First, overall there were very few manner verbs and manner as a verb + path as a preposition combinations in speech. As a result participants produced manner + path conflated gestures only in 9% of the gesture trials. This finding matches with the S-framed language characteristic of Persian (like English). Second, because path and manner information were mostly separated in two clauses as a property of V-framed languages (like Turkish), manner and path information were displayed in separate gestures, if any. Third and novel to this study, gesture sequences followed the same order as their linguistic counterparts. Past research has mostly ignored the effects of word order on gesture use. In Persian, subject–object–verb is the formal word order, but there is high flexibility in ordering words. However, adverbs usually do not come after the main verb (Megerdoomian, 2001). In keeping with this, we found that manner gestures that are expressed as adverbs in speech occurred before path gestures that were mainly expressed as a combination of preposition and light verbs at the final part of the sentence. This finding illustrates the role of language-specific encoding on gesture use, as claimed by the Interface Model.

In summary, the study of Persian, a language unique in its large number of noun + light verb compounds, and possessing the characteristics of both S- and V-framed languages, revealed the same pattern of correspondence between path gestures and the utterances describing them, as English and Turkish. The dominance of path gestures across languages may point to the universality of language-gesture interaction. On the other hand, other expressions such as manner + path conflations with manner + preposition utterances, manner and path information production in two separate clauses as in speech, and manner-path gesture orders paralleling word order in speech are compatible with the influence of language-specific structures on gesture. These findings call for closer inspection of factors involved in language-gesture interaction, and refinement of the Interface Model.

References

Alibali, M. W. (2005). Gesture in spatial cognition: Expressing, communicating, and thinking about spatial information. *Spatial Cognition and Computation*, *5*(4), 307-331.

Alibali, M. W., Kita, S., & Young, A. J. (2000). Gesture and the process of speech production: We think, therefore we gesture. *Language and cognitive processes*, *15*(6), 593-613.

Whorf, B.L. (1956). *Language, thought, and reality*. Cambridge, MA: MIT Press.

Berman, R. A., & Slobin, D. I. (Eds.). (2013). *Relating events in narrative: A crosslinguistic developmental study*. Psychology Press.

Butterworth, B., and Hadar, U. (1989). Gesture, speech and computational stages: A reply to McNeill. *Psychological Review, 96*, 168-74.

Feiz, P. (2011). Traveling through space in Persian and English: a comparative analysis of motion events in elicited narratives. *Lang. Sci*., *33* (3), 401–416.

Gleitman, L., & Papafragou, A. (2005). Language and thought. In K. Holyoak & R. Morrison (Ed.), *Cambridge Handbook of Thinking and Reasoning*. Cambridge: Cambridge University Press.

Gleitman, L.R. & Papafragou, A. (2013). Relations between language and thought. In D. Reisberg (Ed.), *Handbook of Cognitive Psychology*. New York: Oxford University Press.

Goldin-Meadow, S. (2003). *Hearing gesture: How our hands help us think*. Cambridge: Harvard University Press.

Göksun, T., Lehet, M., Malykhina, K., & Chatterjee, A. (under review). Spontaneous gesture and spatial language: Evidence from focal brain injury.

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic bulletin & review*, *15*(3), 495-514.

Karaduman, A. N., Çatak, E. N., Bahtiyar. S., Göksun, T., (March, 2015). *The role of gestures in describing motion in English and Turkish*. Poster accepted to present at the International Conference on Psychological Science, Amsterdam, Netherlands.

Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and gesture*. Cambridge: 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*(1), 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. *Language and Cognitive Processes*,*22*(8), 1212-1236.

Krauss, R. M., Chen, Y., & Gottesman, R. F. (2000). Lexical gestures and lexical access: A process model. In D. McNeill (Ed.), *Language and gesture*. Cambridge: Cambridge University Press

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

McNeill, D. (2005). *Gesture and Thought*. Chicago: University of Chicago Press.

McNeill, David & Susan D. Duncan (2000). Growth points in thinking-for-speaking. In D. McNeill (Ed.), *Language and gesture*. Cambridge: Cambridge University Press.

Megerdoomian, K. (2001). Event Structure and Complex Predicates in Persian: Special issue on the Syntax of Iranian languages. *Canadian Journal of Linguistics 46*(1/2), 97-125.

Özyürek, A., Kita, S., Allen, S. E., Furman, R., & Brown, A. (2005). How does linguistic framing of events influence co-speech gestures?: Insights from crosslinguistic variations and similarities. *Gesture*, *5*(1-2), 219-240.

Papafragou, A., Hulbert, J., & Trueswell, J. (2008). Does language guide event perception? Evidence from eye movements. *Cognition*, *108*(1), 155-184.

Slobin, D.I. (1996). Two ways to travel: verbs of motion in English and Spanish. In: Shibatani, M., Thompson, S.A. (Eds.), *Grammatical Constructions: Their Form and Meaning*. Oxford University Press, Oxford.

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and semantic description*. Cambridge, Cambridge University Press

Talmy, L. (1991). Path to realization: A typology of event conflation. *Proceedings of the Seventeenth Annual Meeting of the Berkeley Linguistics Society*. Berkley, CA.

Wittgenstein, L. (1922). *Logisch-Philosophische Abhandlung* (Ed.). New York, NY: Hardcourt, Brace & Company, INC.