Forming Event Units in Language and Cognition: A Cross-linguistic Investigation

Sarah Hye-yeon Lee (sarahhl@sas.upenn.edu)
Department of Linguistics, University of Pennsylvania
Philadelphia, Pennsylvania, USA

Ercenur Ünal (ercenur.unal@ozyegin.edu.tr)
Department of Psychology, Özyeğin University
Istanbul, Türkiye

Anna Papafragou (anna4@sas.upenn.edu)
Department of Linguistics, University of Pennsylvania
Philadelphia, Pennsylvania, USA

Abstract

Humans are surrounded by dynamic, continuous streams of stimuli, yet the human mind segments these stimuli and organizes them into discrete event units. Theories of language production assume that segmenting and construing an event provides a starting point for speaking about the event (Levelt, 1989; Konopka & Brown-Schmidt, 2018). However, the precise units of event representation and their mapping to language remain elusive. In this work, we examine event unit formation in linguistic and conceptual event representations. Given cross-linguistic differences in motion event encoding (satellite vs. verb-framed languages), we investigate the extent to which such differences in forming linguistic motion event units affect how speakers of different languages form cognitive event units in non-linguistic tasks. We test English (satellite-framed) and Turkish (verb-framed) speakers on verbal and non-verbal motion event tasks. Our results show that speakers do not rely on the same event unit representations when verbalizing motion vs. identifying motion event units in non-verbal tasks. Therefore, we suggest that conceptual and linguistic event representations are related but distinct levels of event structure.

Keywords: event; event segmentation; language production; motion; cross-linguistic analysis; conceptualization

Introduction

The physical world provides humans with an unorderly flux of dynamic experience. Humans, however, are surprisingly adept at understanding what is happening around them. One of the most fundamental abilities of the human mind is the ability to parse and organize this continuous input into discrete, individual event units.

In cognitive theories, events are viewed as perceptual or conceptual units. These units are defined by perceived changes in quality between two breakpoints in the external world (Newtson & Engquist, 1976; Radvansky & Zacks, 2014; Zacks & Tversky, 2001). On an influential account (Zacks et al., 2007), event segmentation happens spontaneously and automatically as the input unfolds, and is thus viewed as occurring prior to linguistic processing. A classic way of studying how viewers segment events is known as the Newtson-task (Newtson, 1973). In that task, participants were asked to watch a movie of an activity performed by an actor and to press a button whenever they thought that one meaningful event ended and another began: depending on the instructions, event boundaries could be identified at both a fine-grained as well as a coarse-grained level. In this context, events were viewed as units between transition points or boundaries.

Theories of event cognition (Radvansky & Zacks, 2014) have largely assumed that event units are discrete and contiguous, forming a contiguous timeline where one unit ends and another begins with no gap or overlap. However, many real-world events do not happen consecutively, due to potential overlaps between them, or interruptions within units. Consider, for example, knitting a sweater while watching TV, or cycling while crossing a forest. How does the idea of an event unit extend to those cases? In the current study, we look at motion events wherein path and manner components temporally overlap.

Mapping Conceptual Event Representations onto Language

Most importantly for present purposes, the nature of the units of nonverbal event apprehension affects the process whereby conceptual event representations are mapped onto language. Theories of language production assume that segmenting and construing an event offer a starting point for speaking about the event. That is, language production is taken to begin with conceptualization (deciding what to say), and later move onto formulation (deciding how to say it), and articulation (saying it; Levelt, 1989). However, little work has addressed the inner workings of conceptualization (Konopka & Brown-Schmidt, 2018).

Furthermore, across languages, there are considerable differences in the way information about events gets encoded into clauses. For example, expressing an event may require only one clause in one language but multiple clauses in another language (Talmy, 1985). In speech production, it has been proposed that the informational units suitable for speech formulation are what can be encoded in a clause in a given language (Bock & Cutting, 1992; Garrett, 1982; Levelt, 1989). It has also been claimed that single-clause sentences imply single unitary events (Croft, 1991; Déchaine, 1997;

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

DeLancey, 1983, 1984, 1991; Frawley, 1992; Goldberg, 1995; Haiman, 1983; Kiparsky, 1997; Rappaport Hovav & Levin, 1997; Shibatani, 1976; Wolff & Gentner, 1996). However, the link between cognitive and linguistic event units (clauses) has not been investigated in detail. In fact, most of the studies on event cognition have been conducted with English speakers (But see Defina, 2016; Swallow & Wang, 2020). Could non-linguistic event unit formation vary across speakers of typologically different languages?

This question connects to the issue of whether important aspects of cognition are universal or shaped by one's native language (Bowerman & Levinson, 2001; Gentner & Goldin-Meadow, 2003; Papafragou, Hulbert, & Trueswell, 2008; Gleitman & Papafragou, 2016; Lupyan, Rahman, Boroditsky, & Clark, 2020; Landau, 2022). It has been shown that, despite these cross-linguistic differences, speakers strikingly converge in some domains of non-linguistic event perception in the absence of message preparation demands (Papafragou, Massey, & Gleitman, 2002; Papafragou et al., 2008; cf. also Ünal et al., 2021). For example, in Papafragou et al. (2008), English and Greek speakers allocated more attention to the component that they were planning to encode in the main verb while viewing the events prior to speaking. Greek speakers attended more to path of motion than English speakers, and English speakers attended more to manner of motion than Greek speakers. Crucially, these cross-linguistic differences in attention allocation that emerged prior to speaking disappeared when participants freely inspected the events without preparing for linguistic encoding. However, this issue remains an active topic of investigation (see e.g., Athanasopoulos & Bylund, 2013; Bylund & Jarvis, 2011; Flecken, 2011; Flecken, von Stutterheim, & Carroll, 2014; von Stutterheim & Nüse, 2003; von Stutterheim, Andermann, Carroll, Flecken, & Schmiedtová, 2012).

Motion Event Units in Language and Cognition

Motion events provide an excellent domain for investigating the relationship between linguistic and cognitive event unit formation because typological differences in how languages encode motion have been well-documented (Allen et al., 2007; Aske, 1989; Kita & Özyürek, 2003; Naigles et al., 1998; Özçalişkan, 2013; Papafragou, Massey, & Gleitman, 2006; Slobin & Hoiting, 1994; Slobin, 1996; Talmy, 1975, 1985, 1991). These differences have direct consequences for the clausal units employed in descriptions of events that contain manner and path components.

Specifically, satellite-framed languages such as English typically convey manner (e.g., *slide*, *walk*) in the main verb and path (e.g., *into the phone booth*) in a non-verbal element. That is, both the manner and path information are expressed within a single clause, as in the English example (1).

(1) English: *one clause*
The woman ran into the phone booth.

Verb-framed languages such as Turkish may convey path in the main verb and express manner in a subordinated verb. For example, the manner and path information are conveyed in separate clauses in example (2).

(2) Turkish: *two clauses*
Kadın koş-arak telefon kulübesi-(n)e gir-di
woman run-CONN phone booth-DAT enter-PST
'The woman entered the phone booth while running.'

Empirical research linking linguistic and non-linguistic event unit formation is limited (cf. Wolff, 2003; Wolff, Jeon, & Li, 2009; Gerwien & von Stutterheim, 2018). These studies have claimed that there exists a close correspondence between linguistic and cognitive event units. Wolff (2003), for example, showed that within the domain of causative events, causal chains that could be described with single-clause expressions were more often construed as single events than chains that could not, suggesting that event units in cognition paralleled linguistic units (clauses). Gerwien and von Stutterheim (2018) also claimed that language-specific structural properties impact the cognitive process of event unit formation, based on patterns of event unit formation by French and German speakers in verbal and non-verbal tasks. However, cross-linguistic differences in the encoding of these events were not fully documented in this study.

Current Study

In this paper, we ask whether well-known typological differences across languages affect non-linguistic event unit formation. We test English and Turkish speakers on verbal and non-verbal event tasks, based on cross-linguistic contrasts in how English and Turkish typically encode motion path and manner. We use stimuli wherein path and manner components temporally overlap.

In Experiment 1, we use a language production task to establish cross-linguistic differences in the linguistic encoding of event units. In Experiments 2 and 3, we use non-linguistic event segmentation and individuation tasks to examine whether English and Turkish speakers form event units along these cross-linguistic patterns in situations where speech planning is not required. If there is a strict mapping between event units in language and cognition, the number of event units formed during non-verbal event perception will parallel the number of units (clauses) formed in the linguistic encoding of events. Alternatively, we may find that the two levels of representation are independent.

Experiment 1

In Experiment 1, we investigated linguistic event units in English and Turkish. We predicted that English and Turkish speakers would differ in linguistic event unit formation when describing motion events involving a manner and path component. Specifically, Turkish speakers should be more likely than English speakers to encode manner and path

information in separate clauses (i.e., form multiple linguistic event units).

Method

Participants We recruited 19 native speakers of English from University of Pennsylvania and 22 native speakers of Turkish from Özyeğin University in Turkey. Participants were granted course credit for participating in the study.

Stimuli The critical stimuli was adapted from the stimuli in Ünal, Manhardt, and Özyürek (2022) and ter Bekke, Özyürek, and Ünal (2022) and consisted of 16 short videos clips depicting Manner-Path events (e.g., running into a phone booth) that depicted a female actor moving with respect to a landmark object along a particular path with a particular manner. Each video clip was 2500ms long and motion lasted throughout the entire 2500ms. The stimuli included four different manners of motion: run, hop, twirl, and tiptoe, and four different paths of motion: to, from, into, and out of. In addition to Manner-Path events, the experiment included Manner-only event videos and non-motion event videos that depicted transitive events of an agent performing actions on objects. Figure 1 shows a screenshot from a sample Manner-Path event stimulus.

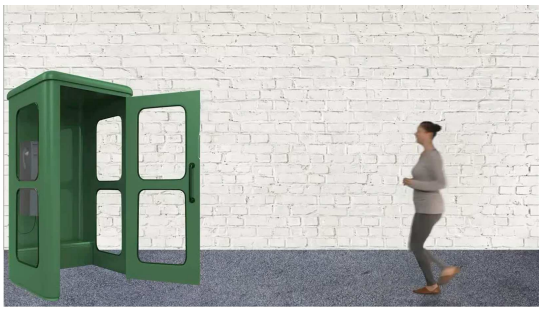

Figure 1. Sample Manner-Path event

Procedure Participants were tested in-person on their university campus. Participants were asked to watch a video and to describe what they saw to a confederate who sat across from them. The computer screen was only visible to the participant and not to the confederate. Before the experimental trials started, participants completed two practice trials (a woman opening an umbrella, a woman bending over and touching her toes), followed by an optional opportunity for questions and feedback. At the start of each main trial, participants saw a fixation cross for 1000ms. Next, an event video clip was played for 2500ms. Afterwards, a gray screen appeared. Once the gray screen appeared, participants were asked to provide a brief verbal description of what had happened in the video, to the confederate. Participants' verbal descriptions were recorded. After the participant had finished the description, the confederate pressed on a button to initiate the next trial. The experiment lasted for approximately 15 minutes.

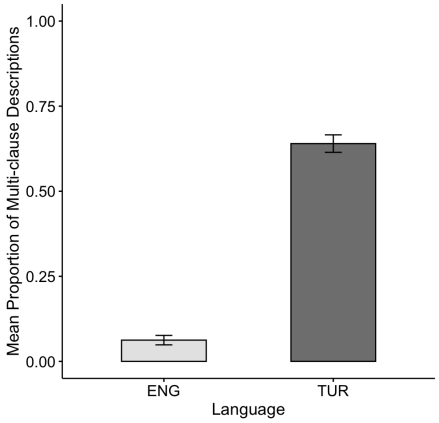

Figure 2: Proportion of multi-clause descriptions across language groups (Error bars represent ±SE).

Results

Participants' verbal responses were transcribed by a native speaker of the language. The responses were coded with respect to the number of clauses used in describing the depicted motion event. Participants' responses (single clause description vs. multiclause description) were analyzed using Generalized Linear Mixed Effects models (*glmer*). We coded Language (English vs. Turkish) using centered contrasts (English=-0.5, Turkish=0.5) and included it as the fixed effect. As random effects, we entered intercepts for subjects and items, as well as by-subject and by-item random slopes for the effects of Language. They were then reduced (starting with by-item effects) via model comparison, wherein only random effects that contributed significantly to the model ($p<.05$) were included (Baayen et al., 2008). The same analytical methods were used in subsequent experiments. The reported model included by-subject random slopes and intercepts, and by-item intercepts.

Results are plotted in Figure 2. As expected, in producing Manner-Path event descriptions, Turkish speakers provided more multiclause descriptions (M=64%) than English speakers (M=6.6%) (*glmer*, $p<.001$). In all multiclause descriptions in both English and Turkish, both manner and path components were mentioned.

Discussion

Data from the linguistic task confirmed the cross-linguistic differences between English and Turkish in terms of the encoding of manner-path motion events. As expected, Turkish speakers were far more likely to encode events involving a manner and a path component in multiple clauses than English speakers. In other words, Turkish speakers were more likely than English speakers to form multiple linguistic event units when encoding events involving a manner and a path component.

Experiment 2

Given the cross-linguistic differences between English and Turkish, in Experiment 2, we examined whether these

differences mapped onto the way that English and Turkish speakers' conceptual system segments events in a non-verbal task (Newtson-task). If there exists a direct mapping between linguistic and non-linguistic event units, Turkish speakers should indicate event boundaries more frequently than English speakers. Otherwise, both groups should perform similarly.

Method

Participants We recruited 38 native speakers of English and 39 native speakers of Turkish from Prolific.

Stimuli We used the same set of video stimuli as in Experiment 1. However, we did not include the non-motion event stimuli in this experiment, in order to avoid introducing any biases about the notion of change.

Procedure We follow methods of the classic Newtson-task. Participants were told that they would view short video clips. They were instructed to press the space bar on the keyboard to indicate when they perceive a change in the situation presented in the clip. In the instructions, "change in the situation" was further clarified with "... whenever something new happens in the scene." Participants were explicitly told not to press the space bar if they did not perceive a change (or changes) in the unfolding situation. In order to ensure that all participants had a chance to press the button, participants viewed each video twice in each trial. During the preview phase (viewing the video the first time), no overt response was required or registered. The actual test phase followed immediately where participants were shown the video once again. Participants were asked to respond during this test phase.

Results

All button presses were logged during each trial. In order to test whether the number of linguistic event units required in participants' native languages maps onto the number of cognitive event units they place, we transformed the button press frequency data into "mean hit probability," following the data analysis approach in Gerwien and von Stutterheim (2018). This analysis has the advantage of reducing inter-subject variability with respect to the number of times a subject pressed the button during one trial. We therefore simply coded whether a participant had pressed the button at least once to indicate a boundary (binary response variable; no = 0, yes = 1). If a participant pressed the button once while viewing a given video, it would indicate that they segment the video into two events. The mean hit probabilities were analyzed using Generalized Linear Mixed Effects models (*glmer*). We coded Language (English vs. Turkish) using centered contrasts (English=-0.5, Turkish=0.5) and included it as the fixed effect. The random effects in the reported model included intercepts for subjects and items.

The mean hit probabilities across language groups are plotted on Figure 3. When we compared the mean hit probability of English (M=0.44, SD=0.50) and Turkish speakers (M=0.50, SD=0.50), they did not differ in a meaningful way (*glmer*, β=-0.60, SE=0.87, z=-0.68, p=.49).

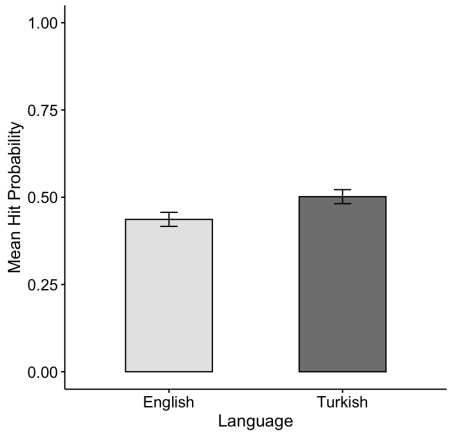

Figure 3: Mean hit probability across language groups (Error bars represent ±SE).

Discussion

Experiment 2 showed that Turkish speakers did not indicate event boundaries more frequently than English speakers. That is, both language groups performed similarly in a non-verbal task when speech planning was not required. Taken together with results from Experiment 1, these results show that linguistic and cognitive event units may diverge.

However, there is a question of whether the classic event segmentation method is suitable for understanding how people segment events that temporally overlap. For example, the manner and path components in our videos unfold over time in a concurrent manner: In an event of a woman twirling towards a tree, she twirls throughout her entire trajectory (manner), and she moves towards the tree throughout her entire trajectory (path). It is possible that a viewer understands this as two separate event units, but this may not be reflected in a method best suited for identifying boundaries between sequential events. In order to address such concerns, we conducted an event individuation task that does not involve placing explicit boundaries around event units as they unfold (Experiment 3).

Experiment 3

In Experiment 3, we conducted an event individuation task to investigate how speakers of English and Turkish construed the videos depicting manner and result components in terms of number of events. Event individuation was measured by having participants map the videos onto symbolic figures depicting either one or two events. Here, we were interested in whether English and Turkish speakers differ, but also whether these speakers understand manner and path components as separate units.

Method

Participants We recruited 46 new native speakers of English and 41 new native speakers of Turkish from Prolific.

Stimuli We used the same set of video stimuli as in Experiment 2.

Procedure Participants saw a series of videos. Each video was presented once followed by the instruction to select between two symbolic images (Figure 4), picking the one that "best depicts what happened in the video." This choice will force participants to judge if two things happened (left) or only one thing (right). No other instructions were given.

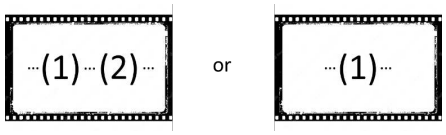

Figure 4: Choices presented in the individuation task

We used this mapping task as opposed to explicitly asking participants whether the video could be construed as one or two events in order to avoid potential biasing effects of language. For example, if we had asked participants whether a video depicted one or two "events", differences across language groups may arise due to subtle differences in the meaning of the word "event" and its translation across languages. (See Wolff et al., 2009 for a similar approach on individuation of causal events across languages.)

Results

The proportion of responses where participants chose the 2-event option is plotted in Figure 5. Participants' responses (2-event vs. 1-event) were analyzed using Generalized Linear Mixed Effects models (*glmer*). We coded Language (English vs. Turkish) using centered contrasts (English=-0.5, Turkish=0.5) and included it as the fixed effect. The reported model included as random effects intercepts for subjects and items. The mean proportion of 2-event responses did not differ across speakers of English (M=0.66, SD=0.48) and Turkish (M=0.65, SD=0.48) (β=-0.02, SE=0.31, z=-0.07, p=0.94). The model revealed a significant intercept (β=0.96, SE=0.41, z=2.35, p=0.019) indicating that overall participants were more likely to report 2-event responses and this did not change across languages.

Discussion

Results from Experiment 3 indicated that Turkish speakers did not identify more event units than English speakers: both language groups were likely to understand motion events involving concurrent manner and path elements as being composed of two units. This echoes findings from the non-verbal event segmentation task in Experiment 2. In the context of our earlier linguistic findings, this pattern points to a lack of a strict parallel between linguistic and cognitive event units: even though both language groups mostly agreed that two things happened in the videos in Experiment 3, the Turkish speakers were more likely to encode these two cognitive units in two clauses (linguistic units) compared to English speakers in Experiment 1.

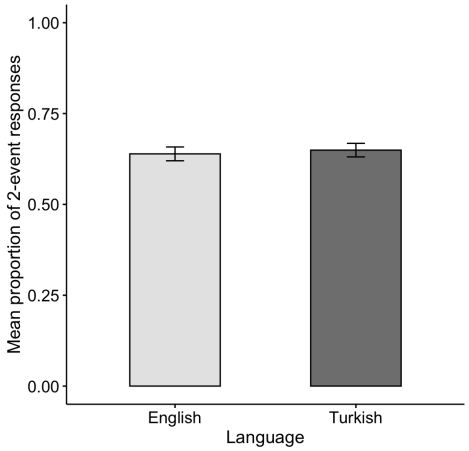

Figure 5: Proportion of 2-event responses across language groups (Error bars represent ±SE).

General Discussion

In this study, we investigated the relationship between conceptual representations and linguistic encoding in the domain of motion events. Specifically, we focused on linguistic and cognitive event unit formation in speakers of English and Turkish, based on well-documented cross-linguistic contrasts between satellite-framed and verb-framed languages in the way manner and path of motion are typically encoded. Our findings indicated that, despite robust differences in the ways English and Turkish speakers formed linguistic event units (Experiment 1), these differences did not map onto non-linguistic unit formation behavior, in either an explicit event segmentation task (Experiment 2) or an event individuation task (Experiment 3). Thus, important aspects of event cognition, such as event unit formation, are shared, possibly universal, regardless of one's native language.

It is important to note that our findings show striking convergence with another study that probed event unit formation in verbal and non-verbal tasks (Gerwien & von Stutterheim, 2018). In this study that tested French and German speakers based on the cross-linguistic differences in how direction is encoded in the respective languages, the potential unit boundaries were aligned with a specific perceptual change, i.e., change of direction. Whether speakers of different languages placed a boundary at these moments of change was largely driven by attention allocation on the direction change. In fact, Gerwien and von Stutterheim suggest the possibility that direction may be more salient for French speakers when identifying relevant factors for event segmentation. However, the idea that attention allocation has broad cognitive consequences even when we are not using linguistic representations as a means of encoding is dubious given other findings within motion event cognition such as Papafragou et al. (2008) and Trueswell and Papafragou (2010). Moreover, our study suggests a more nuanced approach to understanding the event unit formation process, whereby viewers can take a more 'holistic' look and extract

multiple event units from overlapping happening (i.e., in the lack of an obvious perceptual boundary).

Our results bear on theories of event cognition, as well as the broader relationship between language and thought. Despite the importance of event segmentation in event cognition research (Newtson & Engquist, 1976; Radvansky & Zacks, 2014; Zacks & Tversky, 2001), whether event segmentation is subject to linguistic effects has not been well-studied. Whether language affects non-linguistic cognition is a much-debated issue (Bowerman & Levinson, 2001; Gentner & Goldin-Meadow, 2003; Papafragou, Hulbert, & Trueswell, 2008; Gleitman & Papafragou, 2016; Lupyan, Rahman, Boroditsky, & Clark, 2020; Landau, 2022). Our results indicate that speakers of typologically different languages converge on event unit representations, therefore suggesting that event segmentation, at least in the absence of speech planning demands, is a process that precedes, or is independent from, language-related effects. These findings suggest that core aspects of (event) cognition are not shaped by the specificities of one's native language.

Furthermore, the findings presented here provide evidence for nonverbal event conceptualization as a distinct level of representation that differs from the linguistic representation of events. In the context of theories of language production (e.g., Levelt, 1989), our study suggests that the event units and representations that feed the conceptualization stage (deciding what to say) are shared across speakers of different languages. At later stages (formulation: deciding how to say it, articulation: saying it), speakers of different languages may diverge in terms of what information they package into language and how, depending on the constraints of their own language. We consider the possibility that a reorganization of the conceptual material takes place when the speaker starts formulating a preverbal message, thereby shaping the message according to language-specific requirements. Future research can elucidate how transitions from messages to the formulation stage proceed, and the timing and interaction of these processes.

Our studies investigated event unit formation behavior under minimal context because we did not provide participants with specific instructions about the size of the unit. However, event unit formation in both cognition and language is subject to the perceiver's perspective, which can be greatly influenced by the context or by the task. In the same way that event units in cognition can be identified at both fine-grained and coarse-grained levels (Kurby & Zacks, 2011; Newtson, 1973), event units in language are also very flexible. For example, pragmatic context concerning the speaker and the addressee's communicative goals (Clark, 1996; Grice, 1975; Sperber & Wilson, 1986) can influence how speakers form units in language. We plan to explore such effects in ongoing work.

Finally, our study was one of the first in the field to investigate how viewers understand overlapping happenings (events) in terms of units. In theories of event cognition that assume that event units are contiguous in time (Radvansky & Zacks, 2017), it is not expected that viewers would pick out event components that overlap in time as separate units. Thus, our findings suggest an alternative approach to understanding the nature of event units in both human cognition and language.

Acknowledgements

This work is supported by the NSF Grant BCS-2041171 awarded to AP and the TUBITAK grant 121K259 awarded to EÜ.

References

Allen, S., Özyürek, A., Kita, S., Brown, A., Furman, R., Ishizuka, T., & Fujii, M. (2007). Language-specific and universal influences in children's syntactic packaging of manner and path: A comparison of English, Japanese, and Turkish. *Cognition*, *102*(1), 16–48.

Aske, J. (1989). Path predicates in English and Spanish: A closer look. In *Annual Meeting of the Berkeley Linguistics Society* (Vol. 15, pp. 1-14).

Athanasopoulos, P., & Bylund, E. (2013). Does grammatical aspect affect motion event cognition? A cross-linguistic comparison of English and Swedish speakers. *Cognitive science*, *37*(2), 286-309.

Bock, K., & Cutting, J. C. (1992). Regulating mental energy: Performance units in language production. *Journal of Memory and Language*, *31*(1), 99-127.

Bock, K., Irwin, D., & Davidson, D. (2004). Putting first things first. In J. Henderson & F. Ferreira (Eds.), *The interface between language, vision and action: Eye movements and the visual world* (pp. 249–317). New York and Hove: Psychology Press.

Bowerman, M., & Levinson, S. C. (2001). *Language acquisition and conceptual development*. Cambridge: Cambridge University Press.

Bylund, E., & Jarvis, S. (2011). L2 effects on L1 event conceptualization. *Bilingualism: Language and Cognition*, *14*(1), 47-59.

Clark, H. H. (1996). *Using language*. Cambridge university press.

Croft, W. A. (1991). *Syntactic categories and grammatical relations*. Chicago, IL: University of Chicago Press.

Déchaine, R. (1997). Verb serialization and event composition. In R. De chaine & V. Manfredi (Eds.), *Object positions in Benue-kwa: papers from a workshop at Leiden University, June 1994* (pp. 47–66). The Hague: Holland Academic Graphics.

Defina, R. J. (2016). *Events in language and thought. The case of serial verb constructions in Avatime*. Radboud University Nijmegen: Doctoral dissertation.

DeLancey, S. (1983). Agentivity and causation: data from Newari. *Proceedings of the 9th annual meeting of the Berkeley Linguistics Society* (pp. 54–63). Berkeley, CA: Berkeley Linguistics Society.

DeLancey, S. (1984). Notes on agentivity and causation. *Studies in Language*, *8*, 181–213.

DeLancey, S. (1991). Event construal and case role assignment. *Proceedings of the 17th annual meeting of the*

Berkeley Linguistics Society (pp. 338–353). Berkeley, CA: Berkeley Linguistics Society.

Flecken, M. (2011). Event conceptualization by early Dutch–German bilinguals: Insights from linguistic and eye-tracking data. *Bilingualism: Language and Cognition*, *14*(1), 61-77.

Flecken, M., Von Stutterheim, C., & Carroll, M. (2014). Grammatical aspect influences motion event perception: Findings from a cross-linguistic non-verbal recognition task. *Language and Cognition*, *6*(1), 45-78.

Frawley, W. (1992). Linguistic semantics. Hillsdale, NJ: Erlbaum.

Garrett, M. F. (1982). Production of speech: observations from normal and pathological language use. In A. W. Ellis, *Normality and pathology in cognitive functions* (pp. 19-76). London: Academic Press.

Gentner, D., & Goldin-Meadow, S. (Eds.). (2003). *Language in mind: Advances in the study of language and thought*. Cambridge, MA: MIT Press.

Gerwien, J., & von Stutterheim, C. (2018). Event segmentation: Cross-linguistic differences in verbal and non-verbal tasks. *Cognition*, *180*, 225-237.

Gleitman, L., & Papafragou, A. (2016). New perspectives on language and thought. In K. Holyoak & R. Morrison (Eds.), *Oxford handbook of thinking and reasoning* (2nd , pp. 543–568). New York, NY: Oxford University Press.

Goldberg, A. E. (1995). Constructions. Chicago, IL: University of Chicago Press.

Grice, H. P. (1975). Logic and conversation. In P. Cole, & J. L. Morgan (Eds.), *Syntax and semantics: Vol. 3. speech acts* (pp. 41-58). New York: Academic Press.

Haiman, J. (1983). Iconic and economic motivation. *Language*, *59*, 781–819.

Kiparsky, P. (1997). Remarks on denominal verbs. In A. Alsina, J. Bresnan & P. Sells (Eds.), *Complex predicates* (pp. 473–499). Stanford, CA: CSLI.

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.

Konopka, A. E., & Brown-Schmidt, S. (2014). Message encoding. In V. Ferreira, M. Goldrick, & M. Miozzo (Eds.), *The Oxford handbook of language production* (pp. 3–20). Oxford: Oxford University Press.

Kurby, C. A., & Zacks, J. M. (2011). Age differences in the perception of hierarchical structure in events. *Memory & Cognition*, *39*, 75-91.

Landau, B. (2022). Language and thought: The lexicon and beyond. In A. Papafragou, J. C. Trueswell., & L. R. Gleitman (Eds.), *The Oxford Handbook of the Mental Lexicon* (pp. 236–C12.P64). Oxford University Press.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: The MIT Press.

Lupyan, G., Rahman, R. A., Boroditsky, L., & Clark, A. (2020). Effects of language on visual perception. *Trends in cognitive sciences*, *24*(11), 930-944.

Naigles, L. R., Eisenberg, A. R., Kako, E. T., Highter, M., & McGraw, N. (1998). Speaking of motion: Verb use in English and Spanish. *Language and cognitive processes*, *13*(5), 521-549.

Newtson, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology*, *28*(1), 28.

Newtson, D., & Engquist, G. (1976). The perceptual organization of ongoing behavior. *Journal of Experimental Social Psychology*, *12*(5), 436–450.

Özçalişkan, Ş. (2015). Ways of crossing a spatial boundary in typologically distinct languages. *Applied Psycholinguistics*, *36*(2), 485-508.

Papafragou, A., Massey, C., & Gleitman, L. (2002). Shake, rattle, 'n' roll: The representation of motion in language and cognition. *Cognition*, *84*, 189–219.

Papafragou, A., Massey, C., & Gleitman, L. (2006). When English proposes what Greek presupposes: The cross-linguistic encoding of motion events. *Cognition*, *98*(3), B75-B87.

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

Papafragou, A., & Grigoroglou, M. (2019). The role of conceptualization during language production: evidence from event encoding. *Language, Cognition and Neuroscience*, *34*(9), 1117-1128.

Radvansky, G. A., & Zacks, J. M. (2014). *Event cognition*. Oxford University Press.

Rappaport Hovav, M., & Levin, B. (1997). *On the nature of complex events*. Paper presented at the Conference on Semantics, Institute for Advanced Studies, Hebrew University of Jerusalem.

Shibatani, M. (1976). The grammar of causative constructions: a conspectus. In M. Shibatani (Ed.), *Syntax and semantics, vol. 6: The grammar of causative constructions* (pp. 1–40). New York: Academic Press.

Slobin, D. (1996). Two ways to travel: Verbs of motion in English and Spanish. In M. Shibatani & S. Thompson (Eds.), *Grammatical constructions: Their form and meaning* (pp. 195–219). Oxford: Clarendon Press.

Slobin, D., & Hoiting, N. (1994). Reference to movement in spoken and signed languages: Typological considerations. In *Annual Meeting of the Berkeley Linguistics Society* (Vol. 20, No. 1, pp. 487-505).

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition.* Cambridge, MA: Harvard University Press.

Swallow, K. M., & Wang, Q. (2020). Culture influences how people divide continuous sensory experience into events. *Cognition*, *205*, 104450.

Talmy, L. (1975). Semantics and syntax of motion. In J. P. Kimball (Ed.), *Syntax and semantics* (pp. 118–238). New York: Academic Press.

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology*

and syntactic description (pp. 57–149). New York, NY: Cambridge University Press.

Talmy, L. (1991). Path to realization: A typology of event conflation. In *Annual Meeting of the Berkeley Linguistics Society* (Vol. 17, No. 1, pp. 480-519).

ter Bekke, M., Özyürek, A., & Ünal, E. (2022). Speaking but not gesturing predicts event memory: A cross-linguistic comparison. *Language and Cognition*, *14*(3), 362-384.

Trueswell, J. C., & Papafragou, A. (2010). Perceiving and remembering events cross-linguistically: Evidence from dual-task paradigms. *Journal of Memory and Language*, *63*(1), 64-82.

Ünal, E., & Papafragou, A. (2016). Interactions between language and mental representations. *Language Learning*, *66*(3), 554–580.

Ünal, E., Richards, C., Trueswell, J., & Papafragou, A. (2021). Representing Agents, Patients, Goals and Instruments in causative events: A cross-linguistic investigation of early language and cognition. *Developmental Science*, *24*(6), Article e13116.

Ünal., Manhardt, F., & Özyürek, A. (2022). Speaking and gesturing guide event perception during message conceptualization: Evidence from eye movements. *Cognition*, *225*, Article 105127.

von Stutterheim, C. V., & Nüse, R. (2003). Processes of conceptualization in language production: Language-specific perspectives and event construal. *Linguistics*, *41*(5), 851-881.

von Stutterheim, C., Andermann, M., Carroll, M., Flecken, M., & Schmiedtová, B. (2012). How grammaticized concepts shape event conceptualization in language production: Insights from linguistic analysis, eye tracking data, and memory performance. *Linguistics*, *50*(4), 833-867.

Wolff, P., & Gentner, D. (1996). What language might tell us about the perception of cause. *Proceedings of the 18th annual conference of the Cognitive Science Society* (pp. 453–458). Hillsdale, NJ: Erlbaum.

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin*, *133*(2), 273.

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, *127*(1), 3-21.