Effects of Language Background on Gaze Behavior: A Crosslinguistic Comparison Between Korean and German Speakers

Florian Goller [1], Donghoon Lee [2], Ulrich Ansorge [1], and Soonja Choi [3,4]

[1] Faculty of Psychology, University of Vienna, Austria

[2] Department of Psychology, Pusan National University, Republic of Korea

[3] Department of Linguistics and Asian/Middle-Eastern Languages, San Diego State University, United States of America

[4] Faculty of Philological and Cultural Studies, University of Vienna, Austria

ABSTRACT

Languages differ in how they categorize spatial relations: While German differentiates between containment (*in*) and support (*auf*) with distinct spatial words—(a) *den Kuli IN die Kappe stecken* ("put pen *in* cap"); (b) *die Kappe AUF den Kuli stecken* ("put cap *on* pen")—Korean uses a single spatial word (*kkita*) collapsing (a) and (b) into one semantic category, particularly when the spatial enclosure is tight-fit. Korean uses a different word (i.e., *netha*) for loose-fits (e.g., apple in bowl). We tested whether these differences influence the attention of the speaker. In a crosslinguistic study, we compared native German speakers with native Korean speakers. Participants rated the similarity of two successive video clips of several scenes where two objects were joined or nested (either in a tight or loose manner). The rating data show that Korean speakers base their rating of similarity more on tight- versus loose-fit, whereas German speakers base their rating more on containment versus support (*in* vs. *auf*). Throughout the experiment, we also measured the participants' eye movements. Korean speakers looked equally long at the moving *Figure* object and at the stationary *Ground* object, whereas German speakers were more biased to look at the Ground object. Additionally, Korean speakers also looked more at the region where the two objects touched than did German speakers. We discuss our data in the light of crosslinguistic semantics and the extent of their influence on spatial cognition and perception.

KEYWORDS

INTRODUCTION

Does our everyday spatial language influence our perception and cognition? More specifically, does language-specific semantic categorization of spatial relations affect our nonverbal categorization and visual attention to objects? In this study, we investigate these questions comparing German and Korean, two languages that differ significantly in the way they categorize spatial relations. Here, we study for the first time if these language differences also lead to differences in how attention is deployed to *Figure* versus *Ground* objects in action recognition.

Objects can relate to one another in different ways. As shown in Figure 1, an object can be contained, supported (on a horizontal or vertical surface), attached or covered by another, or it can fit with the other tightly or loosely. However, languages differ greatly and significantly in the way they classify these relations (Bowerman, 2007; Choi & Hattrup, 2012), not only across unrelated languages (Levinson, Meira, & The Language and Cognition Group, 2003) but also among related languages, for instance, Germanic languages (Majid, Jordan, &

Corresponding author: Florian Goller, Faculty of Psychology, University of Vienna, Liebiggasse 5, 1010 Vienna, Austria. Phone: +43 1 4277 47148.
E-mail: florian.goller@univie.ac.at

Dunn, 2015). For example, Majid et al. (2015) reported that, among the twelve Germanic languages that they investigated, German and English belong to different language clusters in the way they categorize various types of spatial relation in static scenes. They further reported that the degree of crosslinguistic differences in the spatial domain is significantly more extensive than in other semantic domains, such as body part terms or color terms (Majid et al., 2015). Therefore, space is a good testing ground for investigating the relationship between language and cognition.

The debate on whether language shapes perception and cognition has continued over centuries and has become a core matter in cognitive science, particularly in recent years (Gentner & Goldin-Meadow, 2003; Gleitman & Papafragou, 2013; Wolff & Holmes, 2011). In this debate, crosslinguistic comparisons in the spatial domain have provided critical data. On the one hand, studies have reported data supporting a version of Whorf's (1956) hypothesis, namely that language significantly influences the way we perceive and categorize the world (Boroditsky, Fuhrman, & McCormick, 2011; Levinson, Kita, Haun, & Rasch, 2002; Pederson et al., 1998). On the other hand, studies have supported a *modular theory* claiming that cognition is universal, independent of language, and thus is unaffected by language-specific grammar (Gleitman & Papafragou, 2013; Li & Gleitman, 2002; Munnich, Landau, & Dosher, 2001). According to the latter view, any influence of language on perceptual or cognitive tasks is due to the mediation of language during the tasks, which can be suppressed by a concurrent linguistic activity (e.g., verbal interference). Therefore, effects of language on cognition are thought to be rather shallow (as they happen only online while carrying out a specific task in a specific condition) and do not permeate the underlying universal cognitive organization (Gleitman & Papafragou, 2013; Landau, Dessalegn, & Goldberg, 2010).

However, the depth of language influence may depend on the semantic domain. Recent studies (Athanasopoulos & Bylund, 2013; Choi & Hattrup, 2012; Lupyan, 2009; Thierry, Athanasopoulos, Wiggett, Dering, & Kuipers, 2009) that have examined perception and cognition (e.g., eye-movements, memory, similarity judgment) in different domains showed that language-effects are more automatized and internalized than the modularist view may claim. In particular, Choi and Hattrup (2012) reported that, in a spatial categorization task, English and Korean speakers showed significant, linguistically relevant differences regardless of a "language-interference" condition, where verbal thinking was actively suppressed. The study suggests that at least in the domain of spatial categorization, language has permeated and become an integral part of nonverbal cognition. Another element to consider in the language and cognition debate is that within a semantic domain (e.g., spatial categorization) *both* universal perceptual/cognitive tendencies and language-specific components may contribute to its organization, such that language-specifics affect some parts of a semantic domain more than others. For example, in spatial categorization, languages may categorize containment relations crosslinguistically similarly while they categorize support relations more diversely (Choi & Hattrup, 2012; Levinson et al., 2003; Yun & Choi, under revision). In

sum, recent studies on language and cognition have revealed that the interaction between the two is highly complex.

In investigating nonlinguistic behaviors related to spatial perception and cognition, researchers have studied participants' nonverbal categorizations and eye movements. To assess categorization, studies have examined participants' intuitive judgments about how similar spatial scenes/events are, either by forced choice or by rating degree of similarity (Choi & Hattrup, 2012; Engemann, Hendriks, Hickmann, Soroli, & Vincent, 2015; Gennari, Sloman, Malt, & Fitch, 2002). Studies have also measured participants' eye movements to specific areas of interest that are linguistically relevant (Papafragou, Hulbert, & Trueswell, 2008; Soroli, Hickmann, Hendriks, Engemann, & Vincent, 2015). Note that making a judgment or rating a degree of similarity involves controlled processes, determined by the instructions, whereas eye movements are not controlled by the instructions alone (cf. Flecken, Gerwien, Carroll, & von Stutterheim, 2015; Papafragou et al., 2008; Van Bergen & Flecken, 2017). Examining both types of behavior thus measures two partly independent ways of how language influences spatial cognition —that is, it measures language effects more exhaustively.

In the present study, we measured both similarity ratings and eye movements to assess the relationship between language and nonverbal spatial categorization. Previously, Choi and Hattrup (2012) used a triad design where participants first saw one target event in the middle of the screen for a few seconds. Then, the next screen appeared with two choice events presented simultaneously, one on the left and one on the right side of the screen. Participants were asked to choose which of these two choice events was more similar to the target event, thus engaged in a two-alternative forced-choice (2AFC) similarity judgment task. Similarly, participants in the current study compared two events directly to each other (rather than choosing one over the other in a 2AFC task) and indicated how similar they are on a range from 1 through to 9. Additionally, we also measured eye movements to linguistically-relevant areas of the spatial events while participants engaged in this similarity-rating task. In particular, we examined the allocation of attention to the Figure and the Ground objects as well as the contact area between them (see below). Overall, drawing on Choi and Hattrup's (2012) results on English and Korean, we expected to see a significant language-effect on the similarity-rating and the eye movements.

In the following, we first present critical differences in the spatial semantics between German and Korean and then present hypotheses about possible influences on spatial perception and cognition.

Language-Specific Spatial Categorization in German and Korean

German and Korean differ in classifying dynamic spatial events, such as putting an object into/onto another (see Figure 1). They also differ in the morphology used to categorize spatial relations: prepositions/particles in German and verbs in Korean. In German (similar to English), a major distinction in spatial categorization involves whether an entity is contained (*geben in*[1], "put in") or supported (see Figure 1). Support

relations are typically expressed with *geben auf*, "put on", whether they involve horizontal support, attachment, or covering, forming an abstract category of "support"[2]. In contrast to German, in Korean, a major distinction is made based on the degree of fit between linguistically defined Figure and Ground objects. In linguistics, a "Figure object is a moving or conceptually movable point whose paths or site is (…) variable (…)," while the "Ground object is a reference-point, having a stationary setting within a reference-frame, with respect to (…) the figure" (Talmy, 1978, p. 627). In particular, when a Figure object fits tightly with the Ground object (e.g., put rings tightly on poles; put pegs tightly into matching holes) Korean speakers use the same expression, *kkita* (or *kkiwu-ta*, with the causative suffix *-wu*), "fit tightly/interlock,"[3] collapsing across containment or support into one semantic category (see Figure 1; Bowerman & Choi, 2003; Yun & Choi, under revision). When the relation does not involve a tight-fit, a distinction is made between loose containment (*nehta*) and loose support (*nohta*). Yet again, Korean differs from German: The category of *nehta* (a word generally

referring to loose containment) includes loose encirclement as well, for example, a big ring *on* a thin pole. Thus, the division between containment and support is again blurred in the two loose-fit categories in Korean (*nehta* and *nohta*). Figure 2 summarizes how German and Korean semantically categorize the four types of spatial relation, loose-in, tight-in, loose-on, and tight-on, and shows the primary difference between the two languages: While German categorizes in terms of containment and support, thereby distinguishing between tight-in and tight-on, Korean collapses the two tight-fit relations into one semantic category, *kkita*.

It is important to note that when tight-fit is involved, Korean speakers use the verb *kkita* to denote tight-fitness between the Figure and Ground objects, disregarding the topological spatial relations between them, for example, containment or support. In contrast, German speakers consistently encode the topological relation between the Figure and the Ground, regardless of the degree of fit. Sentences 1A-2B (see Table 1) illustrate these crosslinguistic differences. Consider events of joining a pen cap and a pen: One can either move the cap to the pen or move the pen to the cap. Of course, one can also move both objects to join them, but the current study does not concern symmetric movement. In German, to express that one moves a pen cap (as a Figure) to cover
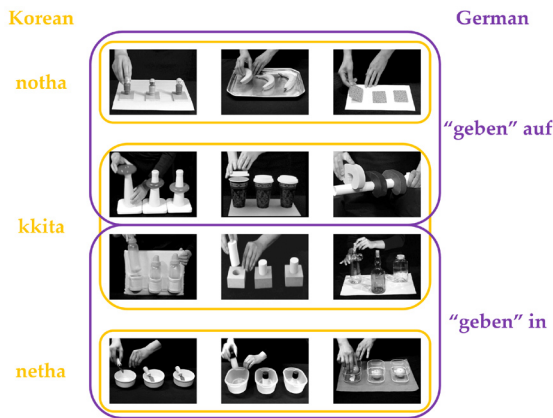
FIGURE 1.

Examples of the different spatial relations used in the current study. From top to bottom, example scenes for loose-on, tight-on, tight-in, and loose-in are shown. Left (in orange) is the Korean categorization, right (in purple) the German categorization.
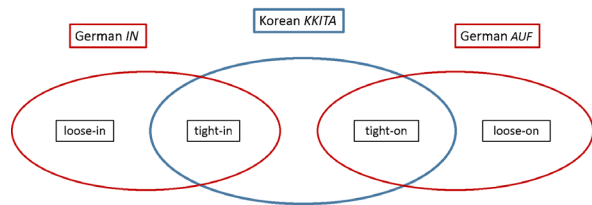
FIGURE 2.

Depicted is an abstract representation of the category memberships of or similarities between different video depictions of diverse spatial relations that were used in the present study. The major point of interest is that spatial relations in one and the same video that are similar according to Korean language (enclosed by the blue circle) fall into separate categories in German (red circles).

TABLE 1.

Examples of crosslinguistic differences between German and Korean

German	1A Pen cap moving:	Sarah steckt die	AUF	den Kuli
		'Sarah puts the cap	on	the pen.'
	1B Pen moving:	Sarah steckt den Kuli	IN	die Kappe
		'Sarah puts the pen	in	the cap.'
Korean	2A Pen cap moving:	Sara-ka	*pen-ttwukkeng-ul*	*pen-ey KKI-ta.*
		Sara-SUBJ	pen-cap-OBJ	pen-LOC tight-fit-DECL
	2B Pen moving:	Sara-ka	*pen-ul pen-ttwukkeng-ey*	*KKI-ta.*
		Sara-SUBJ	pen-OBJ pen-cap-LOC	tight-fit-DECL

Note. SUBJ – Subject marker, OBJ = Object marker, LOC = Locative marker, DECL = Declarative ending marker.

a pen, one encodes the spatial relation with AUF ("on") as in Sentence 1A, but when one moves a pen to insert it into a pen cap, one expresses it with IN ("in") as in Sentence 1B. But Korean typically uses the same spatial verb *kkita*, regardless of which object is moving (Sentences 2A and 2B), to denote the fitness between Figure and Ground.

Spatial semantics are essential to our everyday language: We frequently communicate with others about where things (Figures) are relative to a reference point (Ground). Developmentally, infants explore the physical properties of spatial relations (e.g., containment, support, tight-fit) virtually from the beginning of life and start categorizing and generalizing them from the preverbal period (Casasola & Cohen, 2002; Hespos & Baillargeon, 2001). Not surprisingly then, children produce spatial words from the one-word stage onwards and use them according to their language's specific semantics (Choi, McDonough, Bowerman, & Mandler, 1999). Furthermore, there is much evidence that language-specific semantics influence or even guide nonlinguistic spatial categorization from an early period (Casasola, Cohen, & Chiarello, 2003; Choi, 2006; McDonough, Choi, & Mandler, 2003) into adulthood (Choi & Hattrup 2012).

Given the fundamental nature of spatial cognition and early acquisition of spatial expressions, based on previous findings, we hypothesized that the critical differences summarized above between German and Korean—in spatial semantic categorization and (non-) distinction between linguistically defined Figure and Ground—have significant effects on speakers' nonverbal spatial categorizations particularly in those behaviors that are directly related to the linguistic expressions in question, namely similarity ratings and in eye movements to objects in dynamic spatial events.

SPECIFIC PREDICTIONS

If language influences spatial perception/cognition, we predicted the following results: In similarity ratings, we predicted that compared to Korean speakers, German speakers give a higher rating for the tight-in/loose-in pair and for the tight-on/loose-on pair but a lower rating for the tight-in/tight-on pair. We expected no significant differences between the two language groups for the tight-on/loose-in pair, a pair of relations that differ in both the tight-loose and the containment-support dimensions.

For eye movement behaviors, we expected Korean speakers and German speakers to differ in (a) the amount of looking to Figure versus Ground objects, and (b) areas of contact between Figure and Ground. We also expected that these crosslinguistic differences are particularly pronounced in tight-fit events compared to loose-fit events (as it is the tight-fit domain which the two languages categorize differently, see Figure 1), such that Korean speakers will attend to Figure and Ground equally often to ascertain the tight-fitness between the two objects, whereas German speakers may bias their attention to the Ground because the Ground is more likely to provide critical information about the topological relation: A concave container as Ground will feature a containment relation whereas a non-container Ground (e.g., flat or convex surface) will result in a support relation. With respect to contact

areas, Koreans should attend to them much more than German speakers do, again particularly for tight-fit events.

METHODS

Participants

We tested 15 participants (nine female, six male, M_{age} = 21.09; SD_{age} = 1.45) that were recruited among students of the Pusan National University (Republic of Korea) and 15 participants (ten female, five male, M_{age} = 23.36; SD_{age} = 3.48) that were recruited among students of the University of Vienna (Austria). The sample size was based on an a-priori power calculation using G*Power (Faul, Erdfelder, Buchner, & Lang, 2009), assuming a moderate effect size and a statistical power of 80%. This power analysis was based on a design with one two-step between-participants factor and one two-step within-participant factor. Based on the literature, an interaction between language and fitness (tight-fit versus loose-fit) was reasonable to assume. We only conducted one general power analysis for all data analyses reported in this paper.

All participants were native speakers of their respective language and were raised monolingual. Furthermore, all participants were naïve with respect to the research hypothesis, had normal or corrected to normal visual acuity, and received partial course credit. We adhered to the Declaration of Helsinki and to the ethical guidelines for human subject testing of the respective universities. Informed consent was obtained from all participants and, together with a language survey, a full debriefing followed the experiment. From the Korean-speaking sample, one participant was excluded due to excessive eye blinks which resulted in a data loss of more than 75% for that participant. One additional participant from the Korean sample and one participant from the German sample were excluded because the language survey indicated a bilingual upbringing. The final sample consisted, therefore, of 13 Korean speakers and 14 German speakers.

Apparatus

In the Pusan and the Vienna laboratories, we tested our participants under very similar conditions. All videos were displayed on a 19 in. monitor at a resolution of 1,024 × 768 pixels and a vertical refresh rate of 60 Hz. Viewing distance was kept stable at 64 cm by chin and forehead rests. Eye movements of the participants' dominant eye were recorded using an EyeLink 1000 Desktop Mount eye-tracker (SR Research Ltd., Kanata, Ontario, Canada) at a sampling rate of 1,000 Hz and an average accuracy of 0.15° of visual angle. The eye-tracker was calibrated using a 13-point calibration procedure. Prior to each trial, a drift check was performed, requiring participants to fixate on a centrally presented target circle. Recalibrations were performed if recorded fixation gaze average was outside a 4° radius of the pre-trial drift check target circle. The experimental procedure was implemented in Experiment Builder (SR Research Ltd., Kanata, Ontario, Canada), and the experiment was run on a computer under the Windows operating system. Manual re-

sponses were recorded as button presses with the right index finger on a keyboard.

STIMULI

Among all possible pairs of combinations involving the four spatial relations (tight-in, tight-on, loose-in, and loose-on), we selected four pairs (see Table 2). More specifically, we focused on three pairs (1-3 in Table 2) for which the two languages differ in semantic categorization and included one pair for which the two languages do not differ. In Pair 4, both languages distinguish the two relations (tight-on vs. loose-in) as they are maximally different in that they share neither tight-fit nor containment (or support) features.

We created a set of 32 videos (eight videos for each type of relation), each lasting for 4 s. We made multiple videos with different objects for each of the four spatial relations (tight-in, tight-on, loose-in, loose-on), each video containing a simple manual action, such as putting play-ing cards on a table (loose-on) or putting corks in bottles (tight-in, see Appendix 1). All actions were performed by a single female performer. The performer was dressed in black and filmed in front of a black back-ground. In all videos, only her hands were visible.

Each video consisted of three Figure objects (e.g., three cards) and one or more larger Ground objects (e.g., a table). All videos started with the first Figure already placed in or on the Ground (e.g., from the start, one sees the first card on the table). The performer put the second and the third Figures serially on or in the Ground over the course of the video. This redundancy in the spatial action in the videos (i.e., having three Figures) was intended to help the participants to fully perceive the action and the relation involved in the spatial event in question, which would be critical for performance in the later rating task. The performer's hand holding an object came into view from the top of the video screen. On average, the hand with the second Figure appeared on the screen about 100 ms after video onset, and the hand with the third Figure appeared on the screen about 1 s after video onset. Due to the diversity of objects used in our videos, these timespans varied between the videos.

The videos were shot using a Canon EOS 550D at a frame rate of 50 frames/s. The lighting conditions were kept constant for all videos. We decided to use grayscale videos to minimize the effects of salient

colors that varied across videos and are known to attract attention in an automatic manner (Itti, Koch, & Niebuhr, 1998; Theeuwes, 1991, 1992). We feared that too many of such salience influences could have equated the eye-movement behaviors of our participants so much as to potentially mask all language-specific differences.

PROCEDURE

All instructions were given in the native language of the partici-pant, that is, Korean for Korean participants and German for Austrian participants. Each trial started with a central fixation dot that was used for the eye-tracker drift check (see the Apparatus section). Afterwards, participants were shown two videos in succession. The video pairs were determined in advance to make sure that each spatial relation is equally often compared to the other spatial relations. The presentation-order of the respective videos in the pair (first versus second video) was, however, random and counterbalanced across all participants within their language group. The two videos were separated by a central fixation dot that was shown for 2 s. After the second video, a rating scale, ranging from 1 to 9, was presented on the screen. Participants were instructed to rate the similarity of the two videos they just saw. Importantly, participants were not told on what specific features or dimensions they should rate the videos. They were encouraged to give an intuitive and quick rating. After participants gave their rating, the next trial started. The overall experiment lasted for about 40 min, including preparation, instructions, and debriefing of the participant. After the main experiment, participants filled out a survey to confirm their language background and, most importantly, whether they were raised monolingual.

Eye-Tracking Coding and Data Processing

Eye-tracking samples were time-locked to the onset of each video. Since we tracked the eyes with 1,000 Hz, we had a possible maximum of 4,000 samples per stimulus. However, we had to exclude all samples that were recorded during eye blinks. Furthermore, we excluded all samples that were recorded during saccades. Using the SR Research algorithm, saccades were identified as a change in the recorded gaze direction of more than 0.15°, with an eye movement velocity above 30°/s, and an acceleration exceeding 8,000°/s². Overall, we had to ex-clude 17.27% of all samples, leaving us with an average of about 3,308 samples per video and participant to analyze.

We analyzed the time (i.e., the percentage of samples) that par-ticipants spent looking at the Figure and the Ground. In line with the linguistic definition, the Figure was always defined as the moving object, while the Ground was stationary. Figure and Ground objects were hand-coded as interest areas (or regions of interest) separately for each frame of each video. The Figure was always in the foreground especially if it moved over the Ground. The only exceptions are loose-in events where a concave container could partly obstruct the view of a Figure that was put into it. Additionally, the hands and the background were never part of the Figure or the Ground. See Figure 3 (left side) for an illustration.

TABLE 2.

Semantic Categories of the Stimuli Pairs as a Function of Language

Stimuli Pair	Semantic category in German	Semantic category in Korean
1. tight-in / tight-on	different (IN/AUF)	same (kkita)
2. tight-in / loose-in	same (IN)	different (kkita/netha)
3. tight-on / loose-on	same (AUF)	different (kkita/notha)
4. tight-on / loose in	different (AUF/IN)	different (kkita/netha)
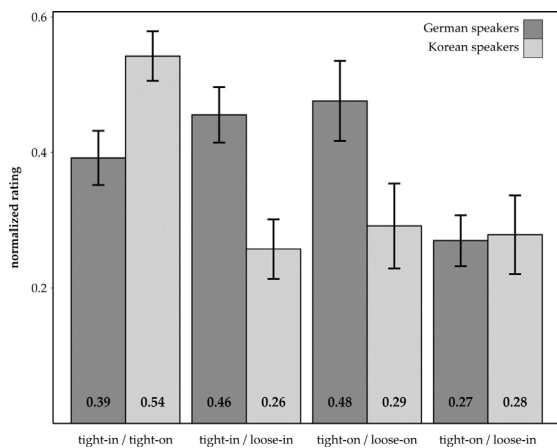
FIGURE 3.

Mean normalized ratings (*y* axis) as a function of spatial relation pairs (*x* axis) and language (separate bars: German speakers in dark grey and Korean speakers in light grey). The numbers on the bottom of each bar, indicate the mean for the respective condition. Error bars represent the *SEM*.

The Ground was always bigger than the Figures (both measured in pixels), $t(31) = 4.26$, $p < .001$, $d = 1.06$. This result is based on the point in time where all three figures were already present in the videos. To ensure that the Ground was consistently larger than the Figure, the size of the Figures was subjected to an analysis of variance (ANOVA) with the between-participants variables Fitness (tight; loose) and Topological Relation (IN; ON). No significant effects were found, all *F*s < 2.67, all *p*s > .133, indicating that the size of the Figures did not vary significantly across the different conditions. The same analysis on the Grounds rendered the same result, all *F*s < 0.55, all *p*s > .249.

As in each video the Figure objects were put in, on, or around a Ground object, separate from the Figure-Ground analysis, we also coded the Figure-Ground contact areas. This coding was also based on every frame of the videos. See Figure 3 (right side) for an illustration. By the end of the video, there were three contact areas present. The contact area was drawn around the immediate area where Figure and Ground joined or touched (such as the part of a container where an object that was inserted into it, see also Figure 3). The size of the contact areas varied, but this variation was equal across all conditions. This was ensured by an ANOVA of the size of the contact areas with the between-participants Variables Fitness (tight; loose) and Topological Relation (IN; ON) which yielded no significant results, all *F*s < 1.15, all *p*s > .201. Half of the defined contact area was part of the Figure and the other half was part of the Ground. The definition of contact areas was quite straightforward for tight-fit events because there is a clearly defined, visible touching area. As before, loose-in events were a bit problematic because the Ground partly obstructed the view of the Figure that was put into it. In this case, the contact area only covered the visible part of the Figure.

RESULTS

Similarity Rating

To make the rating data more comparable between the different language groups, we first normalized the data separately for each participant. Each individual rating was recalculated as (V − min V)/(max V − min V), where V represents the value of the rating in the original data set. This method allowed us to have ratings with different means and *SD*s but equal ranges. For an illustration of the results, see Figure 4.

The mean normalized ratings per participant and pair were subjected to a mixed ANOVA, with the within-participant variable Pair (tight-in/loose-in; tight-in/tight-on; tight-on/loose-in; tight-on/loose-on) and the between-participants variable Language (German; Korean). If the Mauchly test indicated that the assumption of sphericity was violated, *p* values were adjusted with the Greenhouse-Geisser correction. Note that the similarity ratings are based on pairs of video stimuli, not individual stimuli. First of all, we found no main effect of language, $F(1, 25) = 1.25$, $p = .228$, $\eta_p^2 = .06$, indicating that Korean and German speakers did not differ regarding their average similarly ratings across different pairs. However, there was a main effect of pair, $F(3, 75) = 7.29$, $p = .001$, $\eta_p^2 = .23$, and an interaction between pair and language, $F(3, 75) = 8.02$, $p = .001$, $\eta_p^2 = .24$. The tight-in/loose-in pair was rated as more similar by German speakers compared to Korean speakers, $t(25) = 3.31$, $p = .003$, $d = 1.27$. An analogous result was found for the tight-on/loose-on pair, $t(25) = 2.15$, $p = .042$, $d = .83$. Only the tight-in/tight-on pair was rated as more similar by Korean speakers compared to German speakers, $t(25) = 2.76$, $p = .011$, $d = 1.06$. As expected, there was no significant difference in the tight-on/loose-in pair, $t(25) = 0.13$, $p > .249$, $d = .05$.

We also checked whether there were significant differences between similarity ratings of each pair. Pairwise Bonferroni corrected comparisons were performed, separately for each language group. Korean speakers showed a significant difference between the pair tight-in/tight-on when compared to all other pairs (all *p*s < .007). No other differences were found for Korean speakers (all non-significant *p*s > .249). In contrast, for German speakers, we found significant differences between tight-on/loose-in and tight-in/loose-in (*p* = .032) as well
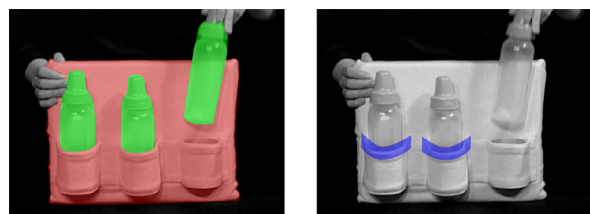
FIGURE 4.

Examples of the analysis of Figure (green) and the Ground (red), left side, and contact area (right side in blue). Figure was always defined as the moving object, while Ground was stationary. The hands and the black background were excluded from both coding methods.

as tight-on/loose-in and tight-on/loose-on ($ps < .013$). No other differences were found for German speakers (all nonsignificant $ps > .249$).

Eye-tracking Data

### LOOKING TIME AT FIGURE VERSUS GROUND

From all eye-tracking samples, 71.80% were on either Figure or Ground. From these data, we computed the proportion of samples (which corresponds to the proportion of viewing time) that was directed to the Ground. The results are illustrated in Figure 5. All proportions were arcsine transformed to approximate homogeneity of the variances. The proportions of samples directed to the Ground were subjected to an ANOVA with the within-participant variables Fitness (tight; loose) and Topological Relation (IN; ON) and the between-participants variable Language (German; Korean). If the Mauchly test indicated that the assumption of sphericity was violated, p values were adjusted with the Greenhouse-Geisser correction. Note that, unlike the similarity ratings, the analysis of the eye tracking data was based on individual stimuli and not on pairs of stimuli.

The ANOVA yielded significant main effects of language, $F(1, 25) = 11.71$, $p = .002$, $\eta_p^2 = .32$, and topological relation, $F(1, 25) = 72.77$, $p < .001$, $\eta_p^2 = .74$. These main effects, as well as the interactions between language and fitness, $F(1, 25) = 67.18$, $p < .001$, $\eta_p^2 = .73$, and fitness and topological relation, $F(1, 25) = 6.20$, $p = .020$, $\eta_p^2 = .20$, are best explained by resolving the also significant interaction of all factors of this ANOVA, $F(1, 25) = 4.75$, $p = .039$, $\eta_p^2 = .16$. Since this paper mainly focuses on crosslinguistic differences, we resolved this three-way interaction based on between-languages comparisons. We conducted such an analysis with separate ANOVAs for loose-fit and tight-fit events. For the loose-fit events, we found no significant results, all $Fs < 2.47$, all $ps > .129$, indicating no crosslinguistic differences. For the tight-fit events, we found a significant main effect of language, $F(1,$
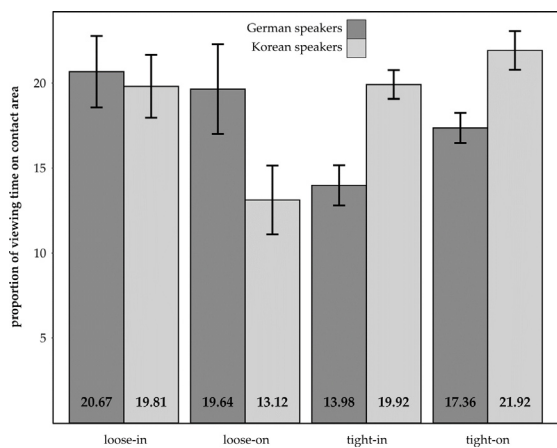

FIGURE 5.

Mean proportion of viewing time directed to the Ground (*y* axis) as a function of spatial relation pairs (*x* axis) and language (separate bars). The numbers on the bottom of each bar indicate the mean for the respective condition in percent. Error bars represent the *SEM*.

$25) = 41.04$, $p < .001$, $\eta_p^2 = .62$, showing that German speakers looked significantly more often at the Ground (68.65%) than Korean speakers did (49.03%). The interaction between language and topological relation showed a non-significant numerical trend, $F(1, 25) = 3.77$, $p = .063$, $\eta_p^2 = .13$. Korean speakers showed no significant difference between IN and ON events, $t(25) = -1.24$, $p = .237$, $d = -0.49$. The same was true for German speakers, $t(25) = 1.58$, $p = .139$, $d = 0.60$. No main effect of topological relation was found for the tight-fit events, $F(1, 25) = 0.02$, $p > .249$, $\eta_p^2 < .01$. In sum, these results indicate that in tight-fit events German speakers were more biased towards the Ground than to the Figure while Korean speakers distributed their viewing time more equally between Figure and Ground.

For the sake of completeness, we also analyzed whether each of the two language groups differed across the steps of the factors Topological Relation and Ritness. Therefore, we conducted separate ANOVAs for Korean and German speakers with the within-participant factors of Topological Relation and Fitness. Korean speakers showed an interaction between topological relation and fitness, $F(1, 12) = 9.68$, $p = .009$, $\eta_p^2 = .45$. No main effects were found, both $Fs < .09$, both $ps > .249$. In the IN relation, they looked slightly more at the Ground in loose-fit (51.04%) compared to tight-fit events (47.41%), $t(12) = 2.52$, $p = .027$, $d = 0.99$. In the ON relation, we found the opposite: a smaller proportion of viewing time to the Ground in loose-fit (46.36%) than in tight-fit events (50.66%), $t(12) = -2.26$, $p = .043$, $d = -0.89$. German speakers, in contrast, showed only a main effect of fitness, $F(1, 13) = 105.00$, $p < .001$, $\eta_p^2 = .89$, resulting from a higher proportion of viewing time to the Ground in loose-fit (51.80%) than in tight-fit events (68.65%). No other effects were found for the German speakers, $F < 1.28$, $p > .249$ in all cases.

We also checked post hoc whether the video presentation order (first video or second video of the pair) had an influence on gaze behavior. It was reasonable to assume that the participants' behavior might differ between the first and the second video. In the first video, participants may just freely look at the videos, while in the second video they may actively compare it to the first video. Analyzing the data separately for the first and second video may provide us with additional information that is not captured by an analysis collapsed over the video presentation order. Therefore, we repeated the main analysis separately for the first video and for the second video that was presented. The results were essentially the same, meaning that the crucial interaction between language, topological relation and fitness was significant in both cases, $F(1, 25) = 8.37$, $p = .008$, $\eta_p^2 = .25$, for the first video and $F(1, 25) = 6.08$, $p = .004$, $\eta_p^2 = .20$, for the second video.
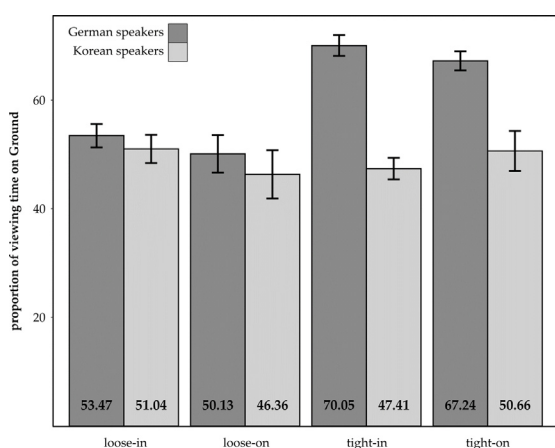
LOOKING TIME AT CONTACT AREA

This analysis was based on the same data as the Figure-Ground analysis, meaning that these two analyses are not independent. We consider this analysis as both complementary to and more fine-grained than the Figure-Ground analysis because it concentrates on an important area for identifying the spatial relation. We computed the percentage of viewing time of the whole video (4 s) that was directed to the contact area(s). Mean arcsine transformed percentages were subjected

FIGURE 6.

Mean proportion of viewing time directed to the contact area (*y* axis) as a function of spatial relation pairs (*x* axis) and language (separate bars). The numbers on the bottom of each bar indicate the mean for the respective condition in percent. Error bars represent the *SEM*.

to an ANOVA, with the within-participant variables Fitness (tight-fit; loose-fit) and Topological Relation (IN; ON), and the between-participants variable Language (German; Korean). If the Mauchly test indicated that the assumption of sphericity was violated, *p* values were adjusted with the Greenhouse-Geisser correction. The results are illustrated in Figure 6. We found a significant interaction between language and fitness, $F(1, 25) = 24.82$, $p < .001$, $\eta_p^2 = .50$, indicating that Korean speakers looked less into the contact area in loose-fit events (16.47%) than German speakers (20.16%), $t(25) = -2.27$, $p = .032$, $d = -0.87$. In tight-fit events, in contrast, Korean speakers looked more at the contact area (20.92%) than German speakers (15.67%), $t(25) = 4.26$, $p < .001$, $d = 1.64$. Furthermore, we found an interaction between topological relation and fitness, $F(1, 25) = 5.31$, $p = .030$, $\eta_p^2 = .18$. Only in the tight-fit events, a difference between IN and ON relations was present (16.84% vs. 19.56%), $t(26) = -3.54$, $p = .002$, $d = -0.96$, but not in the loose-fit events (20.26% vs. 16.50%), $t(26) = 1.42$, $p = .169$, $d = 0.39$.

As before, we also conducted separate ANOVAs for Korean and German speakers with the within-participant factors Topological Relation and Fitness. Korean speakers yielded a significant main effect of topological relation, $F(1, 12) = 9.36$, $p = .010$, $\eta_p^2 = .44$, as well as an interaction between topological relation and fitness, $F(1, 12) = 7.91$, $p = .016$, $\eta_p^2 = .40$. No main effect of fitness was found, $F(1, 12) = 3.02$, $p = .108$, $\eta_p^2 = .20$. For loose-fit events, Korean speakers looked more at the contact area in IN (19.81%) than in ON (13.12%) events, $t(12) = -2.44$, $p = .031$, $d = -0.96$. For tight-fit events, this effect was numerically yet non-significantly reversed (19.92% vs. 21.92%), $t(12) = 2.10$, $p = .058$, $d = 0.82$. German speakers, in contrast, yielded only a main effect of fitness, $F(1, 13) = 17.27$, $p = .001$, $\eta_p^2 = .57$, indicating a higher proportion of viewing time on the contact area in loose-fit (20.16%) than tight-fit (15.67%) events. No other results were found for the German speakers ($F < 0.89$, $p > .249$ in all cases).

As in the Figure and Ground analysis, we checked whether the video presentation order (first video or second video of the pair) led to

differential effects. Separate analysis for the first video and the second video that was presented yielded essentially the same results as above. The crucial interaction between language and fitness was significant for the first video, $F(1, 25) = 30.02$, $p < .001$, $\eta_p^2 = .55$, and for the second video, $F(1, 25) = 11.43$, $p = .002$, $\eta_p^2 = .31$.

GENERAL DISCUSSION

We have examined possible influences of language-specific semantic categorizations of spatial relations on three types of nonverbal behavior: similarity ratings, visual attention to Figure and Ground, and amount of looking time to contact areas between Figure and Ground objects. The results confirmed our overall hypothesis that language-specific semantic categorization has a significant impact on these behaviors.

As predicted, differences in similarity ratings between German and Korean speakers corresponded to the differences in semantic categorization between the two languages: Korean speakers perceived tight-relations (tight-in and tight-on) to be significantly more similar to each other than did German speakers. In contrast, German speakers perceived the two types of containment (tight-in and loose-in) and the two types of support (tight-on and loose-on) to be significantly more similar than did Korean speakers. In other words, the two language groups perceived the degree of similarity along the dimension delineated by their language-specific semantics.

In this study, we also examined possible relationships between spatial semantics and visual attention to Figure and Ground. As predicted, Korean speakers' looking behavior was significantly different from German speakers' behavior particularly in relation to tight-fit events (i.e., tight-in and tight-on events for which the two languages differ in their semantic categorization): Korean speakers spent similar amounts of time looking at Figure and Ground whereas German speakers looked at the Ground more than the Figure. To decipher the resulting topological relationship (i.e., containment or support), which is relevant to the categorization in German, the Ground gives more information than the Figure: A concave container as Ground will result in a containment relation whereas a non-container Ground (e.g., flat or convex surface) will result in a support relation. It is interesting that the longer looking time to the Ground by German speakers is restricted to tight-fit events only. For loose-fit events, German speakers did not look more at the Ground than at the Figure. Attending to Ground may be particularly necessary for tight-fit events since in these events Figure and Ground interlock with each other and thus the contour of individual objects is not salient. In contrast, these speakers do not need to focus so much on Ground when viewing a loose-fit event because the identity of the Ground is readily detectable.

There is another linguistic element (besides the German spatial prepositional system) that may have influenced German speakers to allocate their attention more to the Figure in loose events in comparison to tight-fit events: In German, a distinction is typically made between horizontal (*liegen/legen*) and vertical (*stehen/stellen*) orientation of a Figure object relative to the Ground. Such linguistic distinction may have promoted German speakers to attend to the Figure object (cf. Van

Bergen & Flecken, 2017). This may be the case in particular in loose-fit events because either orientation is possible when an object is loosely put on a surface (e.g., stellen/legen die Flasche auf den Tisch—"make stand/lay down a bottle on the table").[4] To resolve between the two possibilities, a study could be conducted that systematically contrasts between fitness (tight vs. loose) and posture (vertical vs. horizontal).

The results of looking times to contact areas also showed significant differences between the two languages. In tight-fit events, Korean speakers spent more time looking at contact areas than did German speakers. This difference implicates that contact areas between Figure and Ground are important for Korean speakers for assessing tight-fitness. By contrast, such an assessment is less important for speakers of German, probably because all they need to identify is whether the Figure goes into or onto Ground, which can be achieved by looking at the Ground, as discussed above. In loose-fit events, overall, Korean speakers looked less at the contact areas than German speakers. It is unclear how to interpret these data. For loose-in events, both groups looked at the contact area rather extensively. It is possible that speakers of both groups followed the Figure until it went to the bottom of the container or support (i.e., the contact area). For loose-on events, while German speakers showed a strong tendency to follow with their eye gaze until the end point of the Figure's trajectory (where the Figure touches the Ground surface), Korean speakers did not. Perhaps Korean speakers are less interested in continuing their eye gaze until the end point of motion, because the corresponding verb in Korean *nohta*, "put loosely on surface," (see Figure 1) also has the meaning of "release x from hand grip." Given the latter meaning of *nohta*, Korean speakers may just have attended to the Figure being released from the performer's hand. We also mentioned that the definition of the contact area is rather difficult and less strict for loose-fit events. Therefore, the results for the contact area in loose-fit events should be taken with a grain of salt.

Overall, our eye-movement data have shown an interaction between language and spatial relation: Crosslinguistic differences on nonlinguistic behaviors were significantly more pronounced for tight-fit events than for loose-fit events. Interestingly, this result corresponds to a recent study by Yun and Choi (under revision) on English and Korean, which reports greater crosslinguistic differences in semantic categorization for tight-fit events than for loose-fit events. Future studies need to examine the extent of this décalage (i.e., higher degree of language-specificity for tight-fit relations than for loose-fit relations) in other languages and explore possible cognitive implications of the phenomenon.

This brings us to the limitations of the current study. Our participants performed the similarity rating task in a silent environment. The study did not involve an interference condition (e.g., repeating nonsense syllables), which would have hindered or minimized possible verbally supported thinking during the task. However, studies that juxtaposed silent and interference conditions during a nonverbal task have reported conflicting results in terms of possible differences between the two types of conditions and the type of interference that would effectively suppress verbally supported thinking. For example,

studies have reported that articulatory tasks, such as counting numbers or repeating syllables, suppress verbal thinking (i.e., show no impact of language on nonverbal tasks) while other tasks, such as tapping, do not lead to the same result (cf. Gennari et al., 2002; Trueswell & Papafragou, 2010). It should be pointed out that most of these studies investigated the semantic domain of motion expressions in different languages– that is, whether a language highlights the path of motion (e.g., *into*, *out of*) or the manner of motion (e.g., *walk*, *run*) in its grammar (but see Roberson & Davidoff, 2000; Winawer, Witthoft, Frank, Wu, Wade, & Boroditsky, 2008, for studies on color; Meilinger & Bülthoff, 2013, for a study on spatial memory). Interestingly, in the domain of spatial categorization Choi and Hattrup (2012), who tested nonverbal similarity judgments in English and Korean speakers, found a language effect in *both* silent and interference (i.e., repeating syllables) conditions, and thus found no differences between the two conditions. The present study, which has examined spatial categorization in German and Korean speakers, should be extended further and include differential conditions to examine whether the relationship between language and cognition/perception differs across different semantic domains. Additionally, an interference task could also answer the question of whether our results are restricted to (nonconscious) verbal thinking or whether they extend to a deeper and more general level of influence of language on spatial perception. However, such experiments are beyond the scope of the present research, as we aimed to explore how spatial relations might influence nonverbal categorization and visual attention to objects involved in spatial events.

Additionally, the time course with which participants deploy their attention to Figure and Ground might be also of interest for future studies. In the current study, we defined all of the three Figures as one area of interest. A more fine-grained analysis could define the different Figures (and Grounds as well as the contact areas) as different interest areas. By doing so, one could answer the question at what point in time the crosslinguistic differences in the viewing behavior become apparent.

Last, one further extension of our experiment needs to be discussed. Our results are obtained from a similarity rating task. We did not instruct participants to base their rating specifically on the spatial relations but to give an intuitive rating. However, participants might have picked up on spatial relation as an implicit rating dimension as it was the only dimension that was consistently present in all videos. All other perceivable dimensions (such as size or shape of the objects) varied randomly and across the videos and therefore may have not been a feasible basis for a rating. As a result, we may have pushed the effects of spatial language on perception. More compelling tests in the future could use a similarity rating task with objects or actions allowing more than one classification to see if a linguistically marked classification influences the ratings even where other obvious but not crosslinguistically marked features (e.g., object colors, action directions) would invite alternative categorizations of the objects. In addition, a control condition, in which no crosslinguistically marked actions or objects are used in a similarity rating task could be employed to confirm that in such a condition no differences between the Korean and German language

exist. However, note that the rating task of the present study already contained one such condition, namely, the comparison between tight-on and loose-in relations (i.e., Pair 4 in Table 2). These two relations are different both in terms of fitness (tight vs. loose) and of topological relation (ON vs. IN). Thus, the two relations should be categorized as being different by both Korean and German speakers, a prediction that was supported by the much weaker language-dependent behavioral differences in this condition.

In summary, the present study has shown that speakers of German and Korean diverge significantly in nonverbal categorization and attentional behaviors in correspondence to the semantic differences between the two languages. More generally, the present study has shown that the spatial categorization we use every day has a significant impact on our nonverbal behaviors that are directly relevant to it—the way we nonverbally categorize spatial relations and the kinds of things we pay attention to in a spatial event. To that extent, the present study supports the Whorfian hypothesis (1956). Importantly, the present study has also revealed that the language effect on nonverbal behaviors, specifically eye movement behaviors, varies across subdomains: The effect occurred most prominently for tight-fit relations, for which the two languages differed critically in their semantic categorization. In comparison, nonverbal behaviors for loose-fit relations did not generate significant crosslinguistic differences. As mentioned earlier, this may reflect a higher degree of similarity in the way languages categorize loose-fit relations than for tight-fit relations.

As discussed earlier, studies have reported both universal cognitive/perceptual tendencies and language-particular components in the way languages categorize the semantic domain of space (Choi & Hattrup, 2012; Levinson et al., 2003; Yun & Choi, under revision). In particular, Yun and Choi (under revision) have proposed greater crosslinguistic differences in semantic categorization for tight-fit events than for loose-fit events. The present study coheres with this proposal in that crosslinguistic differences did not occur across the board in nonverbal behaviors, but rather in the subdomain of tight-fit relations where language seems to be the principle guide for categorization (Choi & Hattrup, 2012). Thus, there is a complex interaction between language-specific semantics and cognition/perception. However, we limit our claim on the specific nature of interaction to the domain of space, and in particular the domain of spatial categorization. To understand the relationship between language and cognition in other domains, an in-depth analysis of the semantics of the target languages in those domains should be conducted hand in hand with systematic investigation of relevant cognitive and perceptual behaviors.

AUTHOR NOTE

FG, UA, and SC developed the study concept and the study design. FG programmed the experiment, performed the data collection in Austria, and the data analysis of both language groups. DL performed the data collection in Korea. The results were interpreted by FG and SC. FG and SC drafted the manuscript, DL and UA provided critical revisions. All authors approved the final version of the manuscript for submission.

FOOTNOTES

[1] We admit that the verb *geben* may sound slightly odd in the context of the prepositions *in* and *auf*, but the German language uses a variety of verbs to describe spatial relations (Berthele, 2012), and *geben* is not so uncommon in Austrian German—the language of many of our participants. For illustration purposes, here, we used the more general verb *geben* because it is rather neutral and fits all relations.

[2] German speakers may use *an* for some type of attachment such as hanging or hooking. However, our stimuli did not include hanging or hooking events.

[3] It has been discussed in the literature that *kkita* involves "effort-fullness" between two objects and therefore has a component of manner of motion (e.g., Kawachi, 2007). By virtue of putting x in a tight-fit relation with y, many of tight-fitting actions involve effort or friction. While we note that *kkita* has such manner component, our view, however, is that *kkita* expresses foremost a spatial relation (i.e., it is a path term) that contrasts semantically and paradigmatically with other path terms (e.g., *nehta*, *nohta*). This has been shown empirically in linguistic description studies involving various types of spatial relation (cf. Chang, Choi, & Ko, 2015; Yun & Choi, under revision).

[4] We thank an anonymous reviewer for suggesting this possibility.

REFERENCES

Athanasopoulos, P., & Bylund, E. (2013). Does grammatical aspect affect motion event cognition? A cross-linguistic comparison of English and Swedish speakers. *Cognitive Science, 37*, 286–309. doi: 10.1111/cogs.12006

Berthele, R. (2012). On the use of PUT verbs by multilingual speakers of Romansh. In B. Narasimhan & A. Kopecka (Eds.), *Events of putting and taking: A crosslinguistic perspective* (pp. 145–165). Amsterdam, The Netherlands: John Benjamins.

Boroditsky, L., Fuhrman, O., & McCormick, K. (2011). Do English and Mandarin speakers think about time differently? *Cognition, 118*, 123–129. doi: 10.1016/j.cognition.2010.09.010

Bowerman, M. (2007). Containment, support, and beyond: Constructing topological spatial categories in first language acquisition. In M. Aurnague, M. Hickmann, & L. Vieu (Eds.), *The categorization of spatial entities in language and cognition* (pp. 177–203). Amsterdam, The Netherlands: John Benjamins.

Bowerman, M., & Choi, S. (2003). Space under construction: Language-specific spatial categorization in first language acquisition. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and cognition* (pp. 387–428). Cambridge, MA: MIT Press.

Casasola, M., & Cohen, L. B. (2002). Infant categorization of containment, support and tight-fit spatial relationships.

Developmental Science, 5, 247–264. doi: 10.1111/1467-7687.00226

Casasola, M., Cohen, L. B., & Chiarello, E. (2003). Six-month-old infants' categorization of containment spatial relations. *Child Development, 74*, 679–693. doi: 10.1111/1467-8624.00562

Chang, F., Choi, Y., & Ko, Y. (2015). Why loose rings can be tight: The role of learned object knowledge in the development of Korean spatial fit terms. *Cognition, 136*, 196–203.

Choi, S. (2006). Influence of language-specific input on spatial cognition: Categories of containment. *First Language, 26*, 207–232. doi: 10.1177/0142723706060748

Choi, S., & Hattrup, K. (2012). Relative contribution of cognition/perception and language on spatial categorization. *Cognitive Science, 36*, 102–129. doi: 10.1111/j.1551-6709.2011.01201.x

Choi, S., McDonough, L., Bowerman, M., & Mandler, J. M. (1999). Early sensitivity to language-specific spatial categories in English and Korean. *Cognitive Development, 14*, 241–268. doi: 10.1016/S0885-2014(99)00004-0

Engemann, H., Hendriks, H., Hickmann, M., Soroli, E., & Vincent, C. (2015). How language impacts memory of motion events in English and French. *Cognitive Processes, 16*, 209–213. doi: 10.1007/s10339-015-0696-7

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149–1160. doi: 10.3758/BRM.41.4.1149

Flecken, M., Gerwien, J., Carroll, M., & von Stutterheim, C. (2015). Analyzing gaze allocation during language planning: A cross-linguistic study on dynamic events. *Language and Cognition, 7*, 138–166. doi: 10.1017/langcog.2014.20

Gennari, S. P., Sloman, S. A., Malt, B. C., & Fitch, W. T. (2002). Motion events in language and cognition. *Cognition, 83*, 49–79. doi: 10.1016/S0010-0277(01)00166-4

Gentner, D., & Goldin-Meadow, S. (2003). *Language in mind: Mapping the nature of change in spatial cognitive development*. Oxford, England: Oxford University Press.

Gleitman, L., & Papafragou, A. (2013). Relations between language and thought. In D. Reisberg (Ed.), *Oxford handbook of cognitive psychology* (pp. 504–523). New York, NY: Oxford University Press.

Hespos, S. J., & Baillargeon, R. (2001). Reasoning about containment events in very young infants. *Cognition, 78*, 207–245. doi: 10.1111/1467-9280.00324

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20*(11), 1254–1259. doi: 10.1109/34.730558

Kawachi, K. (2007). Korean putting verbs do not categorize space contrastively in terms of "tightness of fit." *Lingua, 117*, 1801–1820.

Landau, B., Dessalegn, B., & Goldberg, A. M. (2010). Language and space: Momentary interactions. In V. Evands & P. Chilton (Eds.) *Language, cognition and space: The state of the art and new directions* (pp. 51–78). London, England: Equinox.

Levinson, S. C., Kita, S., Haun, D. B. M., & Rasch, B. H. (2002). Returning the table: Language affects spatial reasoning. *Cognition, 84*, 155–188. doi: 10.1016/S0010-0277(02)00045-8

Levinson, S. C., Meira, S., & The Language and Cognition Group (2003). Natural concepts in the spatial topological domain—Adpositional meanings in crosslinguistic perspective: An exercise in semantic typology. *Language, 79*, 485–516. doi: 10.1353/lan.2003.0174

Li, P., & Gleitman, L. (2002). Turning the tables: Language and spatial reasoning. *Cognition, 83*, 265–294. doi: 10.1111/1467-9280.00324

Lupyan, G. (2009). Extracommunicative functions of language: Verbal interference causes selective categorization impairments. *Psychonomic Bulletin & Review, 16*, 711–718. doi: 10.3758/PBR.16.4.71

Majid, A., Jordan, F., & Dunn, M. (2015). Semantic systems in closely related languages. *Language Sciences, 49*, 1–18. doi: 10.1016/j.langsci.2014.11.002

Meilinger, T., & Bülthoff, H. (2013). Verbal shadowing and visual interference in spatial memory. *PLOS ONE, 8*:e74177. doi: 10.1371/journal.pone.0074177

McDonough, L., Choi, S., & Mandler, J. (2003). Understanding spatial relations: Flexible infants, lexical adults. *Cognitive Psychology, 46*, 229–259. doi: 10.1016/S0010-0285(02)00514-5

Munnich, E., Landau, B., & Dosher, B. A. (2001). Spatial language and spatial representation: A cross-linguistic comparison. *Cognition, 81*, 171–207. doi: 10.1016/S0010-0285(02)00514-5

Papafragou, A., Hulbert, J., & Trueswell, J. (2008). Does language guide event perception? Evidence from eye movements. *Cognition, 108*, 155–184. doi: 10.1016/j.cognition.2008.02.007

Pederson, E., Danziger, E., Levinson, S. C., Kita, S., Senft, G., & Wilkins, D. (1998). Semantic typology and spatial conceptualization. *Language, 74*, 557–589. doi: 10.2307/417793

Roberson, D., & Davidoff, J. (2000). The categorical perception of colors and facial expressions: The effect of verbal interference. *Memory & Cognition, 28*, 977–986. doi: 10.3758/BF03209345

Soroli, E., Hickmann, M., Hendriks, H., Engemann, H., & Vincent, C. (2015). Language effects on spatial cognition? Cross-linguistic evidence and eye-tracking. *NINJAL International Symposium: Typology and Cognition in Motion Event Descriptions, Jan 2015*, Tokyo, Japan.

Talmy, L. (1978). Figure and ground in complex sentences. In J. Greenberg, C. Ferguson, & H. Moravcsik (Eds.), *Universals of hu-*

man language (pp. 627–649). Stanford, CA: Stanford University press.

Theeuwes, J. (1991). Exogenous and endogenous control of attention: The effect of visual onsets and offsets. *Perception & Psychophysics, 49*, 83–90. doi: 10.3758/BF03212219

Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception & Psychophysics, 51*, 599–606. doi: 10.3758/BF03211656

Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., & Kuipers, J. (2009). Unconscious effects of language-specific terminology on preattentive color perception. *Proceedings of the National Academy of Sciences of the United States of America, 106*, 4567–4570. doi: 10.1073_pnas.0811155106

Trueswell, J., & Papafragou, A. (2010). Perceiving and remembering events cross-linguistically: Evidence from dual-task paradigms. *Journal of Memory and Language, 63*, 64–82. doi: 10.1016/j.jml.2010.02.006

Van Bergen, G., & Flecken, M. (2017). Putting things in new places: Linguistic experience modulates the predictive power of placement verb semantics. *Journal of Memory and Language, 92*, 26–42. doi: 10.1016/j.jml.2016.05.003

Whorf, B. L. (1956). *Language, thought and reality: Selected writings of Benjamin Lee Whorf.* (Ed. J. B. Carroll). Cambridge, MA: MIT Press.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences of the United States of America, 104*, 7780–7785. doi: 10.1073/pnas.0701644104

Wolff, P., & Holmes, K. (2011). Linguistic relativity. *Wiley Interdisciplinary Review of Cognitive Science, 2*, 253–265. doi: 10.1002/wcs.104

Yun, H., & Choi, S. (under revision). *How languages differ in the semantic organization of spatial relations: A comparison between English and Korean.* Manuscript submitted for publication.

APPENDIX A

List of all stimuli used in the experiment.

fitness	relation	Figure(s)	Ground(s)	picture of the finished action
loose	in	Legos	container	
loose	in	knives	bowls	
loose	in	small deodorants	containers	
loose	in	mandarins	containers	
loose	in	pegs	bowls	
loose	in	candies	containers	
loose	in	shapes	container	
loose	in	wallets	containers	
loose	on	donuts	blocks	
loose	on	men	slabs	
loose	on	leaves	paper	
loose	on	playing cards	paper	
loose	on	bananas	plate	
loose	on	cups	saucers	
loose	on	stuffed animals	paper	
loose	on	shapes	blocks	
tight	in	bottles	holder	
tight	in	corks	bottles	
tight	in	creditcards	wallet	
tight	in	pegs	holder	
tight	in	candles	candleholders	
tight	in	small pegs	small holders	
tight	in	legomen	legocars	
tight	in	usb stick	usb port	
tight	on	donuts	poles	
tight	on	lids	coffeecups	
tight	on	clamps	pole	
tight	on	noppisteck	noppisteck	
tight	on	wheels	sticks	
tight	on	pencaps	pens	
tight	on	grips	grips	
tight	on	CDs	CD cases	