Cognitive Science 31 (2007) 705–719

Attention to Endpoints: A Cross-Linguistic Constraint on Spatial Meaning

Terry Regier, Mingyu Zheng

University of Chicago

Received 14 May 2005; received in revised form 15 October 2006; accepted 8 January 2007

Abstract

We investigate a possible universal constraint on spatial meaning. It has been proposed that people attend preferentially to the endpoints of spatial motion events, and that languages may therefore make finer semantic distinctions at event endpoints than at event beginnings. We test this proposal. In Experiment 1, we show that people discriminate the endpoints of spatial motion events more readily than they do event beginnings—suggesting a non-linguistic attentional bias toward endpoints. In Experiment 2, speakers of Arabic, Chinese, and English each described a set of spatial events displayed in video clips. Although the spatial systems of these languages differ, speakers of all three languages made finer semantic distinctions at event endpoints, compared to event beginnings. These findings are consistent with the proposal that event endpoints are privileged over event beginnings, in both language and perception.

Keywords: Spatial language; Semantic universals; Attention; Spatial motion; Arabic; Chinese; English; Language and thought

1. Introduction

Spatial language may shape spatial attention. Languages differ in their structuring of the spatial world (e.g., Choi & Bowerman, 1991; Talmy, 2000a), and these differences appear to sometimes have cognitive and perceptual consequences (Whorf, 1956). For instance, while many languages encode the difference between "to the left" and "to the right," Tzeltal does not—and Tzeltal speakers seem to pay less attention to left/right distinctions in the world around them, compared with speakers of languages that make this distinction (Levinson, 1996; Pederson, Danziger, Wilkins, Levinson, Kita, & Senft, 1998; see also Li & Gleitman, 2002 for a critique of this work, and Levinson, Kita, Haun, & Rasch, 2002 for a rejoinder). Although spatial language and spatial cognition often do not correspond directly (Crawford,

Correspondence should be addressed to Terry Regier, Department of Psychology, University of Chicago, 5848 S. University Ave. Chicago, IL 60637. E-mail: regier@uchicago.edu

Regier, & Huttenlocher, 2000; Munnich, Landau, & Dosher, 2001; Gennari, Sloman, Malt, & Fitch, 2002; Papafragou, Massey, & Gleitman, 2002), under some circumstances, habitual use of a language may draw its speakers' attention to particular aspects of the spatial world.

There may be an influence in the opposite direction as well: of attention shaping language. On this view, universal and relatively non-malleable aspects of attention may leave their mark on the world's languages, giving rise to universal tendencies in the semantics of space. For instance, it has been proposed that people attend more to the *endpoints*[1] of spatial motion events than to their beginnings—and that languages accordingly may make finer semantic distinctions at event endpoints than at beginnings. This has been suggested by a computational model of spatial term learning (Regier, 1996), on the following rationale. Once a person has seen a spatial event, the resulting final spatial configuration is more recent in memory than the rest of the event, and is thus more accessible and salient. Because of this posited greater mental salience of endpoints, fine-grained spatial semantic distinctions may be easier to pick up on when they appear at endpoints—and languages in general may therefore provide more semantic detail at endpoints than at comparatively under-attended event beginnings. For example, consider placing a picture on a wall, vs. laying it flat on a tabletop. The end results of these two joining actions are different—contact with a vertical vs. horizontal surface—and in German (unlike English) the two actions are described by different prepositions: "an" and "auf" respectively. However, German does not make this distinction at event *beginnings* for the analogous acts of separation—taking a picture off a wall vs. taking it off a tabletop: the preposition "von" is used in both cases. In this example, German makes a finer semantic distinction at event endpoints than at event beginnings. The prediction is that this pattern will tend to hold across languages.

Endpoints do seem to be privileged in some way. Three-year-old children find it easier to understand requests concerning where an object has gone *to*, than where it has come *from* (an "allative bias" in comprehension; Freeman, Sinha, & Stedmon, 1981). A similar bias appears in production: children with and without Williams syndrome, and normal adults, when describing simple events, tend to specify where an object is going to, more often than where it is coming from (Lakusta & Landau, 2005). The same pattern is found in the communicative gestures of deaf children in both Chinese- and English- speaking communities (Zheng & Goldin-Meadow, 2002). There is also some evidence of an endpoint effect on semantic breadth. Bowerman (1996:418) notes in passing that children learning English, Dutch, Korean, and Tzotzil all overgeneralize words of separation (e.g., "open," "off"), relative to the adult pattern, more than they do words of spatial joining (e.g., "close," "into"). One possible explanation for this pattern is that children attend preferentially to event endpoints—and can therefore make finer semantic distinctions there than at event beginnings. If this is indeed the reason that children overgeneralize words of separation more than analogous words for joining, we might expect the attentional bias toward endpoints to also cause the adult language itself to show broader terms of separating than of joining. To our knowledge, this hypothesis has not yet been directly tested. We do so here.

Experiment 1 tests whether there is a *non-linguistic* attentional bias favoring event endpoints, using a visual discrimination task. Experiment 2 tests whether there is a corresponding *linguistic* endpoint bias across three languages with differing spatial systems: Arabic, Chinese, and English. To preview our results, we find evidence for greater attention to endpoints both in perception, and in semantic breadth across all three languages. These findings are consistent with the proposal that attention to endpoints allows finer semantic detail at event endpoints. In the discussion, we also consider alternative interpretations of our findings.

2. Experiment 1a: Attention to endpoints in perception

Do people pay greater attention to the endpoints of spatial motion events than to their beginnings? To test this, we showed people pairs of motion events and asked whether the two events were the same or different. We predicted that people would be better at detecting differences when they occurred at the end, rather than the beginning, of the events.

In describing our experiments, we will speak of a *figure* object being spatially located relative to a *ground* object (e.g., Talmy, 2000a). For instance, in "The book is on the table," the book is the figure, and the table is the ground. We take the trajectory of a spatial motion event to consist of three parts: the *beginning* (or *source*: here, the spatial configuration of the two objects before the motion), the *path* taken by the figure while it is moving, and the *endpoint* (or *goal*[2]: here, the spatial configuration of the objects after the motion is completed). For instance, in "She put the book on the table," the beginning of the trajectory is left unspecified, the path leads toward the tabletop, and the endpoint is the resulting spatial configuration of the book lying on the table. We refer to such an event as a *joining* event (Choi & Bowerman, 1991; Bowerman, 1996), since it involves motion into a configuration in which the figure and ground are closely spatially joined. Taking the book off the table would be a *separating* event, since it involves motion out of such a configuration, separating the two objects.

Participants. 16 participants were recruited. All were University of Chicago students or affiliates, and were native speakers of U.S. English. They were paid for their participation.

Materials. A set of eight video clips was created, each portraying a hand manipulating objects on a tabletop. Four of these clips involved a small Tupperware container and its lid: (1) placing the lid *on* the container; (2) taking the lid *off* the container; (3) placing the lid *in* the container; (4) taking the lid *out of* the container. The other four video clips involved a toy bowl and a small shelf: (1) placing the bowl *on* the bottom shelf, (2) taking the bowl *off* the bottom shelf, (3) placing the bowl *on* the second shelf from the bottom, (4) taking the bowl *off* the second shelf from the bottom. Within each set of four, the two joining events (1 and 3) differed in the resulting endpoint spatial configuration (lid in versus on the container, or bowl on bottom shelf versus second shelf from bottom). Similarly, the two separating events in each set (2 and 4) differed in the *beginning* of the event (again, lid in versus on the container, or bowl on bottom shelf versus second shelf from bottom). At the beginning of all joining events, and at the end of all separating events, the figure object was on the tabletop, away from the ground object. Thus, discriminating two joining events required attention to endpoints, while discriminating two separating events required attention to event beginnings.

We controlled the video clips in the following ways. First, the duration of each video clip was the same. To make the task challenging, we sped up the clips to play at 6 times normal speed; after having been sped up, each clip was exactly 1 second long (30 frames), with 3 extra black frames attached at the beginning and the end of the clip. Second, the amount of time spent in the distinguishing spatial configuration (e.g., lid in versus on the Tupperware container) was the same—half a second, or 15 frames—for all clips. Thus, for joining events, the endpoint, or final resulting spatial configuration, was reached exactly halfway through the clip. For separating events, the initial spatial configuration lasted until exactly halfway through the clip. Thus, if these joining events are easier to discriminate than these separating events, it cannot be because of greater exposure to the distinguishing spatial configuration.

Procedure. Participants were shown pairs of video clips and were asked whether they were the same or different. In each pair, the two clips were either identical, or different but drawn from the same stimulus set (tupperware vs. shelf) and direction (joining vs. separating). For instance, if one of the clips showed the lid being placed *on* the tupperware container, the other clip would either show the same event, or the lid being placed *in* the container. This method of pairing ensures that to discriminate a pair of joining events requires attention to event endpoints, while to discriminate a pair of separating events requires attention to event beginnings. For each stimulus set (tupperware or shelf), there were 8 unique pairs of clips: each of the 4 clips, paired either with itself or with its alternate in that set. Each unique pair was repeated 5 times to create a total of 40 pairs of experimental stimuli for each set of stimuli (tupperware or shelf).

Half of the participants first viewed pairs of clips drawn from the Tupperware set, and later viewed pairs of clips from the shelf set, while for the other half of the participants this order was reversed. Within each set, the video clip pairs were presented in randomized order.

In each presentation, the two video clips were positioned vertically one above the other on a computer screen and were played simultaneously. After viewing each pair of video clips, the participant was asked whether the two events were the same or different, and we recorded whether the response was correct or not. The expectation was that people would discriminate joining events more accurately than separating events, because of greater attention to endpoints.

Treatment of the data. For each participant, for each stimulus set (Tupperware, shelf), we recorded how many errors (misses plus false alarms) that participant made in discriminating joining events in that stimulus set, and how many errors the same participant made in discriminating separating events in that stimulus set.

Results and discussion. We analyzed responses to the two stimulus sets separately. A paired t-test on the Tupperware data revealed that participants made fewer errors in discriminating joining events ($M = 5.813$) than separating events ($M = 9.125$), $t(15) = -5.199$, $p < 0.001$, two-tailed. An analogous paired t-test on the shelf data revealed the same pattern ($M = 4.438$ for joining events, $M = 7.75$ for separating events, $t(15) = -5.908$, $p < 0.001$, two-tailed)[3]. These results are displayed in Fig. 1. In general, participants appeared to discriminate joining events more accurately than separating events.

3. Experiment 1b: Attention to endpoints in perception, with reversed stimuli

To control for possible bias in the stimuli themselves, we re-ran Experiment 1a, but played all stimuli backwards, so that what was previously a separating action now appeared as a joining action and vice versa.

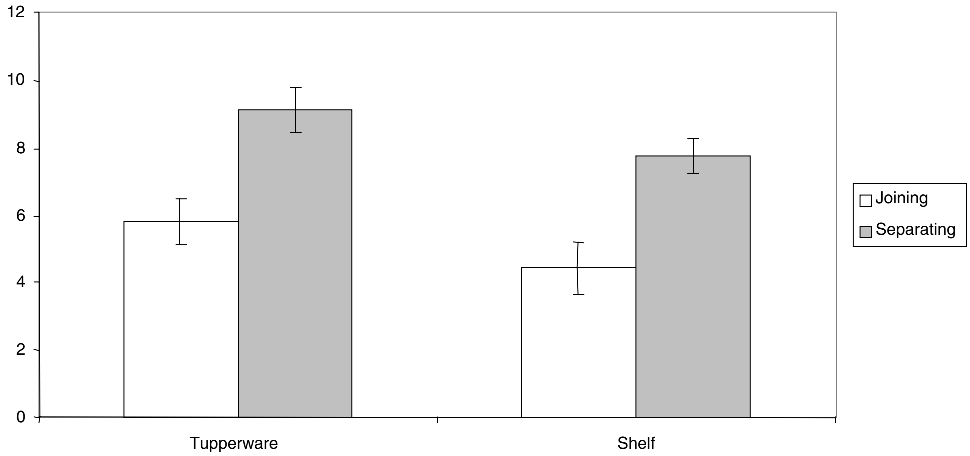

Fig. 1. Mean number of errors in discriminating joining events and separating events. Joining events are discriminated more accurately in both stimulus sets, suggesting greater attention to event endpoints than beginnings. Error bars show standard error of the mean.

Participants. 16 participants were recruited; they were different from those who participated in Experiment 1a. All were University of Chicago students or affiliates, and were native speakers of U.S. English. They were paid for their participation. Of these subjects, all performed both the Tupperware and shelf tasks, except for one who elected to participate in only the shelf task.

Materials, procedure, and treatment of the data. We used the same eight video clips as in Experiment 1a, but played them backwards. For instance, the video clip of a hand taking a lid from the tabletop and placing that lid on a Tupperware container, when played backwards, appeared to participants as a hand taking a lid *off* the container and placing it on the tabletop. The actions portrayed in the resulting backwards clips appeared natural, and generally comparable to those in the forward-playing clips. The construction of stimuli involving these clips, the experimental procedure, and the treatment of the data were all analogous to those in Experiment 1a.

Results and discussion. We obtained the same qualitative results: fewer errors in discriminating joining events (i.e., separating events played backwards) than separating events (i.e., joining events played backwards). For the Tupperware stimuli, $M(\text{joining}) = 5.133$; $M(\text{separating}) = 7.933$; $t(14) = -3.055$, $p < 0.01$, two-tailed, while for the shelf stimuli, $M(\text{joining}) = 3.688$; $M(\text{separating}) = 6.5$; $t(15) = -3.337$, $p < 0.005$[4]. Followup analyses showed that the mean number of errors did not differ significantly across Experiments 1a and 1b, in any of the four conditions (Tupperware/shelf $\times$ joining/separating).

These findings confirm that participants do seem to discriminate joining events more accurately than separating events, at least for the particular spatial actions considered here—further work would be required to probe the generality of this finding. Still, taken together, the results of Experiments 1a and 1b are consistent with our hypothesis that people attend preferentially to event endpoints, compared to event beginnings—since it is at the endpoints that joining events can be discriminated. Next we explore possible linguistic reflections of such an endpoint emphasis.

4. Experiment 2: Attention to endpoints in language

Are event endpoints emphasized in language as well as in perception? We tested this by showing a set of spatial joining and separating events to speakers of Lebanese colloquial Arabic, Mandarin Chinese, and U.S. English, and asking each speaker to describe each event in his or her native language[5]. If people attend more to endpoints than event beginnings, they should be able to make finer-grained semantic distinctions at endpoints than at beginnings. Therefore, we predicted that, across languages, participants' words for joining events would be semantically finer-grained than comparable words for separating events.

Languages differ in how the path of a motion event, including its beginning and endpoint, is expressed (Talmy, 2000b). In a *satellite-framed* language such as English, this path information is often carried in a verbal "satellite"—that is, an adjunct to the main verb such as a verb particle, or a preposition (e.g., she walked *into* the room, he ran *toward* the school), while the verb itself (here, "walk" or "run") primarily conveys manner, rather than path, of motion. In a *verb-framed* language, in contrast, the path is conveyed in the verb itself (e.g., Spanish, with verbs such as "entrar" = to go into). The languages treated here differ in this respect: Arabic is verb-framed, while Chinese and English are satellite-framed (Talmy, 2000b). There are also other differences among these three languages. For instance, Mandarin Chinese uses different spatial verbs to describe putting on clothing ("chuān") vs. putting on accessories such as jewelry ("dài").[6] Spatial verbs in Lebanese colloquial Arabic do not capture this distinction, but *do* capture a different one: putting clothes or accessories on oneself ("libis") vs. putting them on someone else ("labbas"—the causative form of "libis"). In contrast with both of these languages, U.S. English generally uses the verb "put" with the satellite "on" for all actions of putting either clothes or accessories on either oneself or someone else.[7] Thus, if these three languages all show greater semantic breadth in words of separation than in words of joining, that regularity will occur against a backdrop of cross-linguistic differences.

Participants. Nine native speakers of Lebanese colloquial Arabic, ten native speakers of Mandarin Chinese, and ten native speakers of U.S. English participated in the study. The Arabic speakers were students at the American University of Beirut, in Lebanon, and were recruited and tested there. The Mandarin speakers were students and scholars studying in the U.S., but whose first language was Mandarin, and who had lived in Mainland China for at least 20 years before coming to the U.S. The English speakers were University of Chicago students. All were paid for their participation.

Materials. Following Bowerman (1996), we chose as stimuli a set of joining events and their corresponding separating events. For example, since one of our joining events was a hand putting a hat on a doll's head, we also included the corresponding separating event of a hand taking the hat off the doll's head. All events involved small objects being manipulated on a tabletop; in this regard, the events were similar to those used as stimuli in Experiment 1, although a different set of events was used in this experiment. Specifically, 50 such events were used in this experiment: 25 joining events and the corresponding 25 separating events.

Table 1
Stimuli for Experiment 2. The table shows joining events, their corresponding separating events, and a code for each pair of events

	Joining motions	Separating motions	Code
1	Joining lego pieces (1 onto 2)	Separating lego pieces (1 off 2)	3lego
2	Putting a doll on a towel	Taking a doll off a towel	Towel
3	Joining lego pieces (1 onto 1)	Separating lego pieces (2 split in half)	2lego
4	Putting a watch on a wrist	Taking a watch off a wrist	Watch
5	Putting a bandaid on a hand	Taking a bandaid off a hand	Bandaid
6	Putting a bracelet on a wrist	Taking a bracelet off a wrist	Bracelet
7	Buttoning a jacket sleeve	Unbuttoning a jacket sleeve	Button
8	Putting a cap on a pen	Taking a cap off a pen	Cap
9	Putting a toy car in a box	Taking a toy car out of a box	Car
10	Putting a cassette in its case	Taking a cassette out of its case	Cassette
11	Putting a duck in a bucket of water	Taking a toy duck out of a bucket of water	Duck
12	Putting a hat on a doll's head	Taking a hat off a doll's head	Hat
13	Putting lego pieces in a bag	Taking lego pieces out of a bag	Legobag
14	Putting logs in a toy train car	Taking logs out of a toy train car	Log
15	Putting a nail in a toy block	Taking a nail out of a toy block	Nail
16	Putting a puzzle piece in a puzzle	Taking a puzzle piece out of a puzzle	Puzzle
17	Putting a ring on a pole	Taking a ring off a pole	Ring
18	Putting a rubber band around a book	Taking a rubber band off a book	Rubberband
19	Putting scarf around a doll's neck	Taking a scarf off a doll's neck	Scarf
20	Putting a shoe on a doll's foot	Taking a shoe off a doll's foot	Shoe
21	Putting a sock on a doll's foot	Taking a sock off a doll's foot	Sock
22	Joining 2 train cars together	Separating 2 train cars	2train
23	Joining train cars (1 onto 2)	Separating train cars (1 off 2)	3train
24	Joining pop beads (1 onto 3)	Separating pop beads (1 off 3)	4bead
25	Joining pop beads (3 onto 3)	Separating pop beads (6 split in half)	6bead

These are listed in Table 1, with a code for each pair of events. We created a video clip of each event, lasting in duration from 6.7 to 15.5 seconds—some of the events took longer to enact naturally than others.

Procedure. Each speaker of each of the three languages viewed each video clip. At the end of the video clip, the participant (speaker) was asked to write down, in his or her native language, a description of the event s/he had just seen. The order of presentation of the video clips was randomized for each participant.

Treatment of the data. For each spatial description, we extracted both the primary verb describing spatial motion and any spatial verbal satellite such as a particle or preposition. For instance, given the English description "taking the doll off the towel," we would extract the verb "take" and the satellite "off." The identification of verb (if any) and satellite (if any) was done by a native speaker of the language in question. We then noted for each word of separation or joining used by a particular speaker, how many different spatial configurations (from Table 1) the speaker used that word to label—this is the word's semantic breadth for that speaker. From this, we determined, for each speaker, the average semantic breadth (number

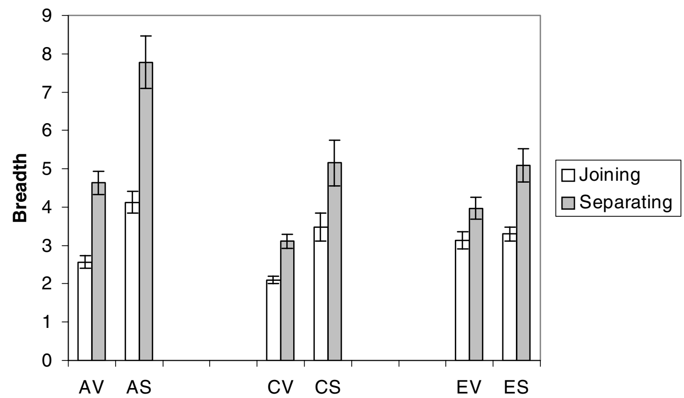

Fig. 2. Mean semantic breadth (number of events named) for joining and separating events as expressed by Arabic verbs (AV), Arabic satellites (AS), Chinese verbs (CV), Chinese satellites (CS), English verbs (EV), and English satellites (ES). Error bars show standard error of the mean.

of different spatial configurations labeled) for each of four different classes of words: verbs of joining, verbs of separation, satellites of joining, and satellites of separation.

Results and discussion. These data were submitted to three separate repeated-measures ANOVAs, one for each language, with Form (verb vs. satellite), and Direction (separating vs. joining) as factors[8].

For Arabic and English, there were significant interactions of Form and Direction (for Arabic: $F(1,8) = 6.9$, $p = 0.031$; for English, $F(1, 9) = 5.6$, $p = 0.043$), reflecting the fact that, in both languages, the difference in semantic breadth of separating and joining terms used was greater for satellites than for verbs. Given this significant interaction, the effect of Direction (separating vs. joining) was examined in separate paired t-tests for each type of form. Paired t-tests revealed that, consistent with our prediction, words for joining were applied to fewer events than were words for separating, for both verbs and satellites (for Arabic verbs, $M = 2.57$ for joining events, $M = 4.63$ for separating events, $t(8) = –7.9$, $p < 0.0001$; for Arabic satellites, $M = 4.13$ for joining events, $M = 7.78$ for separating events, $t(8) = –5.3$, $p = 0.0007$; for English verbs, $M = 3.14$ for joining events, $M = 3.97$ for separating events, $t(9) = –2.5$, $p = 0.0360$; for English satellites, $M = 3.30$ for joining events, $M = 5.10$ for separating events, $t(9) = –3.3$, $p = 0.0089$).

For Chinese, there was no significant interaction of Form and Direction ($F(1, 9) = 2.1$, $p = 0.18$). There was a significant main effect of Direction, reflecting the fact that verbs and satellites describing separating events were applied to a greater number of different events than were those describing joining events, consistent with our prediction ($F(1, 9) = 20.8$, $p = 0.001$). Fig. 2 displays the number of events referred to by verbs and satellites of each direction (joining and separating), for each of the three languages.

These findings show that across all 3 languages, terms of joining are semantically narrower than terms of separating, consistent with our prediction that languages will tend to make finer-grained semantic distinctions at event endpoints than at event beginnings. We also probed the same prediction in another way. Table 2 presents, for each spatial configuration code from Table 1, the Arabic words most often used to describe the joining event into that configuration,

Table 2
Lebanese colloquial Arabic words most often used to describe each of the joining and separating events. The spatial configuration code is from Table 1, VJ = verb of joining, VS = verb of separation, SJ = satellite of joining, SS = satellite of separation

Code	VJ	VS	SJ	SS
3lego	rakkab	fakk	3ala	–
Towel	HaTT	shaal	3ala	3an
2lego	rakkab	fakk	–	–
Watch	libis	shalaH	–	–
Bandaid	lazza’	shaal	3ala	3an
Bracelet	libis	shalaH	–	–
Button	bakkal	fakk	–	–
Cap	sakkar	fatah	–	–
Car	HaTT	shaal	bi	min
Cassette	HaTT	shaal	bi	min
Duck	kabb	shaal	bi	min
Hat	HaTT	shaal	3ala	3an
Legobag	HaTT	shaal	bi	min
Log	HaTT	shaal	bi	min
Nail	shakk	shaal	bi	min
Puzzle	rakkab	shaal	–	min
Ring	fawwat	shaal	bi	min
Rubberband	HaTT	shaal	bi	3an
Scarf	rabaT	fakk	–	–
Shoe	labbas	ShallaH	–	–
Sock	labbas	ShallaH	–	–
2train	rakkab	fakk	–	3an
3train	rakkab	fakk	–	3an
4bead	rakkab	fakk	–	–
6bead	rakkab	fakk	bi	–

and the separation event out of it. Tables 3 and 4 present the analogous information for Chinese and English, respectively. A visual comparison of Tables 2, 3, and 4 confirms that Arabic, Chinese, and English impose different categorization schemes on spatial motion events.

In any language, a given joining term J encodes motion into a range of different spatial configurations, and a given separating term S encodes motion *out of* a range of configurations. If the range of spatial configurations that S encodes motion out of is a subset of the range of configurations that J encodes motion into, we shall say that J *contains* S. Analogously, if the range of spatial configurations that J encodes motion into is a subset of the range of configurations that S encodes motion out of, we shall say that S contains J. For example, as can be seen in Table 4, the range for the English verb of joining “connect” comprises the 3 configurations coded as *2train, 3train*, and *6bead*—that is, “connect” was the modal English

Table 3
Mandarin Chinese words most often used to describe each of the joining and separating events. The spatial configuration code is from Table 1, VJ = verb of joining, VS = verb of separation, SJ = satellite of joining, SS = satellite of separation

Code	VJ	VS	SJ	SS
3lego	fàng	chāi	qǐlái	kāi
Towel	fàng	ná	zài..shàng	–
2lego	fàng	fēn	zàiyīqǐ	kāi
Watch	dài	tuō	–	xià
Bandaid	tiē	sī	zài..shàng	xiàlái
Bracelet	dài	qǔ	–	xiàlái
Button	kòu	jiě	shàng	kāi
Cap	tào	ná	shàng	xià
Car	fàng	ná	dào..lǐ	chū
Cassette	fàng	qǔ	rù	chū
Duck	tuī	ná	rù	chū
Hat	dài	qǔ	shàng	xià
Legobag	fàng	ná	jìn	chū
Log	zhuāng	ná	–	xià
Nail	chā	bá	–	chū
Puzzle	fàng	ná	–	–
Ring	tào	ná	zài..shàng	xià
Rubberband	tào	qǔ	zài..shàng	xiàlái
Scarf	dài	jiě	zài..shàng	xià
Shoe	chuān	tuō	–	xià
Sock	chuān	tuō	shàng	–
2train	lián	fēn	–	kāi
3train	lián	fēn	zàiyīqǐ	kāi
4bead	jiē	ná	shàng	kāi
6bead	lián	fēn	qǐlái	kāi

verb used when describing motion into each of these 3 configurations, and no others. At the same time, the range for the English verb of separation "pull" was *2train* and *3train*. Since the range for "pull" (*2train, 3train*) is a subset of the range for "connect" (*2train, 3train*, *6bead*), we say that "connect" *contains* "pull."

Table 5 presents, for each of Arabic, Chinese, and English, those terms of joining that contain two or more terms of separation, and those terms of separation that contain two or more terms of joining.

In both Arabic and Chinese, there are words of separating that contain multiple words of joining but no words of joining that contain multiple words of separating; this is consistent with our prediction that languages will tend to make finer semantic distinctions at event endpoints than at event beginnings. The English data in this analysis are more equivocal, with one example in each direction.

Table 4
English words most often used to describe each of the joining and separating events. The spatial configuration code is from Table 1, VJ = verb of joining, VS = verb of separation, SJ = satellite of joining, SS = satellite of separation

Code	VJ	VS	SJ	SS
3lego	place	remove	–	–
Towel	place	take	on	off
2lego	put	take	together	apart
Watch	put	remove	on	from
Bandaid	put	remove	on	from
Bracelet	put	take	on	off
Button	button	unbutton	–	–
Cap	place	take	on	off
Car	put	take	in	from
Cassette	put	remove	into	out of
Duck	push	–	into	out of
Hat	place	remove	on	from
Legobag	put	take	into	out of
Log	place	remove	on	from
Nail	push	remove	into	from
Puzzle	put	remove	into	from
Ring	place	take	on	off
Rubberband	put	take	around	from
Scarf	tie	remove	around	from
Shoe	put	take	on	from
Sock	put	remove	on	from
2train	connect	pull	–	–
3train	connect	pull	together	–
4bead	put	remove	to	–
6bead	connect	separate	together	–

In sum, in all three languages, words of joining tend to be semantically narrower than comparable words of separating (Fig. 2), in line with our prediction. Moreover, in two of the three languages, words of joining pick out extensional subsets of words of separation.

5. Discussion

Experiment 1 showed that people visually discriminate event endpoints more accurately than they do event beginnings—suggesting a non-linguistic bias favoring endpoints. Experiment 2 demonstrated an analogous endpoint bias in language: speakers of Arabic, Chinese, and English made finer semantic distinctions at event endpoints than at event beginnings, in line with our predictions.

Table 5
Containment relations among words of motion in Arabic, Chinese, and English. The numbers in parentheses show the number of events for which each word was the modal choice (see Tables 2, 3, 4)

Arabic verbs of joining that contain two or more verbs of separating
None.
Arabic verbs of separating that contain two or more verbs of joining
fakk (8) contains *bakkal* (1), *rabaT* (1).
shaal (12) contains *HaTT* (7), *fawwat* (1), *kabb* (1), *lazza'* (1), *shakk* (1).
Arabic satellites of joining that contain two or more satellites of separating
None.
Arabic satellites of separating that contain two or more satellites of joining
None.
Chinese verbs of joining that contain two or more verbs of separating
None.
Chinese verbs of separating that contain two or more verbs of joining
ná (9) contains *jiē* (1), *tuī* (1), *zhuāng* (1).
Chinese satellites of joining that contain two or more satellites of separating
None.
Chinese satellites of separating that contain two or more satellites of joining
chū (5) contains *dào..lǐ* (1), *jìn* (1), *rù* (2).
kāi (7) contains *qǐlái* (2), *zàiyīqǐ* (2).
English verbs of joining that contain two or more verbs of separating
connect (3) contains *pull* (2), *separate* (1).
English verbs of separating that contain two or more verbs of joining
None.
English satellites of joining that contain two or more satellites of separating
None.
English satellites of separating that contain two or more satellites of joining
from (11) contains *around* (2), *in* (1).

These findings are consistent with the proposal that attention to endpoints yields a possible universal tendency in spatial language, such that across languages, endpoints tend to be more finely semantically differentiated than event beginnings (Regier, 1996). Our findings add to a growing body of work that grounds spatial language in apparently universal aspects of spatial perception and cognition (Landau & Jackendoff, 1993; Hayward & Tarr, 1995; Crawford et al., 2000; Talmy, 2000b, 2000c; Munnich et al., 2001; Regier & Carlson, 2001).

However, there are several respects in which our present findings fall well short of confirming the proposal of an attentionally-created semantic universal. First, we have examined only three languages, leaving us with little basis for claims of universality—research on other languages would be needed to more completely test the idea. Second, even if the cross-linguistic pattern proves to hold across many languages, it would still not be clear that it was *caused* by the perceptual bias we have explored here. We have demonstrated a correlation between perceptual and linguistic endpoint emphasis—but not causation. Finally, even if there is a causal link of some sort between the perceptual and linguistic biases we find, the *nature* of the perceptual bias—and thus the ultimate source of the possible linguistic universal—remains unclear. There are several possible sources for such a bias. As originally posited, it could simply be a matter of memory recency (Waugh & Norman, 1965; Shiffrin, 1973): since endpoints are more recent than other parts of an event, they may tend to be clearer, and thus more accessible, in working memory.

A rather different possibility is that the perceptual endpoint bias could result from attention to the intended *goals* of a human actor, a tendency found even in infants (e.g., Woodward, Sommerville, & Guajardo, 2001). Such attention to goals could account for our findings, for two reasons. First, goals become fully apparent at event endpoints, so a goal bias would highlight the endpoints of human actions. And second, all of our video clips were of human hands manipulating objects; that is, they depicted not merely events, but purposeful human actions which would presumably engage viewers' tendency to attend to goals. If this is the source of the perceptual endpoint bias, one would not expect to find such a bias for events that do *not* concern purposeful human actions. Some recent work is consistent with this idea: Lakusta (2005) finds a non-linguistic attentional preference for event endpoints, much as we do in Experiment 1—and also finds that this preference is weakened when the events contain only inanimate objects. But then what of the linguistic bias? This has not yet been tested using only inanimate objects. It is conceivable that the linguistic tendency would also not generalize to events other than deliberate human actions. However, it is also possible that the linguistic bias is grounded in goal-tracking, but *does* also generalize to other sorts of actions, since language is a system which must apply to both human actions and other events, and since human actions are a particularly salient and important sort of event.

These questions are left open by our present findings. Still, our findings do identify a perceptual asymmetry in event perception, and an analogous cross-linguistic regularity in spatial meaning. They are thus at least consistent with the idea that spatial attention may leave its imprint on aspects of human language.

Notes

1. By the "endpoint" of a motion event we mean the final state that immediately results from that action. For instance, the endpoint of putting a ball in a box would be the resulting state of the ball being in the box. By "attention" we mean any mental process that highlights some features of the world at the expense of others.
2. In the general discussion we consider the role of goal-directed thinking in spatial language, and whether it may account for our findings.
3. For the Tupperware stimuli, the mean number of hits was 8.563 for joining and 7.813 for separating, and the mean number of false alarms was 4.375 for joining and 6.938 for separating. For the shelf stimuli, the mean number of hits was 8.938 for joining and

8.125 for separating, and the mean number of false alarms was 3.375 for joining and 5.875 for separating.

4. For the Tupperware stimuli, the mean number of hits was 8.133 for joining and 7.333 for separating, and the mean number of false alarms was 3.267 for joining and 5.267 for separating. For the shelf stimuli, the mean number of hits was 8.75 for joining and 7.875 for separating, and the mean number of false alarms was 2.438 for joining and 4.375 for separating.
5. These three languages were chosen as a sample of convenience: members of our research group are familiar with these languages, and we had access to speakers of all three.
6. The diacritic mark over the vowel in the transliteration of Chinese words denotes tone: either flat (ā), rising (á), falling-then-rising (ǎ), or falling (à).
7. English can express the on-oneself/on-someone-else distinction, but through word order ("put on a shirt" vs. "put a shirt on someone else") rather than through choice of verb, as in Arabic. In Chinese, the same distinction is expressed by the presence or absence of a separate benefactive marker "gěi," rather than choice of verb.
8. We report subject analyses here. Accompanying item analyses are not required since items were matched across conditions (Raaijmakers, Schrijnemakers, & Gremmen, 1999): for every separating clip there was an analogous joining clip, and vice versa.

Acknowledgments

Thanks to Susanne Gahl for help with the analyses, to Muhammad Ali Khalidi for helping to arrange the collection of Arabic data, to Missan Laycy-Stouhi and Rebecca Keedi for help in coding the Arabic data, and to Andrew Kowalczuk for programming. This paper has benefited from the helpful suggestions of Dedre Gentner, Eric Pederson, and Phillip Wolff. Any remaining errors are ours. This work was supported by grant DC03384 from the National Institutes of Health.

References

Bowerman, M. (1996). Learning how to structure space for language: A crosslinguistic perspective. In P. Bloom, M. Peterson, M. Garrett, & L. Nadel (Eds.), *Language and space* (pp. 385–436). Cambridge, MA: The MIT Press.

Choi, S., & Bowerman, M. (1991). Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. *Cognition, 41*(1–3), 83–121.

Crawford, L., Regier, T., & Huttenlocher, J. (2000). Linguistic and non-linguistic spatial categorization. *Cognition, 75*(3), 209–235.

Freeman, N. H., Sinha, C. G., & Stedmon, J. A. (1981). The allative bias in 3-year-olds is almost proof against task naturalness. *Journal of Child Language, 8*(2), 283–296.

Gennari, S. P., Sloman, S. A., Malt, B. C., & Fitch, W. T. (2002). Motion events in language and cognition. *Cognition, 83*(1), 49–79.

Hare, B., Brown, M., Williamson, C., & Tomasello, M. (2002). The domestication of social cognition in dogs. *Science, 298*, 1634–1636.

Hayward, W., & Tarr, M. (1995). Spatial language and spatial representation. *Cognition, 55*, 39–84.

Lakusta, L. (2005). *Source and goal asymmetry in non-linguistic motion event representations.* Unpublished PhD dissertation, Johns Hopkins University.

Lakusta, L., & Landau, B. (2005). Starting at the end: The importance of goals in spatial language. *Cognition, 96*(1), 1–33.

Landau, B., & Jackendoff, R. (1993). "What" and "where" in spatial language and spatial cognition. *Behavioral and Brain Sciences, 16*, 217–265.

Markson, L., & Bloom, P. (1997). Evidence against a dedicated system for word-learning in children. *Nature, 385*, 813–815.

Medin, D., & Schwanenflugel, P. (1981). Linear separability in classification learning. *Journal of Experimental Psychology: Human Learning and Memory, 7*, 355–368.

Munnich, E., Landau, B., & Dosher, B. A. (2001). Spatial language and spatial representation: A cross-linguistic comparison. *Cognition, 81*(3), 171–207.

Nicholich, L. M. (1981). The cognitive bases of relational words in the single-word period. *Journal of Child Language, 8*, 15–34.

Papafragou, A., Massey, C., & Gleitman, L. (2002). Shake, rattle, 'n' roll: The representation of motion in language and cognition. *Cognition, 84*(2), 189–219.

Raaijmakers, J. G. W., Schrijnemakers, J. M. C., & Gremmen, F. (1999). How to deal with the "Language-as-fixed-effect fallacy": Common misconceptions and alternative solutions. *Journal of Memory and Language, 41,* 416–426.

Regier, T. (1996). *The human semantic potential: Spatial language and constrained connectionism*. Cambridge, MA: MIT Press.

Regier, T., & Carlson, L. A. (2001). Grounding spatial language in perception: An empirical and computational investigation. *Journal of Experimental Psychology: General, 130*(2), 273–298.

Shiffrin, R. M. (1973). Information persistence in short-term memory. *Journal of Experimental Psychology, 100*(1), 39–49.

Talmy, L. (2000a). How language structures space. *Toward a cognitive semantics, Vol. 1: Concept structuring systems* (pp. 177–254). Cambridge, MA: MIT Press.

Talmy, L. (2000b). A typology of event integration. *Toward a cognitive semantics, Vol. 2: Typology and process in concept structuring* (pp. 213–288). Cambridge, MA: MIT Press.

Talmy, L. (2000c). The Windowing of Attention in Language. *Toward a cognitive semantics, Vol. 1: Concept structuring systems* (pp. 257–309). Cambridge, MA: MIT Press.

Waugh, N. C., & Norman, D. A. (1965). Primary memory. *Psychological Review, 72*(2), 89–104.

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

Woodward, A., Sommerville, J., & Guajardo, J. (2001). How infants make sense of intentional action. In B. Malle, & L. Moses (Eds.), *Intentions and intentionality: Foundations of social cognition* (pp. 149–169). Cambridge, MA: MIT Press.

Zheng, M. Y., & Goldin-Meadow, S. (2002). Thought before language: How deaf and hearing children express motion events across cultures. *Cognition, 85*(2), 145–175.