Spatial Terms

Chapter · November 2018
DOI: 10.1093/oxfordhb/9780198791768.013.18

CITATIONS
0

READS
770

2 authors:

Myrto Grigoroglou
University of Toronto
11 PUBLICATIONS 131 CITATIONS

See Profile

Anna Papafragou
University of Delaware
138 PUBLICATIONS 5,022 CITATIONS

See Profile

All content following this page was uploaded by Myrto Grigoroglou on 04 November 2018.
The user has requested enhancement of the downloaded file.
Across the world’s languages, spatial terms are organized around a set of basic, non-linguistic spatial notions. Nevertheless, there is also considerable cross-linguistic variation in terms of both the kinds of linguistic devices used to express spatial relations and the way these devices carve up the semantic domain of space. This chapter reviews literature on spatial terms cross-linguistically focusing on three main subdivisions of the spatial domain: location (i.e., the static position of an object in space), motion (i.e., the dynamic displacement of an object in space), and frames of reference (i.e., abstract spatial-coordinate axes imposed on spatial configurations). The intricate relation between spatial language and non-linguistic spatial cognition is discussed throughout the chapter.

**Keywords:** space, location, motion, frames of reference, language and cognition, universality, language acquisition

**Introduction**

The representation of space is a fundamental cognitive ability and all human languages can and do represent space. There is good evidence that spatial language is organized using a set
of basic principles that include a shared—potentially universal—set of non-linguistic spatial
distinctions (Bowerman, 1996; Mandler, 1992; Miller & Johnson-Laird, 1976; Landau &
Jackendoff, 1993). Nevertheless, careful examination of the means different languages use to
encode space has revealed considerable cross-linguistic divergence (Choi & Bowerman, 1991;
Levinson, 1996, 2003; Levinson & Wilkins, 2006). There is currently a wealth of experimental
evidence on how both shared and language-specific factors conspire to shape the nature of spatial
language and the way spatial terms are acquired and processed. In this chapter, we provide a
selective review of this large literature focusing on three main subdivisions of the spatial domain:
location (i.e., the static position of an object in space), motion (i.e., the dynamic displacement of
an object in space), and frames of reference (i.e., abstract spatial-coordinate axes imposed on
spatial configurations). Towards the end of the chapter we consider the possibility that spatial
language itself could affect the non-linguistic representations of spatial categories.

Location terms

Languages analyze the location of an object in terms of three elements: the object to be
located (figure), the reference object (or ground) and the relationship between the two (e.g.,
containment, as in English in, or support, as in English on; Landau & Jackendoff, 1993; Talmy,
1983).

It is widely recognized that, in several respects, the linguistic encoding of location
reflects a set of shared, pre-existing conceptual notions that constrain both the nature and the
acquisition of spatial vocabulary across languages. Several sources of evidence support this
position. First, there are many similarities in the way the cross-linguistic encoding of location is
organized (Landau & Jackendoff, 1993; Miller & Johnson-Laird, 1976). For instance, there are
principles of figure-ground assignment that characterize all human languages and probably originate with non-linguistic principles of spatial organization (Landau & Jackendoff, 1993; Talmy, 1983). As an example, typically, the smaller, more mobile object in a configuration is treated as the figure and the larger, more stable object as the ground (e.g., *The laptop is on the desk*); reversing this expectation makes a sentence sound odd (e.g., *The desk is under the laptop*).

Second, infants, during their first year of life, already know a lot about the spatial properties of objects in the physical world. Studies using preferential looking time paradigms show that, at 2.5 months, children can reason about containment (e.g., Hespos & Baillargeon, 2001) and, at 3 to 4 months, they can already form a basic representation for the relations ‘above’ and ‘below’ (Quinn, 1994, 2004; Quinn, Cummins, Kase, Martin, & Weissman, 1996). At around 6 months, infants can distinguish between containment, support and occlusion (e.g., Aguiar & Baillargeon, 1998; Baillargeon, Needham, & Devos, 1992; Casasola, 2008; Casasola & Cohen, 2002; Casasola, Cohen, & Chiarello, 2003) and, at 9 to 10 months, they can form a category for the relation ‘between’ (Quinn, Adams, Kennedy, Shettler, & Wasnik, 2003). Preverbal infants can distinguish between relations that their native language does not encode. For instance, infants growing up in an English-speaking community can distinguish between tight-fit and loose-fit containment and support relations, although their native language does not systematically encode this distinction, while other languages do (Casasola & Cohen, 2002; Casasola et al., 2003; Hespos & Spelke, 2004; McDonough, Choi & Mandler, 2003).

Third, and relatedly, studies directly comparing linguistic and non-linguistic understanding of static spatial relations in slightly older children have found that non-linguistic understanding precedes the acquisition of spatial terms. For instance, E. Clark (1973a)
demonstrated that children understood the notions of containment and support when playing with objects earlier than the age at which they fully acquired the meanings of the prepositions in and on. Levine and Carey (1982) reported similar results with the axial terms front and back. Such findings suggest that concepts of location precede (and presumably structure) the acquisition of locative terms in language.

Fourth, children acquire locative terms in a consistent order cross-linguistically (e.g., Ames & Learnerd, 1948; Brown, 1973; E. Clark, 1977, 1980; Grimm, 1975; Johnston, 1984; Johnston & Slobin, 1979; Parisi & Antinucci, 1970; Weissenborn, 1981). In an influential study, Johnston and Slobin (1979) found that children across different languages produced spatial adpositions close in meaning to the English terms in, on, under and beside earlier than the prepositions between, in front of, and behind. It was proposed that this cross-linguistically robust timetable reflected the order in which children develop the corresponding non-linguistic spatial notions: in, on, beside and their synonyms rely on simple topological concepts such as containment, support, and proximity (see Piaget & Inhelder, 1967; but see Coventry & Garrod, 2004 for a more complex picture). Other terms such as between concern the relation between three objects and may, thus, be more complex. Similarly, axial terms such as in front of and behind rely on spatial co-ordinate systems and involve complex computations of figure-ground relations (see section on Frames of Reference terms below; cf. Rabagliati & Shrinivasan, this volume, for a discussion of the relation between words and concepts in non-spatial domains).

Despite these commonalities, languages differ greatly in the ways they express locative information. One difference concerns the formal devices used to mark locative meaning. In English and many other languages, locative information is encoded in adpositions (prepositions or postpositions). Other languages lack adpositions (e.g., the Australian languages Jaminjung
and Arrente) and encode figure-ground relations through locative case marking on the ground Noun Phrase (and an optional positional case on the verb). Yet other languages (e.g., the Mayan languages Tzeltal and Yukatek) have only a limited set of general adpositions and package locative information into a rich inventory of spatial verbs (Levinson & Wilkins, 2006).

More importantly, languages differ in the way they carve up the semantic space of location. In a series of studies, Melissa Bowerman and her colleagues have documented such differences in the domains of containment and support or attachment (Bowerman, 1996; Bowerman & Choi 2001; Bowerman, de Leon & Choi, 1995; Gentner & Bowerman, 2009; see also Levinson & Wilkins, 2006). For example, in English, the preposition in is used for containment (e.g., apple in bowl) and the preposition on is used for a series of support relations: (a) “support from below” (e.g., mug on table), (b) “clingy attachment” (e.g., stamp on envelope), (c) “hanging against” (e.g., poster on wall), (d) “point to point attachment” (e.g., apple on branch), (e) “encirclement with contact” (e.g., ribbon on candle). In Dutch, as in English, a single preposition (in) is used for containment scenes (apple in bowl), but the English on space is partitioned into three prepositions: op used for support-from-below and clingy attachment (a-b above), aan used for hanging support (c-d above) and om used for encirclement (e above). In Spanish, a single preposition (en) is used to describe all the above relations. And in Korean, the degree-of-fit between figure and ground is marked in a way that cross-cuts containment and support: Korean speakers use the verb kkita for tight-fit containment and support relations (e.g., earplug in ear, top on pen) and the verb nehta for loose-fit containment and encirclement (e.g., ball in box, loose ring on pole; Choi & Bowerman, 1991; but see Kawachi, 2007).

These cross-linguistic differences play an important role in the acquisition of locative terms. English-speaking children learn to encode support (via on) earlier than their Dutch-speaking
peers (who have to learn a more complex, three-term system); by contrast, learners in both language groups acquire containment expressions (in) around the same time (Gentner & Bowerman, 2009). Furthermore, by age 2, children already adopt language-specific locative encoding patterns, with English learners organizing spatial meanings around the containment/support distinction and Korean learners organizing spatial meanings around the tight/loose fit distinction (Choi & Bowerman, 1991; see also Bowerman, 1996). Recently, a more systematic comparison of the ways that containment and support are described by children and adults cross-linguistically suggests that one has to look at detailed semantic profiles within each of these relations to capture the intricacies of spatial language and its acquisition (Landau, Johannes, Skordos, & Papafragou, 2016). This work reveals a principled but highly complex interplay of shared and language-specific contributions to how spatial language is used and learned.

**Motion terms**

Languages analyze motion events as the displacement of a moving entity (figure) in relation to a reference object (ground), along a trajectory (path), and in a specific manner (Talmy, 1985). For example, in English the sentence *The cat jumped from the couch into the basket* includes a figure (*the cat*), a manner (*jumped*), and two path expressions specifying the source (*from NP*) and the goal or endpoint of the path (*into NP*), each with respect to a specific ground (*the couch* for the source path and *the basket* for the goal path).

As with locative terms, there are good reasons to assume that linguistic-motion primitives correspond to pre-linguistic, probably universal, conceptual-motion primitives that shape motion vocabulary across languages. First, some basic motion concepts are available early on; infants in
the first year of life detect changes in the path and manner of motion events (Pulverman, Song, Hirsh-Pasek, Pruden, & Golinkoff, 2013) and are able to form a category for the invariant path or manner of motion across different exemplars of manner or path (Pruden, Hirsh-Pasek, Maguire, Meyers & Golinkoff, 2004). Interestingly, infants form path and manner categories independently of the encoding preferences of the linguistic environment in which they are growing up (Pulverman, Golinkoff, Hirsh-Pasek, & Buresh, 2008).

Second, there are homologies between the way motion terms are used and acquired and the way humans process motion non-linguistically. A case in point is a well-documented asymmetry between goal and source paths. In language, both children and adults tend to mention goal path expressions (e.g., into the basket) more often than source path expressions (e.g., from the couch) when describing motion events (Lakusta & Landau, 2005, 2012; Landau & Zukowski, 2003; Papafragou, 2010; Regier & Zheng, 2007). The goal-source asymmetry has also been documented in the speech of brain-damaged patients (Ihara & Fujita, 2000), and children with Williams syndrome, a rare genetic deficit that causes spatial impairment (Landau & Zukowski, 2003), as well as in the spontaneous gestures of congenitally deaf children who have never been exposed to conventional language (Zheng & Goldin-Meadow, 2002). Furthermore, across languages, goals are encoded with greater specificity than sources (Johanson & Papafragou, 2010; Regier & Zheng, 2007) and this asymmetry affects the way both child and adult learners generalize novel motion expressions (Papafragou, 2010). The linguistic source/goal asymmetry has its roots in non-linguistic motion cognition. Both children and adults are better at detecting changes of landmarks or spatial configurations in goal compared to source paths (Papafragou, 2010; Regier & Zheng, 2007). Furthermore, this non-linguistic source/goal asymmetry is already present in 12-month-old infants (Lakusta, Wagner, O’Hearn & Landau,

Despite being rooted in a shared conceptual typology, the linguistic encoding of motion is characterized by intense typological variability. Both the ways motion primitives are lexicalized in spatial vocabularies and the ways these primitives are conflated into sentential structure vary considerably cross-linguistically. For instance, some languages (e.g., Romance, Japanese, Greek, Turkish) tend to encode the path of motion in the main verb (e.g., in French *entrer* ‘enter’, *sortir* ‘exit’, *descendre* ‘descend’) and the manner of motion (optionally) in an additional clause or gerund (e.g., *en courant* ‘running’); by contrast, other languages (e.g., English, German, Russian, Chinese) package manner information in the main verb and path information in particles or prepositions (e.g., *up, into*; Talmy, 1985; see also Chen & Husband, this volume). Several studies have confirmed these cross-linguistic preferences in motion encoding in both adults and children (e.g., Allen et al., 2007; Berman & Slobin, 1994; Hickmann, 2006; Naigles, Eisenberg, Kako, Higher, & McGraw, 1998; Papafragou, Massey, & Gleitman, 2002, 2006; Slobin, 1996, 2003 a.o.). For instance, Papafragou et al. (2002) found that English-speaking 4- to 12-year-old children and adults used primarily manner verbs to describe motion scenes (e.g., *The frog is jumping into the room*), while Greek-speaking participants used primarily path verbs (e.g., *O vatraxos beni sto domatio* ‘The frog is entering the room’).

These language-specific preferences for encoding motion information affect how newly encountered motion terms are interpreted. In one study, adult speakers of English and Spanish watched simple motion events (e.g., a woman skipping towards a tree) and heard a novel motion verb describing the event. Spanish-speaking adults interpreted the novel verb as a path verb, while English-speaking adults as a manner verb, thus following the motion lexicalization
preferences of their language (Naigles & Terrazas, 1998). A similar bias towards language-specific interpretations has also been documented in children from different linguistic backgrounds, at least from age 3 (e.g., Maguire et al., 2010; Papafragou & Selimis, 2010; Skordos & Papafragou, 2014; cf. Hohenstein, Naigles, & Eisenberg, 2004).

**Frames of Reference terms**

Frames of Reference (FoRs) are abstract co-ordinate systems for locating a figure object in space in relation to the axes defined by or imposed onto a reference (ground) object. Languages distinguish three FoRs: the *intrinsic*, the *relative* and the *absolute* (Brown & Levinson, 1993; Levinson, 1996, 2003; Pederson et al., 1998; see also Shusterman and Li, 2016a, for a somewhat different terminology). The *intrinsic* FoR describes the location of a figure object in terms of the inherent properties (e.g., front/back, top/bottom) of the ground object, often one’s own body (e.g., *The tree is in front of the house/me*). The *relative* FoR describes the location of a figure with respect to a ground object that lacks inherent sides (e.g., ball, tree, bottle) in terms of the speaker’s or some other observer’s viewpoint (e.g., *The ball is to the right of the table*). The *absolute* FoR describes the location of the figure with respect to environment-based co-ordinates such as cardinal directions, the solar compass, wind directions, mountain slopes etc. (e.g., *The forest is to the north of the village*).

Different co-ordinate systems for locating objects are available in pre-linguistic infants (see Newcombe & Huttenlocher, 2000; Quinn, 2004). Early studies on spatial orientation showed that infants in the first year of life use their own bodily co-ordinates to code object location and become sensitive to environment-based co-ordinates such as landmarks only later (Acredolo, 1978; Bremner & Bryant, 1977; Rieser, 1979). However, subsequent work
demonstrated that both co-ordinate systems are available early on and that the choice of system is context-dependent (Acredolo, 1979, 1982; Acredolo & Evans, 1980; Bremner, 1978). Different types of co-ordinate systems are also available to non-human species (see Gallistel, 1990; Gallistel & Cramer, 1996, for reviews).

Cross-linguistically, there is considerable variation in the availability or frequency of use of different FoRs. All languages have terms to describe the intrinsic FoR (even in rudimentary form; Levinson & Wilkins, 2006) but the distribution of the other two FoRs differs. English and Dutch make use of both the relative and the absolute FoR but prefer the relative FoR for small-scale arrays (e.g., *The ball is behind the table*). Tzeltal and Arrente mostly make use of the absolute FoR, even for small-scale arrays (e.g., *The ball is to the north of the table*), and lack relative terms altogether (Levinson & Wilkins, 2006).

Even within the set of languages that share a frame of reference, there are differences in how FoRs work. For intrinsic FoRs, languages use different (and often fairly complex) criteria in assigning names to a reference object’s facets (Levinson & Wilkins, 2006). For instance, in English, the “front” of an object is defined by canonical encounter (for people or animals), forward motion (for vehicles), functional orientation (for appliances), etc. (see H. Clark, 1973; Harris & Strommen, 1979; Miller & Johnson-Laird, 1976). However, in Tzeltal, object-part name assignment is completely dependent on the object’s internal geometry: for example, a stone lying down with a flat surface on the ground has its “face” (i.e., the least flat side) upside down (Levinson, 1994; Levinson & Wilkins, 2006). For relative FoRs, one source of cross-linguistic variation is how the viewpoint of the observer is projected onto the ground object. In English, the sentence *The ball is in front of the table* typically means that the ball is located between the observer and the table. Thus, the table has acquired a “front” by the observer through reflection,
as if the table were “facing” the observer. However, in Hausa, the same sentence means that the ball is in the region projected from the furthermost side of the table with respect to the observer (a position which would have been described by the term behind in English). Thus, in the Hausa relative system, the table’s “front” faces the same direction as the observer (Hill, 1982; Levinson & Wilkins, 2006). Finally, absolute FoRs in the world’s languages are extremely diverse.

Arrente has a fully abstract cardinal direction system (e.g., north, south, etc.), Tzeltal uses the terms “uphill” (south) vs. “downhill” (north), Yélî Dnye distinguishes between “up” (east) vs. “down” (west) and “hillwards” vs. “seawards”, and Jaminjung between “upstream” vs. “downstream” (Levinson & Wilkins, 2006).

Most acquisition studies of FoR terms have focused on learners of languages that have relative FoRs. Typically, in such languages, a single set of terms (e.g., front/back, left/right) marks both intrinsic and relative FoRs (Levinson & Wilkins, 2006) and the acquisition of these terms follows a cross-linguistically robust pattern (Johnston & Slobin, 1979; Rigal, 1994, 1996). Children’s earliest knowledge of front and back emerges around age 2 and corresponds to intrinsic FoR instances that take one’s own body as the reference object (Kuczaj & Maratsos, 1975; Levine & Carey, 1982), followed a year later by intrinsic FoR instances that take objects with intrinsic facets as reference objects (E. Clark, 1980; Goodglass, Gleason, & Hyde, 1970; Grimm, 1973; Kuczaj & Maratsos, 1975; Levine & Carrey, 1982; Tanz, 1980). Around the age of 4 or later children show evidence of relative FoR meanings that involve applying their own viewpoint to objects without inherent fronts and backs and extending such uses to incorporate another person’s viewpoint (e.g., E. Clark, 1980; Goodglass, Gleason, & Hyde, 1970; Grimm, 1975; Johnston, 1984; Kuczaj & Maratsos, 1975; Levine & Carey, 1982; Tanz, 1980; Weissenborn, 1981). The acquisition of left/right follows a similar sequence but lags
considerably behind *front/back*, presumably because of additional computations required to
differentiate the secondary left-right axis after the primary front-back axis has been defined
(Elkind, 1961; Harris, 1972; Irwin & Newland, 1977; Rigal, 1994, 1996; Shusterman & Li,
2016b). This sequence of (sub-types of) intrinsic and relative FoRs has been attributed to the
increasing conceptual demands on perspective-taking posed by the relative FoR in its various
incarnations (cf. Piaget & Inhelder, 1967; but see Rubio-Fernandez, this volume; Shusterman &
Li, 2016b for more nuanced discussion).

Less is known about the acquisition of absolute systems. In a longitudinal naturalistic
production study, Brown and Levinson (2000) found that Tzeltal-speaking children start using
absolute terms (‘uphill’, ‘downhill’) around age 2, but use them relationally (“X is uphill of Y”) only by age 3;6, and integrate them into adult-like requests to others for manipulating objects in
a tabletop array only by 7 or 8 (ibid.; see also de León, 1994, on the acquisition of other absolute
systems). Other work shows that absolute (environment-based) representations are also available
to learners of languages that do not prioritize this FoR: Shusterman and Li (2016b) report that 4-
year-old English-speaking children readily map absolute (north/south) meanings onto novel
ambiguous spatial terms (e.g., *It is on the ZIV/KERN side of the room*; cf. also Haun, Rapold,
Call, Janzen, & Levinson, 2006). Beyond these basic patterns, the acquisition of the conventions
regarding how a specific language community derives and uses FoR terms is quite protracted
(Abkarian, 1982; Harris & Strommen, 1979), and there are key language-specific aspects in the
profiles of different languages (see, e.g., Brown & Levinson, 2000; de León, 1994, on the
acquisition of intrinsic terms in learners of absolute languages).

**Does spatial language affect spatial cognition?**
The evidence reviewed throughout this chapter suggests a tight causal relationship between spatial language and cognition, since the linguistic encoding of space builds on antecedently available, pre-linguistic spatial concepts in important ways. One might ask whether this causal relationship could be reversed—i.e., whether spatial language itself might affect the way spatial categories are acquired, perceived, categorized and remembered. If so, cross-linguistic differences in the encoding of space might create cognitive discontinuities among speakers of different languages. This topic has attracted a lot of recent attention within a larger discussion about the role of language in cognition (see Ünal & Papafragou, 2016, for a recent review).

In the domain of location, there is evidence that spatial semantic distinctions do not shape non-linguistic cognition. English makes a distinction between on and above but Japanese and Korean do not; nevertheless, categorization patterns for these spatial relations converge regardless of language background (Munnich, Landau & Dosher, 2001). Other work has argued for the opposite conclusion. Recall that, unlike English that draws a distinction between containment and support in its prepositional system (in vs. on), Korean makes a distinction between tight fit and loose fit in its verbal system that cross-cuts the containment-support boundary (e.g., kkita ‘put tightly in/on/together/around’ vs. nehta ‘put loosely in/around’; Bowerman & Choi, 1991). Both English- and Korean-speaking infants distinguish between tight and loose fit when processing containment and support scenes; Korean-speaking adults continue to attend to the degree-of-fit when categorizing spatial relations of containment and support but English-speaking adults do not (Hespos & Spelke, 2004; McDonough et al., 2003; cf. also Choi, 2006). These results have been taken to suggest that linguistic encoding decreases the cognitive salience of degree-of-fit relations in mature English speakers (ibid.). However, the empirical
picture is complex and does not clearly support this conclusion: Norbury, Waxman, and Song (2008) showed that both English- and Korean-speaking adults were sensitive to the non-linguistic dimension of fit; furthermore, for both groups, tight-fit relations were more salient than loose-fit relations. These results suggest that adults’ non-linguistic representation of fit does not depend on the language they have acquired (if anything, its structure is characterized by a bias that emerges in speakers of different languages and might involve a deeper perceptual-cognitive asymmetry; Norbury et al., 2008).

Similar results have been obtained in the domain of motion. Papafragou et al. (2002) found that, although children and adult speakers of English and Greek described motion events differently, they did not differ in their memory and categorization of these events (see also Gennari, Sloman, Malt & Fitch, 2002). A further study (Papafragou, Hulbert, & Trueswell, 2008) compared attention allocation to motion events (measured by eye movement patterns) in adult speakers of English and Greek. During a verbal description task, when people were preparing to describe these events, they allocated their attention to components of motion events in ways that reflected language-specific encoding patterns (i.e., speakers of English attended earlier to manner of motion and speakers of Greek attended earlier to path; see also Bunger, Trueswell & Papafragou, 2012, for evidence of similar effects in young children). However, during a memorization task, when participants freely inspected the ongoing motion events, the language-specific effects disappeared. These results suggest that motion event perception is not guided by the perceivers’ native language, even though language-specific patterns in attention emerge when the task specifically involves the recruitment of linguistic representations (as in language production).
A striking finding from Papafragou et al. (2008) was that, during the memorization task, after having seen the motion events unfold and when trying to memorize them, speakers of English and Greek attended to different components of the events (i.e., path for speakers of English and manner for speakers of Greek). A subsequent eye tracking study (Trueswell & Papafragou, 2010) showed that these effects disappeared when participants were asked to perform a secondary task that engaged the language faculty (i.e., counting aloud) but not when they performed an equally taxing secondary task that did not engage the language code (i.e., tapping a rhythm). These results suggest that participants in Papafragou et al. (2008) spontaneously recruited language online to support the representation of an event in memory: participants attended to aspects of motion events that were encoded outside the main verb in their native language and, thus, might be forgotten. However, the online recruitment of language to support cognitive operations was flexible and task-dependent; furthermore, such linguistic intrusions could be blocked by secondary tasks that interfered with linguistic encoding (see also Athanasopoulos & Bylund, 2012; Athanasopoulos et al., 2015).

Finally, in the domain of Frames of Reference, Levinson and colleagues (Levinson, 1996; Pederson et al., 1998; see also Majid et al., 2004) compared how speakers of Dutch (a relative FoR language) and speakers of Tzeltal (an absolute FoR language) responded in various spatial tasks. For instance, Pederson et al. (1998) tested participants in the Animals-in-a-Row task, in which participants first studied a line of toy animals on a table, then were rotated 180 degrees and moved to a different table, where they were given the animals and were asked to make it “the same” as what they saw before. Speakers of Dutch solved the task by applying a relative strategy (i.e., maintaining the left-right orientation of the animals in the row) but speakers of Tzeltal solved the task by applying an absolute strategy (i.e., maintaining the cardinal orientation
of the animals in the row). Further studies have confirmed the presence of a correlation between the
dominant FoRs in a specific language community and the use of FoR representations in non-verbal
cognitive tasks in members of that community (Haun et al., 2006, 2011). Such findings have been
taken as indications that language-specific preferences shape non-linguistic spatial cognition (ibid.).

Other work has challenged this view. Li and Gleitman (2002) showed that speakers of English (a relative FoR language) could provide different responses in the Animals-in-a-Row task depending on the testing conditions: when tested indoors without access to external landmarks, English speakers patterned with the Dutch speakers and used the Relative solution; but when tested outdoors or indoors with landmark information present, English speakers were more likely to make use of the Absolute solution, like the Tzeltal speakers. Similarly, Li et al. (2011) showed that Tzeltal speakers could use relative strategies to solve rotation problems when given subtle hints about the solution that was sought by the experimenter; furthermore, Tzeltal speakers were more accurate in Relative than in Absolute solutions, elicited under similar conditions. This work suggests that participants may fall back onto the FoR linguistic conventions of the community when interpreting ambiguous instructions such as “make it the same” (Li et al., 2011) but such biases do not limit the representation of FoR in cognition.

Summarizing, despite the presence of cross-linguistic differences in the domains of locatives, motion, and Frames of Reference, the underlying cognitive representations are remarkably similar in members of different language communities. Nevertheless, spatial language regularly intrudes into cognitive processing, even when it is not explicitly invoked or necessary, especially when the task is cognitively demanding (Papafragou et al., 2008) or ambiguous (Li et al., 2011). As a result, cognitive and linguistic categories of the spatial world,
although dissociable, are often highly correlated. This line of reasoning is consistent with evidence that *overt* use of spatial terms benefits children’s and adults’ performance on a variety of spatial tasks, including spatial categorization (Casasola, 2005; Casasola & Bhagwat, 2007), spatial analogy (Loewenstein & Gentner, 2005), spatial memory (Dessalegn & Landau, 2008, 2013), and navigation (Hermer-Vazquez, Spelke, & Katsnelson, 1999; Pyers et al., 2010; Shusterman, Lee & Spelke, 2011; Spelke & Tsivkin, 2001). In all such cases, spatial language (whether covertly or overtly introduced) may augment representational or processing resources by helping identify, store and/or manipulate spatial information.

**Concluding remarks**

Decades of research on spatial terms have revealed a complex set of factors that shape the nature, use and acquisition of spatial vocabularies. Several pieces of evidence support the conclusion that spatial language—at least in part—reflects a set of non-linguistic, potentially universal, cognitive spatial primitives. Nevertheless, detailed studies of individual linguistic systems make it clear that there are many differences in how individual languages talk about space. The literature we reviewed has highlighted the importance of conducting research with diverse populations (e.g., speakers of different languages, typically and atypically developing children, infants, non-human species) and studying spatial language and cognition with a variety of empirical methods (e.g., linguistic tests of production and comprehension, cognitive tests of memory and categorization, eye-tracking). The current state of the art calls for a nuanced position both on how spatial terms are acquired cross-linguistically (since learning to speak about space does not involve a simple mapping between concepts and spatial terms) and on how spatial language connects to non-linguistic cognition. Future work should pursue the quest for
linguistic-semantic spatial universals through both field work with speakers of many different languages and formal analyses of the semantics of space. Relatedly, future work needs to provide a fuller map of the non-linguistic cognitive presuppositions of spatial language through a variety of empirical methods (including neuroscientific approaches; e.g., see Burgess, 2008; Wolbers & Hegarty, 2010).

A particularly rich avenue for further research in spatial language involves the role of pragmatic inference. Since language is limited and can only express certain aspects of non-linguistic spatial representation (Landau & Jackendoff, 1993; Talmy, 1983, 1985), pragmatic inference plays an important role in both how people choose spatial expressions as speakers and interpret such expressions as comprehenders (Herskovits, 1985; Levinson, 2000). For example, the English preposition in and related containment expressions across languages can be used to convey related but distinct relations such as full containment (“coffee in a cup”) or partial containment (“pencil in a cup”), depending on one’s knowledge about the specific objects in the scene (Herskovits, 1985; Levinson, 1995, 2000). Addressees use implicature (Grice, 1975) or pragmatic enrichment (Sperber & Wilson, 1986/1995; cf. Degen & Tanenhaus, this volume) to add contextual refinement to coarse spatial meanings and speakers anticipate such means of reconstructing the exact spatial configuration conveyed through a spatial description. Such patterns of language use and interpretation have been argued to be impressively consistent across the world’s languages (Levinson, 2000; Levinson & Wilkins, 2006), although language-specific encodings may result in different pragmatic inferences for the same spatial configurations (Bowerman, 1996).

Currently a growing body of experimental work documents pragmatic contributions to spatial meaning. Several cross-linguistic studies have shown that children’s and adults’ choice
of spatial terms to describe space and motion scenes depends on whether these terms make an appropriate and specific informational contribution compared to other alternatives (Grigoroglou, Johanson & Papafragou, in press; Papafragou, Massey & Gleitman, 2006; Papafragou, Viau, & Landau, 2013; Tanz, 1980). A separate strand of research has shown that context provides sets of expectations that guide both the production of spatial descriptions and the interpretation of spatial language in conversation (Andonova, Tenbrink & Coventry, 2010; Carlson & Covey, 2005; Carlson & Kenny, 2006; Coventry, Carmichael & Garrod, 1994; Coventry, Tenbrink, & Bateman, 2009; Li et al., 2011; Morrow & Clark, 1988; Ullman, Xu, & Goodman, 2016). Finally, other work argues that general pragmatic principles affect the shape of cross-linguistic spatial systems themselves (Khetarpal, Majid, & Regier, 2009; Khetarpal, Neveu, Majid, Michael, & Regier, 2013). The integration of these directions with research on spatial semantics and cognition is particularly promising for future research.

**Acknowledgement**

Preparation of this chapter was supported in part by grant #1632849 from the National Science Foundation.
References


doi:10.1037/a0036970


