



HAL
open science

How language influences spatial cognition, categorization of dynamic motion events and gaze behavior: a crosslinguistic comparison

Efstathia Soroli

► **To cite this version:**

Efstathia Soroli. How language influences spatial cognition, categorization of dynamic motion events and gaze behavior: a crosslinguistic comparison. *Language and Cognition*, 2023, pp.1-45. 10.1017/langcog.2023.66 . hal-04350367

HAL Id: hal-04350367

<https://hal.science/hal-04350367>

Submitted on 26 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.


L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

ARTICLE

How language influences spatial thinking, categorization of motion events, and gaze behavior: a cross-linguistic comparison

Efstathia Soroli 

Laboratoire Savoirs, Textes, Langage (STL), University of Lille, UMR 8163 – CNRS, Lille, France

Corresponding author: Efstathia Soroli; Email: efstathia.soroli@univ-lille.fr

(Received 19 March 2023; Revised 17 November 2023; Accepted 27 November 2023)

Abstract

According to Talmy, in *verb-framed languages* (e.g., French), the core schema of an event (Path) is lexicalized, leaving the co-event (Manner) in the periphery of the sentence or optional; in *satellite-framed languages* (e.g., English), the core schema is jointly expressed with the co-event in construals that lexicalize Manner and express Path peripherally. Some studies suggest that such differences are only surface differences that cannot influence the cognitive processing of events, while others support that they can constrain both verbal and non-verbal processing. This study investigates whether such typological differences, together with other factors, influence visual processing and decision-making. English and French participants were tested in three eye-tracking tasks involving varied Manner–Path configurations and language to different degrees. Participants had to process a target motion event and choose the variant that looked most like the target (non-verbal categorization), then describe the events (production), and perform a similarity judgment after hearing a target sentence (verbal categorization). The results show massive cross-linguistic differences in production and additional partial language effects in visualization and similarity judgment patterns – highly dependent on the salience and nature of events and the degree of language involvement. The findings support a non-modular approach to language–thought relations and a fine-grained vision of the classic lexicalization/conflation theory.

Keywords: English/French; eye-tracking; language-thought interface; linguistic and cognitive processing of events; Path/Manner salience; production; reaction times; scene variability; similarity judgments; voluntary motion events

1. Introduction

In cognitive science, the traditional view is that cognitive processing is modular and that high-level thinking (such as categorization, reasoning, and decision-making)

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike licence (<http://creativecommons.org/licenses/by-nc-sa/4.0>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the same Creative Commons licence is used to distribute the re-used or adapted article and the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use.



involves symbolic computations that are not directly linked to our perception and action systems (Fodor, 1983; Mahon & Caramazza, 2008).¹ Within this framework, the cognitive and language systems are also believed to be independent of each other and guided by universal determinants (Chomsky, 1977; Gleitman et al., 2007; Papafragou et al., 2008; Pinker, 1989; Tomasello, 2003).

Over the last two decades, an opposing, situated view has gained scientific support, arguing that cognitive processing is better understood not as a set of isolated computations that take place solely inside the brain, but rather as emergent properties that result from the constant interaction of the brain with the body (sensorimotor system) and the environment. For such views, see, for example, the embodied accounts of cognition proposed by Barsalou (2008) and others (e.g., Gallese & Lakoff, 2005; Willems & Casasanto, 2011; Willems & Hagoort, 2007); the action perception theory proposed by Pulvermüller (2013) (see also Kiefer & Pulvermüller, 2012; Pulvermüller, 2018; Pulvermüller & Fadiga, 2010); and the complex and dynamic systems theory by De Bot (2017). In this line of work, various high-level cognitive processes, including language processing, have been found to actively interact with both perceptual and motor systems (e.g., for visual perception–language interactions, see Anderson et al., 2011; Lupyan & Spivey, 2010; Richardson & Matlock, 2007; for motor action–language interactions, see Pulvermüller et al., 2005; Spivey, 2007; Wispinski et al., 2020; Zgonnikov et al., 2017; and for motor action–visual perception interactions, see Tucker & Ellis, 1998; Richardson et al., 2001).

With respect to the interaction of cognition with the environment, more specifically with the linguistic information in the environment, a more relativist approach holds that human cognition may be additionally shaped by language-specific factors (Boroditsky, 2001; Bowerman & Levinson, 2001; Choi, 2006; Gentner & Goldin-Meadow, 2003; Hickmann, 2006; Lucy, 1992; Whorf, 1956, among others). In this context, recent psycholinguistic research has witnessed a growing interest in the language–thought interface, with particular attention to the possible impact of language-specific properties on our cognitive and even perceptual and action mechanisms of processing (Gibson et al., 2017; Goller et al., 2020; Lupyan et al., 2020; Pulvermüller, 2018; Yun & Choi, 2018).

Although most researchers (mostly from the domain of psychology and neurosciences) ‘do not find plausible the idea that the language system is encapsulated’ (Fedorenko & Varley, 2016, p. 16) and rather suggest that language and thought are two distinct and independent systems (e.g., Monti et al., 2012; Papafragou & Selimis, 2010; Varley et al., 2005), some recent studies in cognitive science and neuroscience tend to admit that language representations might be used in support of reasoning, across a range of domains, particularly under conditions of high cognitive load. More specifically, the role of language has been recognized as important in non-verbal thinking, for example, for the development of certain abilities in understanding others’ mental states (e.g., de Villiers & de Villiers, 2000), for the storage and

¹According to Jerry Fodor (1983), brain computations are achieved within an *amodal symbol system* – a system of mental representations that is not tied to any specific sensory modality. In Fodor’s view, cognitive processes have the flexibility to operate on information from different modalities and involve the manipulation and interpretation of abstract symbols, which are independent of the sensory inputs that they may represent.

manipulation of important quantities of information (e.g., Deldar et al., 2021), or for complex problem-solving (e.g., Baldo et al., 2005). In other words, in this line of research, language is not considered as fundamental for thinking but rather as a system playing an accessory role, especially when the task at hand is rather demanding, that of a facilitator. It has been shown, for example, that thought (e.g., arithmetic processing, inhibition, theory of mind, music processing, and spatial navigation) is possible without language. For instance, individuals with aphasia, who have almost no or partially impaired ability to understand or produce language, are able to add, subtract, and solve logic problems, think about another person's thoughts, appreciate music, and successfully navigate and explore their environments (e.g., Soroli, 2018; Willems et al., 2011). But is thought limited to these functions? What about complex reasoning such as event recognition, analogical thinking, and decision-making?

The focus of this paper is on spatial thinking and reasoning and on the role language plays (central or accessory) during motion event perception and recognition, attention allocation to different spatial components, spatial encoding, categorization, and decision-making.

Space provides a rich and experimentally tractable domain to investigate language–thought relations of this type (Boroditsky, 2001; Levinson, 1996) because it is fundamental to human existence while also characterized by considerable cross-linguistic variability (Talmy, 1985). With respect to the expression of motion events, this variability includes (but is not limited to) asymmetries across and within languages in terms of (a) *types of semantic information* that are preferentially encoded (*semantic focus*) across systems, such as Path and Manner (Jackendoff, 1990; Levin & Rappaport Hovav, 1992) and Source and Goal (Kopecka & Vuillermet, 2021); (b) *types of lexical and grammatical encoding patterns* (*locus/distribution of components*) (Matsumoto, 2003; Talmy, 2000); and (c) *density, frequency, and complexity* (*utterance architecture*) of the encoded information (Hickmann et al., 2017; Soroli & Verkerk, 2017), which reflect differences in the *relative saliency* of spatial components and variation in the potential combinatorial assemblies a language offers (Ibarretxe-Antuñano, 2009; Slobin, 2006).

Some studies suggest that such linguistic asymmetries do not affect underlying *on-line* processing (e.g., Gennari et al., 2002; Munnich et al., 2001; Papafragou et al., 2008), while others argue that perceptual and cognitive mechanisms are fine-tuned by language (e.g., Athanasopoulos & Bylund, 2013; Boroditsky, 2012; Choi et al., 2018; Goller et al., 2017; Levinson, 2003; Majid et al., 2004). The emerging view is that our perceptual and cognitive systems are partly adjustable depending on specific contexts, but the precise nature and role of factors contributing to activate language effects are still not well understood.

The aim of this study is to determine whether specific typological differences affect visual *on-line* event processing in relation to specific features of motion events and the extent to which speakers of different languages may attend to different aspects of events while making non-verbal similarity judgments about them and while integrating them into linguistic structures. After a brief review of previous relevant research (Section 1.1), a list of factors that may influence the degree to which language-specific properties constrain verbal and non-verbal behavior is discussed (Section 1.2). Several hypotheses about the potential relationship between verbal and non-verbal behavior are also formulated, with a special focus on encoding patterns, attention allocation, and non-linguistic processing (Section 1.3). The rationale of the

method used and the results of a production task and two similarity judgment experiments that tested these hypotheses are presented in Sections 2 and 3, respectively. These sections are followed by a discussion and a conclusion (Sections 4 and 5, respectively) that raise broader issues about the language–thought interface and highlight the contribution of the present study to dynamic and multidimensional models that take into account cognitive, perceptual, and linguistic interactions.

1.1. Linguistic and non-linguistic representations of motion across languages

Although a common set of semantic components of motion can be expressed in most languages, such as Manner (e.g., *to jump*, *to crawl*) and Path (e.g., *up*, *into*, *across*), languages provide speakers with a limited number of linguistic means to encode these aspects, acting as ‘filters’ that lead speakers to focus on particular features of scenes and sub-events in their verbalizations and organize them in very different ways (Slobin, 1987). According to Slobin, such differences depend on the subjective component of motion, namely the Manner of motion, and its relative salience from one language to the other (see, i.e., the *Manner cline* proposed by Slobin, 2006). In contrast, according to Talmy (2000), such variation stems from differences in the objective aspects of motion, namely the expression of Path.

According to Talmy’s *lexicalization framework* (LF) (1985), the languages of the world offer different form-to-meaning mappings to their speakers for the expression of the core spatial component: the Path. For instance, based on the semantics of motion verbs, Talmy makes the distinction between so-called *satellite-framed* languages, such as English, which privilege the lexicalization of Manner, expressed in the main verb, leaving Path in satellites² (1), and *verb-framed* languages, such as French, which highlight the Path of motion instead and leave Manner peripheral (2a) or unexpressed (2b).

- (1) *A woman is walking out*
 MANNER PATH
- (2) a. *Elle sort (en marchant).*
 PATH MANNER
 ‘She exits (by walking)’
- b. *Elle sort/part.*
 PATH
 ‘She exits/leaves’

²Talmy’s definition of a satellite involves ‘the grammatical category of any constituent other than a noun-phrase or prepositional-phrase complement that is in a sister relation to the verb root. It relates to the verb root as a dependent to a head’ (Talmy, 2000: 102). In this work, this strict use of the term ‘satellite’ was enlarged to include all linguistic means outside of the verb root contributing to motion expression (e.g., particles, prepositions, adverbials, and gerunds), as also suggested by Hickmann et al. (2017) and Beavers et al. (2010) (for a discussion on this issue, see also Fortis (2010)).



Figure 1. Example of a motion event video.

Later, Talmy focused on the way a complex event is integrated into a clause and formulated his *event integration framework* (EF) (1991). More specifically, to fully and compactly describe a voluntary motion event (Figure 1), the speakers need to integrate into their description two events: (a) an event in which the figure is performing a motion (a displacement, such as the one depicted here: *moving from inside to outside*) and (b) an event in which the figure is moving in a certain Manner and which encodes the *co-event* (here: *walking*). On the two events, the first (the motion event) plays a primary role in the event complex. According to Talmy (2005), this motion event describes the central relationship between the involved figure and the ground (a special configuration that changes during the displacement) and thus is to be considered as the *framing event* with a *core schema* that encodes the Path traced by the moving figure. In the example described above (Figure 1), the ‘walking’ event is the *co-event* that holds a particular supplementary relation to the *framing event* and describes a relation of a specific Manner of displacement.

The EF is different from the LF: The EF focuses on the constituent that encodes the core event schema, the Path, and how this is related to the *co-event* – what is called in this paper: the *Architecture-based framework* that is interested in the relationship between the type of the described events and the constructions used to package them (utterance architecture); the LF focuses on the specific spatial components of motion (e.g., Path, Manner, Cause), more specifically whether these components are lexicalized or not – what is called in this paper: the *focus–locus dimension* that is interested in the kind of spatial component(s) encoded (*focus analysis*) and the specific morphosyntactic units (*loci*) used to express them (*locus analysis*).

From the EF point of view, in a *satellite-framed* language such as English, speakers integrate into one syntactically compact and semantically dense clause both the *co-event* (Manner) expressed in the verb and the *framing event* (Path) expressed in extra-verbal elements, in satellites such as particles (e.g., *out*), as well as in other devices (e.g., prepositions such as *into*). In contrast, in *verb-framed* languages, such as French, speakers tend to adopt a distributed pattern in which information about the *framing event* (Path) is encoded in the verb, but the *co-event* is left in the periphery, expressed with extra-verbal elements (e.g., adverbials), distributed in other clauses (e.g., gerunds, coordinated/juxtaposed propositions), or completely omitted, as illustrated in (2a) and (2b), respectively.³ Such cross-linguistic variation in this *filtering process*

³As indicated by square brackets in the literal translations of the examples, French does not mark aspects in the present tense (e.g., no morphological progressive marker). With some verbs, oral French does not mark

raises questions about the relative impact of language-specific factors on low-level mechanisms contributing to motion event construal and leads to (at least) two main questions: Does this variability mean that people attend to things differently when viewing the same events? Do they focus on different components/different sub-events? If so, what is the psychological reality of these different lexicalization/event integration patterns?

1.2. *Weight of the language factor in relation to scene, task, and event types*

The relationship between visual processing and verbal planning is still not fully understood mainly because non-linguistic event representations are hard to access and specify (cf. Bock et al., 2004; Jackendoff, 1996). Most studies reporting language effects on conceptualization⁴ (see also Levelt, 1989, for a tripartite model of the speaking process) are based on analyses of production data that show different verbal behaviors across languages, suggesting that the particular linguistic resources of speakers' native language invite different event conceptualization and encodings. It is only recently that researchers have begun to distinguish and experimentally explore the relationship between event construal (mechanisms of 'attention direction' that help (re)direct attention toward certain aspects of a situation reflecting the speakers' ability to adjust) and event description (differential selection of linguistic resources for verbal encoding). The systematic investigation of verbal and non-verbal data now shows a surprisingly tight temporal coupling between these two types of behavior (Gleitman et al., 2007) with a large overlap between conceptualization and planning processes (but see Griffin & Bock, 2000, for a sequential account of processes at the conceptualization and formulation levels).

Studies examining various types of non-verbal behavior (e.g., memory, categorization, eye-movements) beyond verbal production present rather divergent results. Some report either no language effect on cognition or effects that are not clear and/or viewed as being superficial (Landau & Lakusta, 2006; Papafragou et al., 2002, 2006). Others find language-specific differences in non-linguistic measures captured immediately after verbalization (e.g., Gennari et al., 2002; Naigles & Terrazas, 1998; Slobin, 2005) and/or when linguistic forms are recruited for explicit encoding (Papafragou et al., 2008; Papafragou & Selimis, 2010), suggesting that the nature or the demands of the task may affect differently the language–thought interaction (see also Soroli et al., 2019, for a recent discussion). For example, preparing to speak (in a verbal production task) constrains not only which components speakers choose to express but also how they allocate visual attention to

singular 1st, 2nd, and 3rd persons or plural 3rd person. Changes of location are marked mostly in the verb rather than in particles or spatial prepositions (e.g., *dans* 'in/into'), except for the use of some additional adverbial phrases (e.g., *rapidement* 'quickly') that often specify further the Manner of motion.

⁴Here, the term *conceptualization* refers to one of the main activities involved in natural language generation/speaking. According to Levelt (1989), speaking involves a *conceptualization* activity during which the speaker selects what their discourse will be about, a *formulation* activity during which the speaker decides how to express it, and an *articulation* activity that consists of actually saying it. In cognitive science, the term *conceptualization* is extensively used to refer to a general sense-making process, the way we conceive and understand the world. Within the embodied approach, *conceptualization* is a hypothesis according to which concept acquisition is constrained by the properties of one's body, suggesting that organisms with different bodies conceive of the world differently (Shapiro, 2011).

these components very early during the visual processing of an event (e.g., Flecken et al., 2014; Soroli et al., 2019). Being instructed to provide a verbal output or process verbal material leads the viewer/speaker to focus on relevant aspects of the scene for sentence planning and sentence comprehension right from the start of stimulus onset and in order to optimize the uptake – the selection of the most adequate construal (for a review, see Divjak et al., 2020; Griffin, 2004; Meyer & Lethaus, 2004). The focus on relevant aspects of a scene is typically captured by gaze measures such as eye-fixations (e.g., numbers and duration of fixations), commonly used to study cognitive processing, attention allocation, intentions, and more generally the on-line strategies of the viewers (e.g., Park et al., 2016).

Despite the fact that many studies that involve preparation for speaking or processing of verbal input tasks report robust effects on how people allocate their visual attention, little is still known about attention allocation when language is not explicitly involved during non-verbal tasks (non-verbal input and output) or the relative weight of language in relation to other factors that may create different pressures on *on-line* processing. For example, some studies have shown that the impact of language-specific features on event exploration depends on the nature of the scenes in which an event occurs. More specifically, in a preliminary study using two types of stimuli (animated cartoons and video clips) Soroli and Hickmann (2010) found differences in production as well as eye-tracking measures not only as a function of language but also as a function of scene types. Similar results are also reported by Hickmann et al. (2017) as well as by Henderson and Ferreira (2004), suggesting that the placement and duration of gaze fixations may depend on the specific kinds of visual information to be processed during and even before verbalization.

Some studies focus on the relative impact of different types of motion components, raising more specific questions about the exact features that induce cross-linguistic variation in verbal and non-verbal processing, underlining the need for a fine-grained analysis of event types. For example, different types of Path (cf. Ibarretxe-Antuñano, 2009; Talmy, 2000) and Manner (cf. Slobin et al., 2014) constrain to different degrees attention to specific dimensions of events even within a given language. According to Talmy (2000) and Ibarretxe-Antuñano (2009), Path is the most basic component defining motion, and its relative salience in a given language affects speakers' verbalizations to different degrees. Slobin et al. (2014) consider that the major distinctive feature that determines the likelihood and lexical richness of spatial expressions in a system is the salience of Manner instead. Soroli (2011) also shows that particular features of Manner, including Manner of motion with an instrument (e.g., cycling) or without (running), constrain differently behavior across language groups, for example, inducing richer Manner expression in English than in French, even in unmarked co-events, such as *walking* (prototypical Manner of movement for humans) (but see also Hickmann et al., 2017).

The assumption that different languages and spatial event properties or types of scenes contribute to different ways of conceptualization options needs, however, further specification and careful operationalization. For example, in the domain of spatial language and cognition some studies report great variability with respect to the type of stimuli used. For example, Hickmann et al. (2017), who used animated cartoons in their design, surprisingly report strong Manner saliency with *walking* events and fail to replicate findings from other similar studies that use real-motion

video events and suggest low saliency for this kind of motion (cf. Flecken et al., 2014; Soroli, 2012). Hickmann and colleagues further discuss the possibility that different types of stimuli (video/cartoons) may induce differences in the sensitivity to specific event properties and lead to misleading results. They admit that although cartoon-like stimuli allow to control for many variables (neutralization of moving backgrounds, control of contrast and speed for better identification of target figures and motion properties, etc.), cartoons are not as ‘ecological’ as other types of stimuli such as recorded films of natural motion in real settings and over-attract the viewer during processing. Similar inconsistencies are reported with material that uses human versus animal motion. With respect to this point, what can be prototypical and unmarked for human motion in terms of affordances (e.g., walking) may be highly salient and artificial for an animal motion event, as nonhuman animals do not walk, at least not as humans do (cf. Gibson, 1979).

To conclude, more fine-grained research is necessary to avoid the tendency of some studies to overgeneralize motion material processing (with cartoon-like/video motion, human/nonhuman) to any kind of motion processing and to constrain analysis in one of two ways: (1) by exclusively focusing on Manner or Path salience and (2) by simplifying event properties at the risk of proposing partially misleading conclusions, for example, by reducing Manner of motion to easy distinguishable but rather artificial body movements (e.g., use of figures without limbs; see Bohnemeyer et al., 2001; Montero-Melis et al., 2017) or by reducing Path to only one of its components (e.g., use of *Endpoints*; see Carroll & von Stutterheim, 2011; Papafragou et al., 2008).

1.3. Aims, research questions, and predictions

Although most approaches agree that typological differences affect speakers’ verbal behavior, there is no consensus about the contexts in which such differences arise, which specific levels of verbal encoding are most influenced by such differences, and whether they also affect non-verbal behavior. More specifically, the aim of this paper is to examine the language–cognition relationship by investigating (i) whether the differences in two typologically contrasted languages (English and French) constrain participants’ verbal and non-verbal performance; (ii) whether any variation in performance depends on differences in event types (with varied degrees of saliency of Path/Manner) and/or scene types (natural vs. artificial/cartoon-like voluntary motion sets); (iii) whether typological differences are reflected across encoding dimensions (semantic, lexical/morphosyntactic, utterance levels); and (iv) more importantly, whether any cognitive influences (during attention allocation and decision-making) arise both when language is and is not explicitly involved.

According to a first strong universalist hypothesis, language and cognition are two autonomous systems; thus, participants’ native language should not have an influence on the *on-line* processing of spatial scenes (neither on attention allocation patterns nor on decision-making). According to a strong relativistic hypothesis, language and cognition constantly interact; thus, the specific properties of individual languages, together with other external perceptual features, should leave their traces in all tasks and measures (verbal and non-verbal). According to a more moderate relativistic hypothesis, any potential language effects should only occur in tasks

involving (or requiring the processing of) explicit linguistic information or in perceptually salient contexts (e.g., scenes that involve non-prototypical or artificial events such as *jumping* or *crawling* humans, cartoon-like motion, etc.).

2. Method

2.1. Participants

A total of 49 native speakers were tested, and 40 participants were included in the analysis⁵ (20 participants per language and per experiment), half males and half females with comparable socioeconomic status. Participants were all university students, native, monolingual speakers of English or French, right-handed, above 18 years of age, and without any known acquired or developmental disorder. They all had early exposure to only one language, had not spent more than six months in a foreign country, and eventually learned (if any) a second language after the age of 10 (compulsory teaching at school). The recruitment and testing of the participants took place in two universities: English speakers were tested at Cambridge University (United Kingdom) and French speakers at the University of Paris (France). The participants received course credit compensation for their participation.

2.2. Materials and procedure

The participants were tested in three experiments involving different types of voluntary motion events executed in different Manners (with and without instruments) and along different Paths (with and without boundary crossings) and in tasks involving language to different degrees, all coupled with an eye-tracking paradigm. Participants' eye-movements during stimuli exploration were recorded with a Tobii X120 system that was placed in front of an 18,4" laptop monitor at a distance of about 70 cm from the participants.

Experiment 1 was a non-verbal categorization task in which participants had to perform non-verbal similarity judgments. They first watched a fixation cross (+) on the screen and then had to watch a 4-second target video film depicting a natural human voluntary motion event followed by a *beep* and two video variants: a variant depicting a similar Manner-congruent event and a variant depicting a similar Path-congruent event. They had to choose among the two variants the video clip that looked most like the target, as fast as possible (Figure 2). In total, 54 triads of this type were presented: 3 training items; 7 distractors involving motion of inanimate objects; 14 control items in which one of the variants was Manner- or Path-congruent to the target but performed by another figure; and 30 experimental items involving voluntary motion events in 5 types of Path (involving or not the presence and crossing of a boundary with and without a change of state) \times 3 types of Manner (with instruments, without instruments, and a default-Manner of motion) \times 2 versions/exemplars (each

⁵Given the answers provided by the participants to a sociolinguistic questionnaire and a few cases of insufficient datapoints for eye-movements, nine participants had to be excluded (from the French-speaking group, four were bilinguals and two had insufficient datapoints in their eye-tracking recordings; from the English-speaking group, two were bilinguals and one had insufficient eye-gaze datapoints).



Figure 2. Non-verbal categorization task: example of an experimental item involving a target (jump-out-of) and two variants: jump-into (Manner congruent) versus walk-out of (Path congruent).

Manner–Path combination was presented in two versions: one performed by a man and one performed by a woman).⁶

More specifically, the different Manners and Paths were selected with the following rationale. Manners varied along a continuum from the most to the least salient Manner of moving for humans – riding a bicycle, riding a scooter, roller skating > crawling, jumping, running > walking – and were expected to have different effects on attention allocation and non-verbal behavior in general. For example, Manners involving instruments were expected to be the most attractive ones for the viewers (following Hickmann et al., 2017; Slobin et al., 2014), while the least attractive and, consequently, the least salient were expected to be the most prototypical ones in terms of natural human affordance (walk) – a way of moving that we do not necessarily attend to as it is typically inferred/presupposed to be present by default. Paths included some displacements that involved the crossing of intrinsic boundaries and some that did not. They were selected also with respect to their relative salience (change of location with the presence/absence of a boundary, crossing/or not of the boundary, and involvement/or not of a change of state): single-boundary-P into/out-of (change of location + presence of a boundary + crossing of the boundary + change of state) > double-boundary-P across (change of location + presence of a boundary + crossing of a boundary – no change of state) > double-boundary-Default-P along (change of location + presence of a boundary – no boundary crossing – no change of state) > no boundary at all-vertical motion up/down (change of

⁶Stimuli were scattered in a randomized list that equally distributed the distractors and control items in a mixed set that differed from one participant to the other.

location – absence of a boundary – no boundary crossing – no change of state). The presence of intrinsic boundaries, their crossing, and their eventual combination with a change of state (e.g., into/out-of items) were expected to maximally attract speakers' attention to Path in both groups, but to a greater extent in French given its Path-based (verb-framed) lexicalization pattern.

Experiment 2 (verbal version of experiment 1) was a verbal categorization task, during which participants had to perform similarity judgments. They were presented with the same video set but first heard a target sentence instead of watching a target video – a construal encoding both Manner and Path (e.g., *There is someone walking in/On voit quelqu'un qui entre en marchant*) – and then had to choose among two video clips the variant best described by the sentence.

In experiment 3, participants had to describe voluntary motion scenes in a production task. The first set consisted of the same short video clips in which a human agent (man or woman) performed displacements indoors or outdoors, in varied Paths and Manners. The second set of stimuli (hereafter *cartoons*) consisted of short animated drawings showing different figures (animals and humans) moving along three types of Paths (*up, down, and across*), in different Manners (e.g., *climbing, swimming, walking*), and within varied background settings (e.g., *a mouse climbing up a table* – Figure 3). A total of 35 items were presented: 2 training items, 28 experimental items (10 video clips and 18 cartoons), and 5 distractors.⁷ Stimuli from both sets were scattered in a randomized list that equally distributed the distractors every five items in a mixed set that differed from one participant to the other. Participants first watched a fixation cross in the middle of the screen, then watched a clip followed by a beep and a blank/white screen, and finally had to describe what happened in the clip.⁸

2.3. Coding

2.3.1. Linguistic data

The verbal responses collected during experiment 3 (production task) were coded for the main motion components expressed (*focus*), for the linguistic means used to encode them (*locus*), and for the ways these components were organized in utterances (*architecture*). As illustrated in (3)–(7), responses fell into five groups, depending on their *focus*: only Manner (M), only Path (P), both components (PM), or neither (Z for simple motion verbs like *go/get*, as in (7)), as well as some rare cases where speakers did not provide any response or expressed some other spatial information considered as irrelevant, such as simple locative expressions or Cause (coded: NR).

⁷The distractor items showed displacements of inanimate entities (e.g., a ball bouncing, a bottle falling, and a box moving).

⁸The experimental procedure: Participants first saw experiment 1 (non-verbal categorization) to guarantee that participants would perform their similarity judgments without being influenced by any verbal input. Experiment 1 was followed by experiment 3, the production task. Participants saw experiment 2 at the end of the procedure, and to get maximal verbal contamination in this version of similarity judgments, they were expected to be influenced by both their own verbal productions (experiment 3/production task) and the sentences presented in target position during experiment 3.

- | | | |
|-----|--|----|
| (3) | <i>A man is jumping</i> | M |
| (4) | <i>Une femme traverse la rue</i>
'A woman crosses the street' | P |
| (5) | <i>A man is walking up the hill</i> | PM |
| (6) | <i>Il a grimpé</i> PM
'He climbed.up' | |
| (7) | <i>The man goes there.</i> | Z |

The *locus* analysis distinguished between verb versus other peripheral devices and the spatial components expressed in them (e.g., Manner verb in (3); Path verb in (4), Manner verb + Path peripheral device in (5) versus fused Path + Manner verb in (6)). The *architecture analysis* focused on how spatial components were packaged or distributed in the utterances: *Tight simple* (TS) constructions consisted of one simple, independent clause that encoded both the framing and the co-event in a compact structure (8); *tight complex* (TC) constructions contained at least one dependent clause or gerund (9); *loose simple* (LS) constructions were constructions in which information was spread over at least two juxtaposed or coordinated clauses (10); and *loose complex* (LC) constructions were constructions in which information was expressed in two or more clauses with at least one dependent element (11).

- | | | |
|------|---|----|
| (8) | <i>A man is cycling down to the hill</i> | TS |
| (9) | <i>La fille traverse la rue en courant</i>
'The girl crosse[s] the street by running | TC |
| (10) | <i>Une femme monte et court vite jusqu'en haut de la colline</i> LS
'A woman ascend[s] and run[s] quickly to the top of the hill' | |
| (11) | <i>La fille court jusqu'à l'autre côté, traverse la rue en faisant du roller</i> LC
'The girl run[s] to the other side, crosse[s] the street by skating' | |

The prediction was that speakers' descriptions would reflect the typological features of their language: They were expected to (a) express Manner (in the verb) and combine it with Path (in other devices) more frequently in English than in French and (b) use syntactically simpler and semantically more compact (TS) constructions in English than in French.

2.3.2. Categorization data

In the experimental items of the two categorization tasks (experiments 1 and 2), there was no correct answer. In order to check for unmotivated, biased, or random responses, control items were inserted in the tasks, in addition to the distractor items. The data collected within these experiments were analyzed with particular attention to (a) the number of correct responses (*accuracy rates*) in the control items; (b) variation in *Manner-congruent choices* across and within groups and item types in the experimental items; (c) *reaction times* (RTs) (in milliseconds) from stimuli onset

until participants' response; (d) the *fixation counts* to specific areas of interest (AoI) (Path vs. Manner areas); and the *fixation lengths* to those areas (in milliseconds).

2.3.3. Eye-tracking data

Participants' eye-movements were recorded while participants were exploring the scenes – the video variants in experiments 1 and 2 and the main video and cartoon clips in experiment 3 – in such a way as to measure their attention allocation to various aspects of motion events during decision-making and before verbalization, respectively. In order to better capture the dynamic nature of eye-movements, the coding of these data was not limited to static dimensions, such as the posture of the figure (for Manner) or the Endpoints of motion (for Path) as was done in most previous studies. Rather, spatially distinct components were defined in such a way as to represent dynamic areas during the actual displacement. As illustrated in Figures 3 and 4, the stimuli were divided into AoI, the coding of which corresponded to specific features of the event:

- (a) Path areas (S, P, G): Following Talmy's definition, Path was divided into three parts – an initial, an intermediate (median), and a final region, each one corresponding to an AoI: Source (S), (P), and Goal (G), respectively. Fixations were coded as *Pbroad* fixations (dashed area in Figures 3 and 4) when they fell into the three S -, P -, and G -AoI, excluding the moving limbs of the figure (the other half of the target scene covering the legs, see Manner area in (b)).
- (b) Manner area ($P \pm M$): Eye-movements to areas corresponding to the legs of the moving figures and to instruments were considered as (mainly) Manner fixations even if these areas also involved some aspects of the traced trajectory.

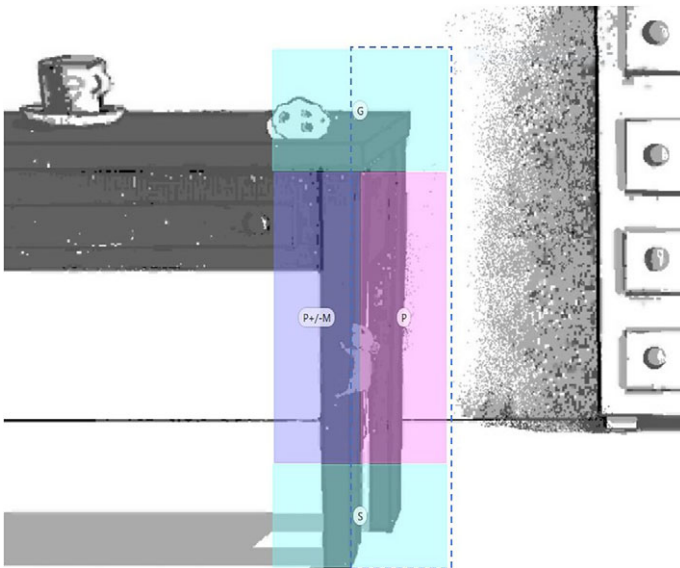


Figure 3. Areas of interest for events without boundary crossing (up/down) – cartoon scenes.

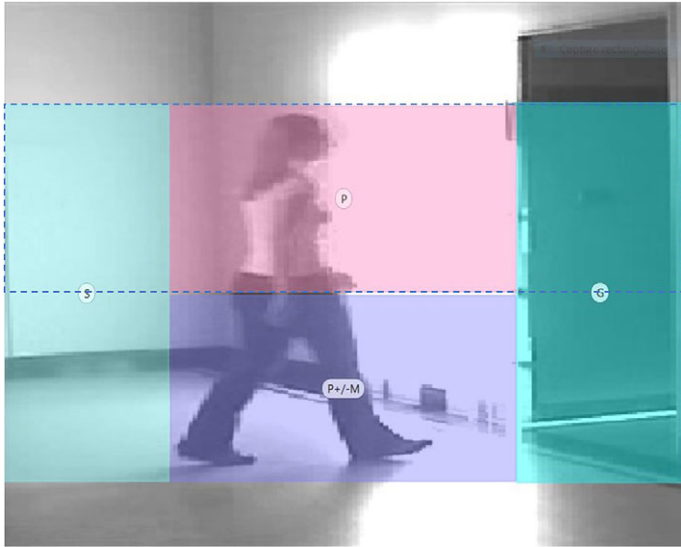


Figure 4. Areas of interest for single-boundary-crossing events (into/out-of) – video clips.

- (c) Additional M-congruent versus P-congruent areas were distinguished for the video variants involved in experiments 1 and 2 that corresponded to the Manner-congruent and Path-congruent choices the viewers were presented with, respectively.

The analysis of fixations involved measures of counts (numbers of fixations) and lengths (durations) but also more qualitative evaluations of the gaze scanpaths (gazeplots). Typically, fixation counts are indicators of the efficiency of information search and uptake during visual scanning; fixation lengths are measures that indicate attention maintenance; and gazeplots provide information about attention distribution and decoding complexity (more steps in the scanning process relate to more cognitive load) (Holmqvist et al., 2011).

According to the typological hypothesis, English viewers were expected to be more sensitive to Manner distinctions and thus focus more and longer on areas involving Manner ($P \pm M$) than French viewers. In addition, although all viewers were expected to focus on Path (P_{broad} areas), French viewers were expected to fixate these areas more often and for a longer duration. A main distinction was made between AoI that involved mostly Manner and those that did not. The analysis of Path fixations was twofold: First, Path fixations were considered separately (e.g., fixations falling into S-, G-, and P-AoI) in order to obtain their relative distribution across Path areas defined in the scenes; then, a distinction was made between P and P_{broad} fixations: the first corresponding to an analysis that considered only fixations that fell to the intermediate P area and the latter to an analysis that merged fixations falling to intermediate P together with fixations to S and G parts (marked in dashed lines in the Figures above). The anonymized quantitative summaries of the data used for the analyses can be found at https://osf.io/2hdxg/?view_only=9408ab7e47844ed2b302ef5eedc621b3.

3. Results

3.1. Verbal measures

The productions collected during experiment 3 were analyzed in several ways in order to determine which motion components were expressed (*focus*), with which linguistic means (*locus*, particularly main verbs vs. other devices), in which structures (*architecture*), and with which types of scenes. Separate analyses of variance (ANOVA) for videos and cartoons examined the effects of language (as between-subject factor) and of core-event-type (as within-subject factor) on several dependent variables.

3.1.1. Focus of information

For cartoons, as expected, PM responses were more frequent in English (83%) than in French (42%) where speakers expressed more often Path alone (55%). Mixed ANOVA examined the effects and interactions of the language factor (English, French) and of the core-event-type factor (up, down, across, along events) on PM responses.⁹ The results show significant main effects of language ($F(1,36) = 75.50$, $p < 0.0001$) and of core-event-type ($F(3,108) = 37.62$, $p < 0.0001$), as well as a significant interaction between these two factors ($F(3,108) = 8.27$, $p < 0.0001$). Additional specific contrasts between boundary-crossing events (across) and displacements without boundary crossing (up and down) show that P-only responses were mostly given by French participants, as opposed to the systematic PM responses of English participants (Figure 5). When French participants provided either fused or distributed PM responses, it was mostly with double-boundary crossings (across) and vertical (upward) events: PM responses with *across* events were more frequent than with *up* ($F(1,18) = 5.67$, $p = 0.02$) or *down* events ($F(1,18) = 153.72$, $p < 0.01$) and PM responses with *up* events were significantly more frequent than with *down* events ($F(1,18) = 49.78$, $p < 0.001$). In English, similar differences occurred in speakers' PM responses, except that PM responses with *up* and *down* events did not differ significantly ($up < across$: $F(1,18) = 9.26$, $p < 0.01$, $down < across$: $F(1,18) = 11.08$, $p < 0.01$, but up vs. $down$: $p = 0.13$ ns). French speakers' PM responses in *across* events

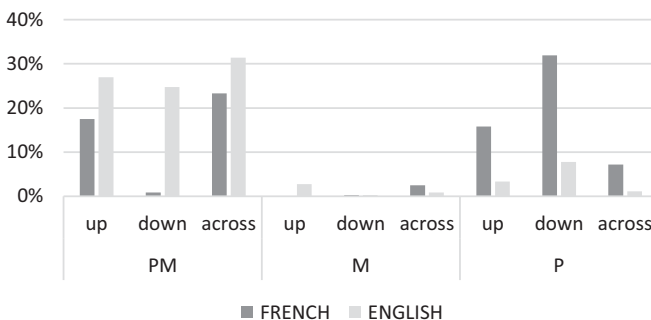


Figure 5. Focus of verbal responses across languages and core-event-types – cartoons.

⁹According to the gender similarities hypothesis (Hyde, 2005), no gender differences were expected in this domain. A preliminary analysis showed that gender did not have any significant effect ($p > 0.05$) and was therefore disregarded in subsequent analyses. Exact *p-values* are reported rounded to two decimal places in (close to) marginally significant and non-significant cases (ns).

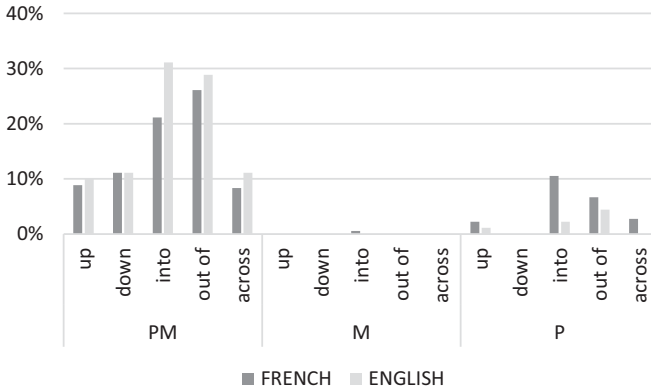


Figure 6. Focus of verbal responses across languages and core-event-types – videos.

were mostly due to distributed P and M encodings and in *up* events to uses of some verbs such as *grimper* ('to climb up'), which lexicalize both upward motion and Manner in a monomorphemic way (12).

- (12) *La souris grimpe pour aller chercher du fromage sur la table*
'The mouse climbs.up to go get some cheese on the table'

With respect to the *focus* of speakers in the video set, the analysis showed again significant main effects of language ($F(1,36) = 24.04, p < 0.0001$) and core-event-type ($F(5,180) = 5.84, p < 0.0001$), as well as an interaction between these two factors ($F(5,180) = 5.02, p < 0.001$). While overall PM responses were still significantly more frequent in English (92%) than in French (76%), speakers showed a general preference for PM responses in both languages, even in French where P-only responses with these items were quite rare (22%). As illustrated in Figure 6, further analyses showed that PM conflation in the video scenes was mostly due to single-boundary-crossing events (*out-of* items: 26% and *into* items: 21%) in both language groups and that the language effect stemmed mainly from the strong preference of English speakers to encode both Path and Manner, especially with *into* ($F(1,36) = 14.59, p < 0.001$) and *across* ($F(1, 36) = 6.08, p = 0.01$) events, as opposed to French speakers who did so but to a lesser extent.

3.1.2. Locus of information

In order to examine the effects and interactions of the three main factors (language, core-event type, and locus) on the expression of P, M, and PM components with cartoon items, three mixed ANOVA were performed. The analysis showed significant main effects of language (except for PM, see below), core-event-type, and locus (verb vs. other devices). More specifically, the expressed components varied significantly as a function of locus (PM: $F(1,36) = 63.63, p < 0.0001$; P: $F(1,36) = 90.11, p < 0.0001$; M: $F(1,36) = 337.39, p < 0.0001$) and core-event-type (PM: $F(3, 108) = 15.30, p < 0.0001$; P: $F(3,108) = 122.68, p < 0.0001$; M: $F(3, 108) = 106.55, p < 0.0001$). Collapsing languages, verbs encoded either Manner (45%) or Path (38%) and only rarely both (10%), while other devices encoded mostly Path (68%). However, the locus for P and M components (but not for their

combination) also varied significantly as a function of language (M: $F(1,36) = 75.06, p < 0.0001$; P: $F(1,36) = 26.91, p < 0.0001$; PM: ns). Figure 7 illustrates the distribution of the spatial components in verb and other devices as expressed by the two language groups: French speakers mainly encoded Path information in the verb (74%), sometimes double-marking it in peripheral devices as well (43%) or without any other information in the periphery (37%), and only rarely expressed Manner in the verb (6%) or in the periphery (18%). In contrast, Manner information was expressed massively in the main verb by the English speakers (84%), systematically combined with other linguistic devices that encoded Path (92%), as expected. Further analyses showed a significant interaction between locus and core-event-types (PM: $F(3,108) = 16.24, p < 0.0001$; P: $F(3,108) = 7.27, p < 0.001$; M: $F(3,108) = 32.99, p < 0.0001$). Speakers only rarely conflated both Path and Manner components in verbal devices (less than 20%), and when they did so, it was mostly the French speakers who opted for such confluations, especially with upward motion events (16%). Manner was encoded more often by the English speakers and lexicalized with *across > up > down* events (32%, 29%, and 23%, respectively), while Path lexicalization was more frequent in the French encodings, especially with *down > across > up* events (32%, 26%, and 16%, respectively).

Figure 8 illustrates how the target spatial components were distributed across different *loci* with videos. The analyses with this type of scenes showed again a

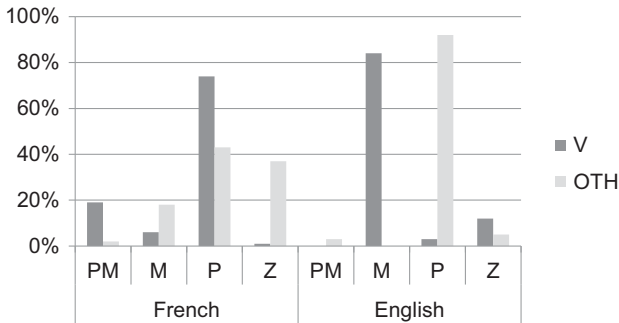


Figure 7. Locus of spatial information as expressed in verbs (V) and other devices (OTH) across languages – cartoons.

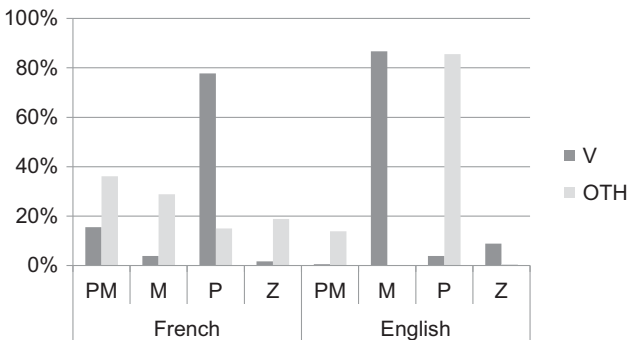


Figure 8. Locus of spatial information as expressed in verbs (V) and other devices (OTH) across languages – videos.

significant main effect of core-event-type but only a partial language and locus effect. More specifically, although the expression of all three components varied as a function of core-event-type (PM: $F(5,180) = 9.37, p < 0.0001$; P: $F(5,180) = 37.80, p < 0.0001$; M: $F(5,180) = 2.69, p = 0.02$), only P and M encodings varied significantly as a function of locus (P: $F(1,36) = 93.19, p < 0.0001$; M: $F(1,36) = 31.23, p < 0.0001$; PM: ns) and as a function of language (P: $F(1,36) = 16.36, p < 0.001$; M: $F(1,36) = 4.85, p < 0.05$; PM: ns). In French, as predicted, speakers mainly expressed Path in the verb (78%) and some Manner outside of the verb (29%). In contrast, English speakers systematically encoded Manner in the main verb (87%) and Path in the periphery (86%). Finally, the locus of information interacted with the core-event factor in the videos (PM: $F(5,180) = 5.06, p < 0.01$; P: $F(5,180) = 8.15, p < 0.0001$; M: $F(5,180) = 17.80, p < 0.0001$). Conflation of both Path and Manner components was relatively infrequent and mainly occurred in peripheral devices (36% in French and 14% in English).

3.1.3. Architecture

Figures 9 and 10 show how speakers decided to package spatial information in different types of constructions (TS, TC, LS, and LC) with cartoon and video scenes, respectively. A mixed ANOVA examined the effects of language and core-event-type

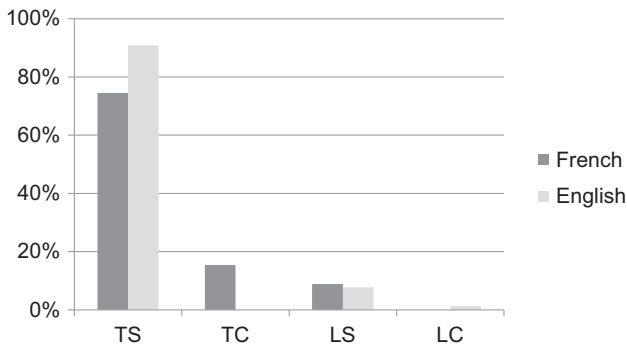


Figure 9. Response architecture with cartoons.

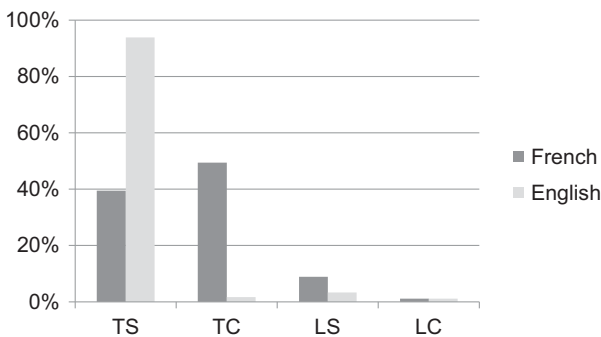


Figure 10. Response architecture with videos.

on TS scores. The results show significant main effects of language (in both cartoons and videos: $F(1,36) = 37.33, p < 0.0001$, and $F(1,36) = 146.96, p < 0.0001$, respectively) and of scene types (cartoons: $F(3,108) = 38.64, p < 0.0001$; videos: $F(5,180) = 20.43, p < 0.0001$). As expected, TS responses were overall significantly more frequent in English (91% and 94%, respectively) than in French (75% and 39%), while TC responses were more frequent in French than in English, but only with the videos (49% vs. 2%). Further analysis revealed language effects with some core-event-types: (a) TS constructions were used in the cartoon set significantly more frequently in English than in French with double-boundary-crossing/*across* items (89% vs. 46%, $F(1,36) = 101.03, p < 0.0001$), but not with *up/down* events; (b) in the video set, significant language effects occurred with all boundary-crossing events: single-boundary crossing such as *into* ($F(1,36) = 73.48, p < 0.0001$) and *out-of* items ($F(1,36) = 103.10, p < 0.0001$) and double-boundary crossings with *across* items ($F(1,36) = 72.00, p < 0.0001$) as well as with *down* events ($F(1,36) = 104.04, p < 0.0001$).

To summarize, although overall compact constructions (TS) were the most frequent across languages and stimuli sets, some variation is noted in motion expression with some scenes (e.g., videos) and core-event-types (e.g., single-boundary-crossing items). More specifically, although English speakers systematically expressed Manner in verbs together with Path in other devices within the same compact structures and across stimuli sets, as in (13a,b), French speakers' motion descriptions followed the prototypical (one lexicalized core component) pattern, focusing on Path, lexicalized in verbs within TS constructions, and omitting Manner information, but only with the cartoon set, as in the example (13c). With the video set, Path primacy was not that evident for French speakers. With these items that involved exclusively natural human voluntary motion events, French speakers tended to add quite frequently Manner information in their utterances by means of other devices and thus organize them within TC constructions and some LS constructions, as in (13d) and (13e), respectively. This phenomenon occurred especially with non-default-Manners (such as run, jump, ride a scooter, ride a bike, roller skate) and especially with one-boundary-crossing events (*into/out-of*), which involved a change of state (invisible/visible figure/agent – cf. Slobin, 2006): an agent enters (appears) and exits (disappears), respectively, in these items, inviting the speaker to specify the particular way (the Manner) this change of state occurs, if necessary/salient, with the addition of some peripheral devices (e.g., *en courant/à toute vitesse* 'running/in all speed', *en sautant/pieds joints* 'by jumping/with joint feet', *à vélo* 'on a bike', *en rollers* 'on roller skates', *avec une trottinette* 'with a scooter') or with the distribution of components across clauses (e.g., coordination). Partly, this tendency to add Manner within the video set can also be explained by the fact that two of the above-mentioned salient Manners (*riding a scooter* and *jumping* events) were present in this set but not in the cartoon set.

To conclude, although these results are in line with the general prediction that language-specific properties affect different aspects of speakers' verbal behavior (*focus, locus, architecture*) supporting the typological asymmetry documented in prior experimental and theoretical studies, the findings further suggest within-language variation in the verb-framed language (French). More specifically, with some scenes that involve human natural motion (videos) and in which the Manner of motion is not the prototypical one for humans (walking) and/or the core event involves a change of state (e.g., single-boundary-crossing events), French invites

speakers to add Manner specifications and organize relevant information in more distributed ways, with TC and LS, as in (13d) and (13e) respectively.

- | | | |
|---------|---|----|
| (13) a. | <i>The mouse is climbing up the leg of the table</i> (cartoon)
[Manner in the verb, Path in a satellite] | TS |
| b. | <i>A girl is jumping up the hill.</i> (video)
[Manner in the verb, Path in a satellite] | TS |
| c. | <i>Une fille traverse les rails</i> (cartoon)
[Path in the verb, no Manner]
'A girl crosses the rails' | TS |
| d. | <i>Un homme rentre dans une pièce en sautant.</i> (video)
[Path in the verb, Manner in a gerund]
'A man enters in a room by jumping' | TC |
| e. | <i>Une femme traverse un chemin et fait du vélo</i> (video)
[Path and Manner separate verbs, coordinated clauses]
'A woman crosses a path and rides a bike' | LS |

3.2. Non-verbal measures

Experiments 1 and 2 involved 14 control items each, in which one of the variants corresponded to a correct answer (either Manner or Path congruent to the target but performed by another figure). The analysis of the control items showed ceiling performance in the responses of the participants allowing to move on to the analysis of the non-verbal measures in the experimental items of these tasks: similarity judgments and RTs.¹⁰

3.2.1. Similarity judgments

The framing event involving Path (considered as the core spatial component, according to Talmy) was expected to be the main similarity judgment criterion in the responses of all participants. Any language difference was expected to emerge with respect to the *non-core* component of events (Manner). Thus, the similarity judgment data were analyzed using mixed ANOVA on Manner responses, including gender (male, female)¹¹ and language (French, English) as across-subject factors and Path-item-type and Manner-item-type as within-subject factors. An additional ANOVA was conducted to evaluate the global sources of variation across tasks with a categorization-type variable (non-verbal/CatNV, verbal/CatV) as an additional within-subject factor. The analysis showed first a significant categorization-type effect, in that Manner choices were significantly more frequent in the verbal task than in the non-verbal one ($F(1,36) = 57.42, p < 0.0001$). With respect to the main

¹⁰Inclusion criterion for this task: less than three errors per participant. None of the participants were excluded from the analysis.

¹¹The results showed no significant gender effect; thus, this factor was discarded from the subsequent analyses.

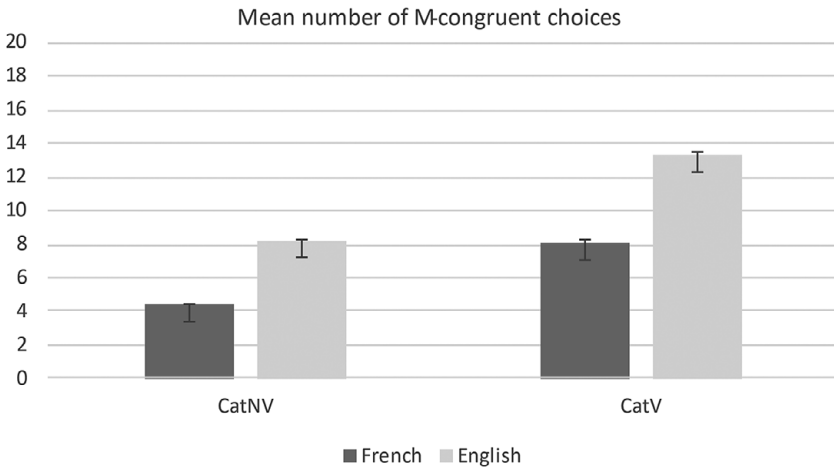


Figure 11. Mean number of Manner choices by French and English participants across categorization tasks. Note: Error bars indicate mean \pm SE.

language factor, all ANOVA revealed a significant language effect across tasks ($F(1,36) = 16.71, p < 0.001$) as well as within each task (non-verbal: $F(1,36) = 7.03, p = 0.01$; verbal task: $F(1,36) = 17.42, p < 0.001$). As expected, Manner-congruent variants were chosen almost twice more frequently by the English participants. More specifically, the mean was calculated by recording the number of Manner matches out of 20 individuals in each group. Figure 11 illustrates the mean number of Manner-congruent choices as performed by French and English participants in the non-verbal and verbal tasks. In the non-verbal categorization task (CatNV/experiment1), although French participants were clearly guided by the Path-congruency in the variants (P congruent: 78%; M congruent: 22% of their choices), the English participants were guided much less by this criterion (P-congruent: 59%; M-congruent responses: 41%), and analogously, Manner choices were significantly more frequent in the English dataset than in the French one. With respect to the similarity judgments in the verbal categorization task (CatV/experiment2), the results show that Manner-congruent choices were selected more frequently in the verbal task than in the non-verbal one by both French ($F(1,18) = 26.14, p < 0.001$) and English ($F(1,18) = 31.30, p < 0.001$) participants; however, French participants continued to show stronger *Path-congruent* preferences and significantly less Manner-guided choices (P: 59% vs. M: 41%) as compared to the English participants (M: 67% vs. P: 33%, ($F(1,36) = 16.71, p < 0.001$).

Further comparisons showed two significant item type effects within and across tasks: an effect of Path-item-type (non-verbal task: $F(3,108) = 7.43, p < 0.001$; verbal: $F(3,108) = 47.82, p < 0.001$; across tasks: $F(3,108) = 27.16, p < 0.001$) and an effect of Manner-item-type (non-verbal task: $F(2,72) = 6.39, p < 0.01$; verbal: $F(2,72) = 18.481, p < 0.001$; across tasks: $F(2,72) = 4.38, p = 0.01$), as well as a significant interaction of these two factors (Path X Manner) in the non-verbal task (CatNV), in the verbal

(CatV), and across tasks (CatNV: $F(6,216) = 5.82, p < 0.001$ CatV: $F(6, 216) = 5.79, p < 0.001$; across tasks: $F(6,216) = 9.60, p < 0.001$). More specifically, with respect to the Manner-item-type effect, in CatNV, Manner-congruent variants were chosen more often when targets involved either Manner-without-instrument (*jump, run*, etc.) or default-Manner (*walk*) as compared to Manner-with-instrument (*cycle, roller skate*, etc.): default-M versus M-with-instr.: $F(1,36) = 5.03, p = 0.02$; M-without-versus M-with-instrument: $F(1,36) = 15.55, p < 0.001$. In the CatV, however, the Manner-congruency criterion was chosen more often when targets involved Manner-with-instrument as compared to the other two item types (M-with- vs. M-without-instrument: $F(1,36) = 14.58, p < 0.001$; M-with- vs. default-M: $F(1,36) = 31.22, p < 0.001$). Further tests carried out within each language showed additional variation stemming from the Manner type of the items. More specifically, although in French a Manner-item-type effect was found both in the non-verbal task ($F(2,38) = 6.36, p < 0.01$) and in the verbal task ($F(2,36) = 7.12, p < 0.01$), in the English responses the Manner-item-type effect appeared only in the verbal one ($F(2,36) = 11.93, p < 0.001$). More specifically, partial comparisons within French responses in the non-verbal task revealed that items that involved either default-Manner or Manner-without-instrument events elicited more Manner choices than motion events that depicted Manner-with-instrument (default-M vs. M-with-instr.: $F(1,18) = 8.56, p < 0.01$; M-without-instr. vs. M-with-instr.: $F(1,18) = 11.37, p < 0.01$). In contrast, in English, only the comparison between Manner-without- and Manner-with-instrument was significant ($F(1,18) = 4.99, p = 0.03$), with Manner-congruent choices being more frequent with the first type (items without instruments) than with the latter (items with instruments), as illustrated in Figure 12. Further tests across languages revealed significant language differences with Manner-with-instrument ($F(1,36) = 10.36, p < 0.01$) and Manner-without-instrument event types ($F(1,36) = 6.58, p = 0.01$) in the non-verbal task and with all Manner types in the verbal task (default-M: $F(1,36) = 9.36, p < 0.01$; M-without-instr.: $F(1,36) = 15.38, p < 0.001$; M-with-instr.: $F(1,36) = 13.15, p < 0.001$).

Further item type comparisons revealed significant differences in the Manner choices of the participants as a function of different Path types. Overall, Manner choices were more frequent with one-boundary-crossing events (into/out-of) and two-boundary-crossing (across) Paths than with default (along) and vertical (up/down) ones. More specifically, and as illustrated in Figure 13, Manner-congruent choices in the non-verbal task were more frequent in French with one- and two-boundary-crossing events as compared to default and vertical Paths (e.g., one-

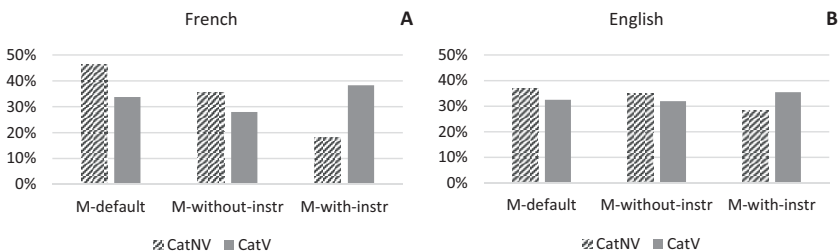


Figure 12. Proportion of Manner choices across different Manner-item-types in French (A) and English (B).

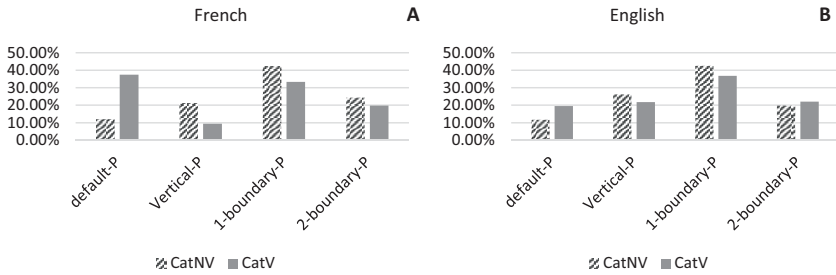


Figure 13. Proportion of Manner choices across different Path-item-types in French (A) and English (B).

boundary-P vs. vertical-P: $F(1,18) = 6.08, p = 0.02$; two-boundary-P vs. vertical-P: $F(1,18) = 5.23, p = 0.03$) with an additional significant increase in Manner choices with default-P (*along*) items (e.g., default-P vs. one-boundary-P: $F(1,18) = 41.41, p < 0.0001$). Partial comparisons in the English responses showed that Manner choices were more frequent when one- or two-boundary-crossing events were involved in the targets, in both tasks, as compared to vertical- and default-Path (*along*) items (e.g., one-boundary-P vs. vertical-P: $F(1,18) = 8.40, p < 0.01$; two-boundary vs. default-Path: $F(1,18) = 4.66, p = 0.04$). Additional tests across languages within specific Path types in the targets showed significant language differences, as well: The Manner-congruency criterion was preferred again significantly more often by the English participants than by the French, especially with vertical ($F(1,36) = 8.02, p < 0.01$) and one-boundary-crossing events ($F(1,36) = 6.69, p = 0.01$) in the non-verbal task, and more broadly with vertical ($F(1,36) = 30.47, p < 0.001$), one-boundary-crossing ($F(1,36) = 15.32, p < 0.001$), and two-boundary-crossing ($F(1,36) = 16.96, p < 0.001$) events in the verbal task.

3.2.2. Reaction times

In experiments 1 and 2, the overall RTs to target video clips were analyzed with a mixed ANOVA with gender (male, female)¹² and language (French, English) as across-subject factors and Path-item-type and Manner-item-type as within-subject factors. An additional ANOVA was conducted to evaluate the global sources of variation across tasks with a categorization-type variable (non-verbal/CatNV, verbal/CatV) as an additional within-subject factor. The analysis showed first a significant categorization-type effect ($F(1,36) = 40.74, p < 0.001$), in that RTs for the selection of Manner-congruent variants were overall significantly longer in the verbal than in the non-verbal task. Figure 14 presents the mean RTs for Manner-congruent choices of participants across tasks (Figure 14). A mixed ANOVA with language as between-subject factor and Manner- and Path-item-types as within-subject factors revealed a

¹²The results showed no significant gender effect; thus, this factor was discarded from the subsequent analyses.

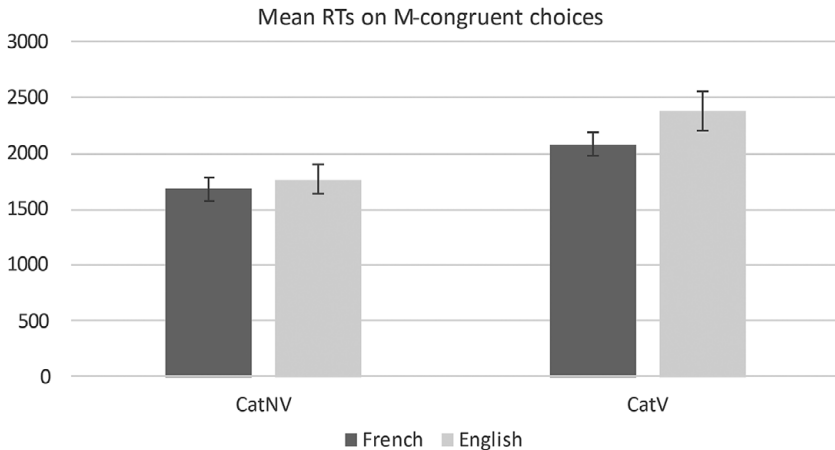


Figure 14. Mean reaction times on M choices of French and English participants across categorization tasks (error bars indicate mean \pm SE).

statistically significant main effect of both Path- and Manner-item-types ($F(2,72) = 19.23, p < 0.0001$; and $F(3,108) = 9.05, p < 0.0001$, respectively), as well as a statistically significant interaction between these two factors ($F(6,216) = 6.28, p < 0.0001$) but no language effect.

With respect to item types, in the non-verbal task, participants took significantly more time to select a variant when exposed to a target item that did not involve any instrument (default-M and M-without-instrument motion) and spent more time with items that involved instruments during the verbal task (task-type effect with default-M targets: $F(1,36) = 30.41, p < 0.001$; M-without-instr.: $F(1,36) = 31.17, p = 0.001$; and M-with-instr.: $F(1,36) = 49.46, p < 0.001$ but no language effect across groups). With respect to the Path-item-type factor, partial comparisons reveal a main Path-item-type effect in the non-verbal task, for both English and French Manner responses ($F(3,54) = 13.16, p < 0.0001$; $F(3,54) = 4.44, p < 0.001$, respectively), but not in the verbal task. Specific comparisons between different item types showed that in the non-verbal task, the items that required more processing time were those involving one-boundary crossing, then those involving vertical motion, and finally those involving either default- or double-boundary crossing, while in the verbal categorization there was no significant variation across Path types. Overall, the analysis shows that differences in the mean RTs of French and English-speaking participants were statistically different across groups only in the verbal task (experiment 2) in that Path-congruent choices took longer time to be selected by English participants as opposed to French (mean difference: 540, $p < 0.0001$), but this difference was only marginal with Manner-congruent variants (mean difference: 110, $p = 0.05$). Figure 15 summarizes the mean RTs of Path- and Manner-congruent choices in the two categorization tasks as performed by French and English participants.

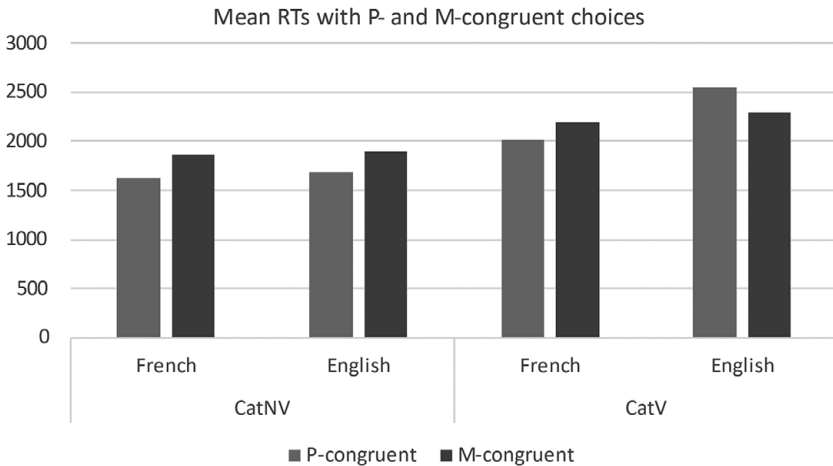


Figure 15. Manner- and Path-congruent choices: reaction times (in msec) of French and English participants across categorization tasks.

3.3. Eye-movements

3.3.1. Production task

The analyses of eye-movements during verbalization (experiment 3) aimed to identify how speakers of typologically different languages not only encode but also allocate their attention to different visual components of motion events (gazes to Manner- and/or Path-related areas, treated as a within-subject factor) during construal selection.¹³

Fixation counts. A mixed ANOVA examined first the effects and interactions of language, core-event-type (up, down, across), and AoI type (P ± M, P, S, G) on the number of fixations with cartoons. The analysis showed significant effects of language ($F(1,36) = 9.04, p < 0.01$) and AoI ($F(4,144) = 203.25, p < 0.0001$), as well as an interaction between these two factors ($F(4,144) = 12.45, p < 0.0001$). Further analysis showed that French participants fixated P broad areas (gazes on P + S + G) significantly more often than the English group ($F(1,36) = 16.43, p < 0.0001$). In contrast, Manner (P ± M) fixations showed no significant language difference (Figure 16).

Although no general effect of core-event-type was found in the cartoon set, this factor had a significant impact in relation to specific Manner and Path components involved in the stimuli. In particular, core-event-type had an impact on both P broad and P ± M fixations ($F(2,72) = 52.38, p < 0.0001$, and $F(2,72) = 23.17, p < 0.0001$, respectively), in that Manner fixations were more frequent with double-boundary-crossing (across) events, while Path fixations were more frequent with vertical ones (up and down). Nevertheless, fixations to different core-event-types did not differ significantly across the two language groups. With respect to the AoI effect, separate

¹³Fixation counts and durations at specific areas of interest (AoI) were calculated in relation to total scene viewing times (and are therefore relative) as opposed to the absolute scores presented in the qualitative gazeplot analysis (see Figures 20–23). The results concern eye-movements as recorded during the exploration of video and cartoon displays, before verbalization.

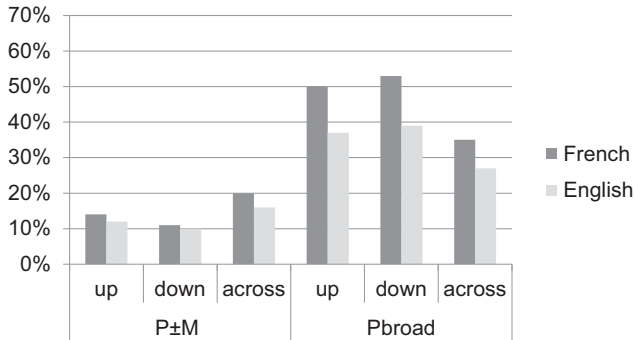


Figure 16. Number of fixations to P ± M and Pbroad AoI with different item types in cartoon scenes.

contrasts show that fixations were significantly more frequent on P-AoI as compared to the other AoI, as summarized below for cartoons:

Fixation counts in cartoons : $P > P \pm M > G = S$

A similar mixed ANOVA was carried out for video scenes. The analysis revealed again a main significant effect of language ($F(1,36) = 8.51, p < 0.01$) and of AoI ($F(4,144) = 63,189053, p < 0.0001$), as well as an interaction between these two factors ($F(4,144) = 5.69, p < 0.0001$). Fixations were more frequent on both P-, P ± M-, and G-AoI as compared to the S-AoI. Fixation counts can be summarized as follows for videos:

Fixation count in videos : $P = P \pm M = G > S$

Although with video scenes both language groups fixated Pbroad areas significantly more than P ± M areas (French: $F(1,18) = 8.78, p < 0.01$; English: $F(1,18) = 15.32, p < 0.01$), French participants fixated Pbroad areas significantly more than the English group ($F(1,36) = 4.76, p = 0.03$). P ± M fixations showed again no significant language difference. Although further analyses on fixation counts to Path and Manner areas revealed no global effect of core-event-type, this factor had a significant impact on specific AoI. As illustrated in Figure 17, this factor had an effect on both P ± M and Pbroad fixations ($F(4,144) = 15.59, p < 0.0001$; and $F(4,144) = 5.48, p < 0.001$, respectively) for both language groups: P ± M fixations were more frequent with *vertical* (no boundary) events (*up* and *down*); Pbroad fixations were more frequent with boundary-crossing events (*into/out-of* and *across*).

To summarize, fixation counts varied as a function of AoI and of Path types in the videos and, in some cases, as a function of language group. Overall, fixations occurred more frequently on Path areas, particularly with vertical motion (both *up* and *down*) events, than on Manner areas, which were mostly associated with *across* in the cartoon set, as well as with *across* and *down* events in the video set. Surprisingly, the language effect initially predicted in relation to the salience of Manner (following the *Manner cline* proposed by Slobin) did not occur. P ± M fixations did not differ between English and French viewers. The only language effect observed concerned

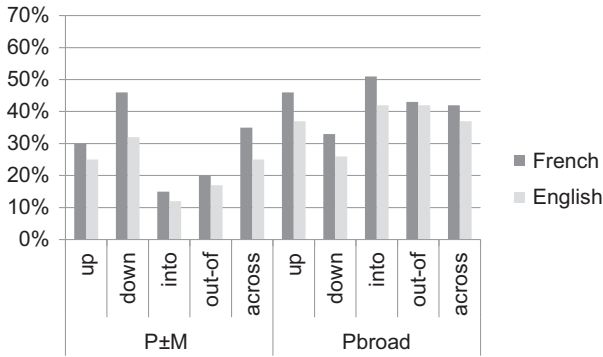


Figure 17. Number of fixations to P ± M and Pbroad Aol with different item types in video scenes.

Path areas, in that French viewers fixated more frequently Path areas, as opposed to the English group, who did so to a lesser extent.

Duration of fixations. The analysis of fixation lengths examined the time spent on different visual components as a function of language groups and event types. In both language groups’ explorations, fixations were longer on P than on the other AoI. Fixation durations with cartoon scenes can be summarized as follows:

$$\text{Duration of fixations in cartoons : } P > P \pm M = G = S$$

Specific comparisons showed that fixations were overall longer on Path areas than on Manner areas in both languages (Pbroad vs. P ± M: $F(1,36) = 6,645783, p = 0.013$) and that the duration of fixations (whether on Manner or on Path areas) did not vary across language groups. Figure 18 shows the proportions of fixation times spent on P ± M and Pbroad areas with cartoons. Further analysis revealed no global effect of core-event-type, yet a significant impact of this factor on specific areas, particularly on fixations to both P ± M and Pbroad areas ($F(2,72) = 16.96, p < 0.001$; and $F(2,72) = 21.00, p < 0.001$, respectively). Overall, Manner fixations were longer with across items and Path fixations with down items.

Although the analysis of the video scenes revealed no significant language effect with respect to fixation durations, a significant main effect of AoI was found

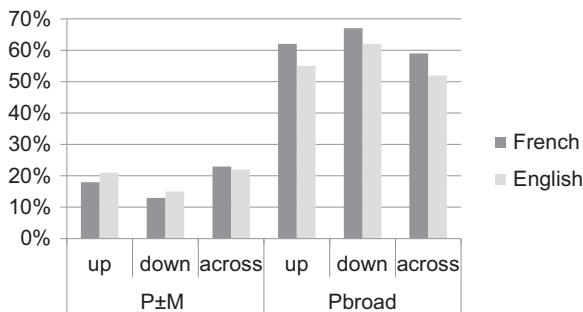


Figure 18. Fixation lengths on P ± M and Pbroad Aol with different item types in cartoon scenes.

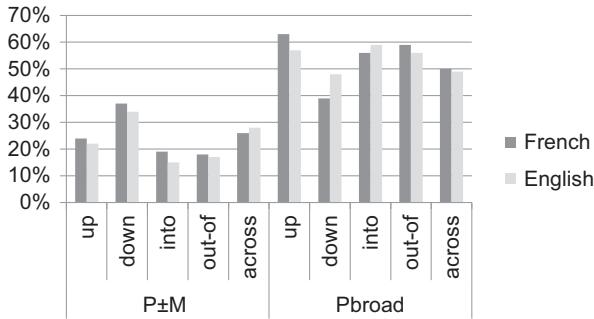


Figure 19. Fixation lengths on P ± M and Pbroad AoI with different item types in video scenes.

($F(4,144) = 20.93, p < 0.001$), as well as an interaction between language and AoI type ($F(4,144) = 2.42, p = 0.05$). More specifically, fixations were longer on Path (P), Manner (P ± M), and Goal (G) as compared to Source (S) areas, while durations were not significantly different for P ± M-, P-, and G-AoI. The results for the duration of fixations in the videos can be summarized as follows:

Duration of fixations in the videoset : $P = P \pm M = G > S$

As shown in Figure 19, further analyses focusing on the length of fixations to the main Manner (P ± M) and Path areas (Pbroad) showed again that, overall in both language groups, fixations were longer on Path than on Manner areas (Pbroad vs. P ± M: $F(1,36) = 19.41, p < 0.001$). Furthermore, within each language, fixations lasted significantly longer on Path AoI than on Manner AoI during both French ($F(1,18) = 12.10, p < 0.01$) and English ($F(1,18) = 7.65, p = 0.012$) explorations. Finally, although the analysis revealed no significant effect of core-event-type, an effect of this factor was found in relation to specific AoI in that Manner fixations were longer with *down* items and Path fixations with *up*, *out-of*, and *into* events.

In sum, the duration of fixations to Path and Manner areas did not depend on language, but varied as a function of Path types in the targets. Overall, fixations lasted longer on Path than on Manner areas, especially with boundary crossing and upward motion events, while the longest Manner fixations occurred during the processing of *across* events in the cartoons and of *down* events in the video set. The data do not support the initial prediction according to which Manner fixations were expected to be longer for English than for French viewers.

Gazeplots. A descriptive analysis examined qualitatively how fixations were distributed, on-line, during participants' visual exploration of events. As shown in Figures 20 and 21, differences across language groups were observed in the number, order (numbered gaze points), and duration (size of gaze points) of fixations. For example, for the same *upward* event (cartoon), French fixations were found to be 'ballistic', going back and forth from S to G areas several times, performing large amplitude saccades (Figure 20, example on the left) in several steps, as opposed to a more minimalist, sequential pattern traced by English viewers who followed the figure's motion step by step, in a linear way (right).

Figure 21 illustrates how French and English viewers allocated their attention through an 'out-of' (video) item. This is a spatial depiction of where participants

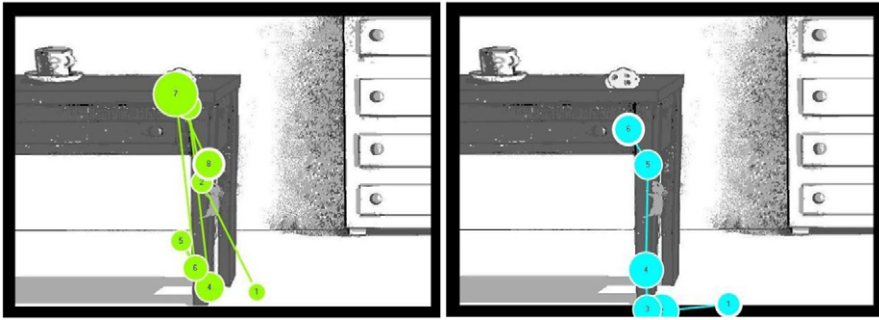


Figure 20. Scene exploration in the production task – ‘climb/up’ event in the cartoon set: ballistic exploration by the French viewers (*left*) and linear exploration by the English viewers (*right*).

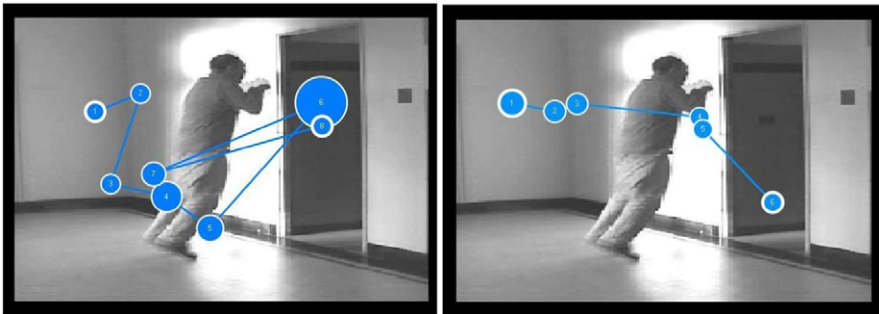


Figure 21. Scene exploration in the production task – ‘jump/out-of’ event in the video set: ballistic exploration by the French viewers (*left*) and linear exploration by the English viewers (*right*).

fixated their gaze and how they navigated through the video stimulus revealing hotspots and the same gaze pattern distribution observed previously: a ‘ballistic’ way of processing by the French viewers characterized by great amplitude in the gaze saccades that drive fixations (*left*) and a rather linear processing of the same event by the English viewers with shorter saccades and less fixations during scanning (*right*).

3.3.2 Categorization tasks

In experiments 1 and 2, two additional AoI were defined: M-AoI corresponding to the Manner-congruent variant presented after the target video and P-AoI corresponding to the Path-congruent variant the participants were shown. All other fixations fell to a general area x that covered the rest of the (blank) screen display.

Numbers of fixations. Mixed ANOVA were conducted on raw numbers of fixations including gender¹⁴ and language as across-subject factors and Path type, Manner type, AoI type, and categorization type as within-subject factors, once for the non-verbal task, once for the verbal task, and once to compare the two. First, the analysis revealed a main AoI effect, significant both within and across tasks (CatNV:

¹⁴Gender did not have any significant effect ($p > 0.05$) and was therefore disregarded in subsequent analyses.

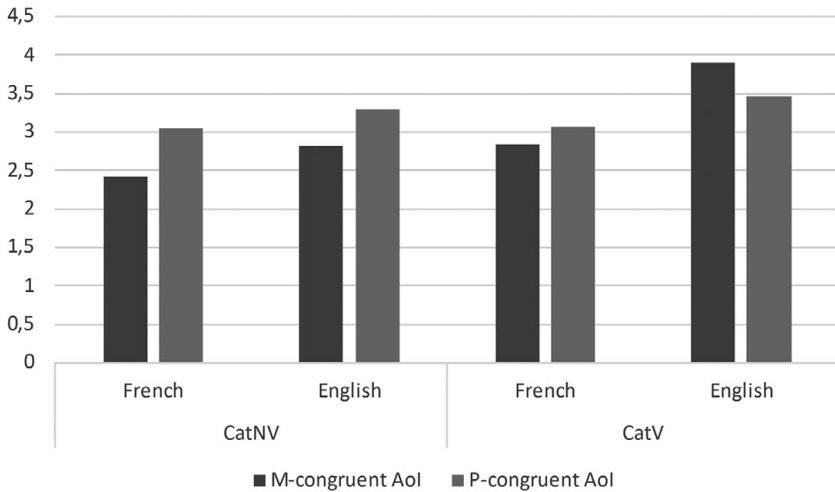


Figure 22. Mean fixation numbers on Manner- and Path-congruent areas by French and English viewers across categorization tasks.

$F(4,144) = 372.51, p < 0.001$; CatV: $F(4,144) = 598.97, p < 0.001$; across tasks: $F(4,144) = 694.74, p < 0.001$), and a Path-type effect, significant only in the non-verbal task ($F(3,108) = 5.40, p < 0.01$) and across tasks ($F(3,108) = 5.50, p < 0.01$), but no Manner type or language effect. However, the language factor was found to interact with the AoI factor in the verbal task ($F(4,144) = 7.17, p < 0.0001$) and across tasks ($F(4,144) = 4.56, p = 0.001$), as well as with the Manner-type factor in the non-verbal experiment ($F(2,72) = 6.75, p < 0.01$) and across tasks ($F(2,72) = 7.50, p < 0.01$). More specifically, overall participants fixated Path-congruent variants more frequently than Manner-congruent variants in the non-verbal experiment (P vs. M : $F(1,36) = 37.25, p < 0.001$), but not in the verbal one, during which English participants fixated more the Manner-congruent variants than the Path-congruent ones, and this is significantly more than the French viewers (Figure 22). With respect to the categorization-type factor, the analysis showed a significant interaction between categorization types and AoI types on the number of fixations to each specific AoI (task effect on M -AoI: $F(1,36) = 18.50, p < 0.001$; and on P -AoI: $F(1,36) = 60.13, p < 0.001$), in that there were overall significantly more fixations in the verbal task as opposed to the non-verbal one, suggesting the difficulty the viewers had in this task to make a choice between M- and P-congruent variants.

Further analysis on specific item types showed no significant Manner effect, either within or across tasks, but a significant global effect of Path type in the non-verbal task ($F(3,108) = 5.40, p < 0.01$), as well as across tasks ($F(3,108) = 5.50, p < 0.01$). The analysis also revealed a significant AoI-type effect on each individual core-event-type in both tasks: in the non-verbal categorization (default-P: $F(4,144) = 165.77, p < 0.001$; vertical-P: $F(4,144) = 284.96, p < 0.001$; one-boundary-crossing-P: $F(4,144) = 181.99, p < 0.001$; two-boundary-crossing-P: $F(4,144) = 149.11, p < 0.001$) as well as in the verbal categorization (default-P: $F(4,144) = 358.30, p < 0.001$; on vertical-P: $F(4,144) = 316.76, p < 0.001$; on one-boundary-crossing-P: $F(4,144) = 343.40, p < 0.001$; on two-boundary-crossing-P: $F(4,144) = 115.33, p < 0.001$). More specifically, Path fixations were found to be more frequent with vertical (*up*, *down*) and

one-boundary-crossing (*into*, *out-of*) events, while Manner fixations were more frequent with default (*along*) and double-boundary-crossing (*across*) events, in both tasks. However, further tests exploring how different groups allocated their attention with different core-event-types in the targets revealed no significant language effect.

Fixation lengths. Similar mixed ANOVA were conducted on *fixations' duration*, showing no significant effects of gender, Path-, or Manner-item-type in either task; thus, these factors were discarded from the following discussion of the results. With respect to the AoI-type and categorization-type factor, the durations of the fixations were found to vary significantly depending on the variant involved both within and across tasks (in the non-verbal task: $F(4,144) = 567.78, p < 0.001$; in the verbal task: $F(4,144) = 775.55, p < 0.001$; and across tasks: $F(4,144) = 1028.73, p < 0.001$). In addition, the AoI-type factor interacted with the language factor, again within and across tasks (marginally in the CatNV: $F(4,144) = 2.38, p = 0.05$; significant in the CatV: $F(4,144) = 9.88, p < 0.0001$; and across tasks: $F(4,144) = 7.91, p < 0.0001$), in that in both tasks, fixations were overall longer for Path-congruent areas than for Manner-congruent areas (P vs. M in CatNV: $F(1,36) = 9.11, p < 0.01$; and in CatV $F(1,36) = 15.95, p < 0.001$, respectively), with French viewers spending more time exploring P-congruent video variants in both categorization tasks (experiments 1 and 2), as illustrated in Figure 23.

Gazeplots. A descriptive analysis examined qualitatively how fixations were distributed on-line during participants' visual exploration during the exploration of target videos in the categorization tasks. As shown in Figure 24, the same differences observed in the production task (experiment 2) were also observed in the exploration of the target events during experiments 1 and 2. For example, for the same *out-of* event, French viewers fixated the scene in a 'ballistic' way, going back and forth, from S to G areas several times, as opposed to the linear pattern of the English fixations that followed the figure's motion step by step.

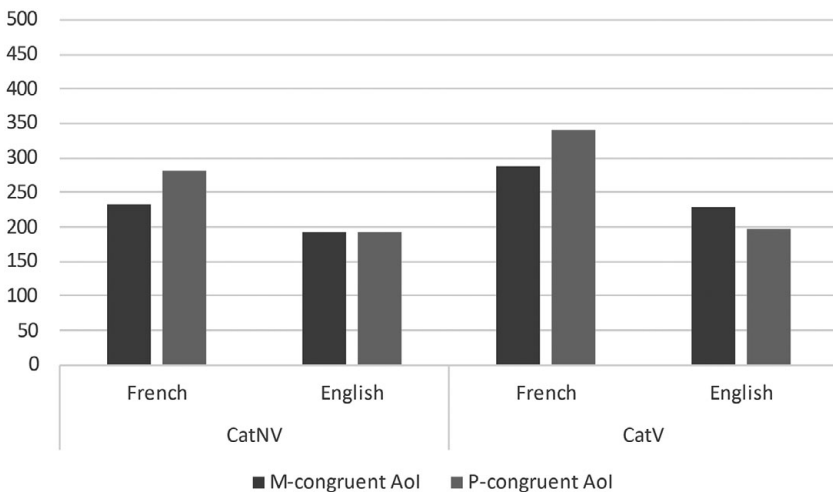


Figure 23. Mean fixation lengths (msec) on M- and P-congruent areas by French and English viewers across categorization tasks.



Figure 24. Target scene exploration during non-verbal categorization: ballistic exploration by the French viewers (*left*) and linear exploration by the English viewers (*right*).

4. General discussion

The current study investigated how speakers from two typologically different languages, English and French, conceptualize and represent voluntary motion events in tasks that involved language to different degrees (non-verbal categorization, verbal categorization, production) and in varied event and scene types. Two main questions were addressed: (i) whether the differences of these two contrasted languages constrain participants' verbal performance in specific ways according to different contexts (varied event saliency and scene nature contexts) and (ii) whether cognitive influences (during attention allocation and decision-making) arise both when language is and is not explicitly involved. A first null hypothesis predicted no major language effects of any kind, whereas a second typological hypothesis predicted language effects to various degrees (either in all measures or at least during verbalization). Overall, the four sets of analysis (similarity judgment data, RT data, production data, and eye-movement data) offer the possibility for a deep investigation and a fine-grained account of the impact language may have on non-verbal processing.

4.1. Effects with maximal involvement of language

With respect to the production data, based on two different scene settings (video clips and animated cartoons) involving different types of voluntary motion events (boundary crossing, vertical motion events), participants' verbal descriptions and their on-line visual strategies during preparation for verbalization were examined in order to determine whether the verbal and non-verbal behavior of speakers differs as a function of their linguistic background and, if so, with which scene contexts and item types. As expected by the typological approach, speakers' verbalizations differed as a function of their language type. Responses were overall semantically denser in English than in French, and this difference was much more striking with boundary-crossing events than with other item types. More specifically, English speakers systematically combined Manner (encoded in the verb) and Path (in other devices) following the typical Satellite-framed pattern of their language (PM conflation), while French speakers followed the typical verb-framed pattern, focusing mostly on Path-alone (lexicalized in the verb) constructions, notwithstanding frequent PM combinations in some contexts (e.g., with upward motion in cartoons; with single-boundary-crossing events in

videos involving a change of state). The results concerning response architecture showed that TS responses were predominant across groups with the cartoon set, but not with the video set. Video scenes elicited the expression of both Manner and Path components, thus inviting French speakers, for whom Manner is normally optional, to produce TC utterances (typically with the addition of gerunds) or LC utterances (typically distributing information in coordinated or juxtaposed clauses). Additional language-specific strategies emerged when scenes involved highly salient components that are otherwise downplayed, such as Manner for French. For example, a video showing a figure *riding a bicycle* or *riding a scooter uphill* made Manner much more salient than is typically perceived by French speakers. These participants therefore paid more attention to this component than they would do habitually, expressing Path in a main clause and Manner in the periphery using complex structures involving a dependent element (TC). These data support both the Manner salience and Path salience hypotheses, additionally underlying the need to combine lexicalization and event integration theories, in the Talmyan sense, when studying language asymmetries of this type. Indeed, Manner was found to be more salient and central in English and Path more salient in French, but the way these components were combined or distributed (combined with the use of other peripheral devices or lexicalized in separate clauses) reveals the bigger picture of how language constrains encoding patterns across different dimensions. For example, English invites the expression of both Manner and Path within compact and dense constructions. In comparison, when French speakers want to express Manner in addition to Path, they often have to do it periphrastically adopting complex or loose constructions.

The eye-tracking data provided additional information concerning the on-line processing of these events as participants were preparing their verbalizations. As predicted by the typological hypothesis, the number of fixations showed that, depending on the language group, spatial information is not only encoded differently, but also partially filtered visually in different ways. The frequency of fixations, but not their duration, differed across groups as a function of language background. In line with predictions based on Talmy (2000) and Ibarretxe-Antuñano (2009), this language effect stemmed from a difference in the Path salience of the events and not from salience in Manner (cf. Slobin, 2006). In addition, although both groups fixated Path more than Manner (especially in the absence of intrinsic boundaries), French viewers fixated more frequently Path areas than the English group, especially with the cartoon scenes. Surprisingly, the duration of fixations did not depend on language, but rather on the motion components that were visually displayed, and on the specific Paths involved in the stimuli: fixations were significantly longer on Path areas, particularly with vertical motion (*up*, *down* events), and this difference was again even more striking in the cartoon set. Finally, in both languages, *gazeplots* also showed a general preference for Path over Manner fixations during participants' exploration of events, while also providing further information about the on-line distribution of attention allocation during the visual scanning of motion: ballistic fixations by the French group as compared to the linear exploration strategy of the English group.

4.2. Effects with minimal or no involvement of language

The analysis of the two categorization tasks (experiments 1 and 2) took into account the degree of involvement of language (minimal verbal input involvement with the

presented audio sentences in the verbal categorization task – experiment 2 and no involvement of language in either the input or the output of the non-verbal categorization task – experiment 1). Overall, both the off-line and on-line data support the Path salience hypothesis according to which Path, as the core (framing) event, attracts the attention of all groups but constrains even more the mechanism of perceptual processing when the viewers come from a Path-oriented (verb-framed) language such as French. The responses to the experimental items showed indeed that, when categorizing events, participants had a general preference for the framing event (the core/universal component: Path) and, as expected, this criterion was the main categorization criterion across groups. However, with respect to the Manner-congruent responses, the results show a clear language effect in that when Manner-congruent variants are selected, they are selected twice as much time as French by the English participants in the verbal task, and this preference is even more striking with items involving boundary-crossing contexts and instruments. This language effect was clear in the similarity judgment measure of the verbal task but not significant in the RT measure.

More specifically, given the systematic encoding of both Manner and Path in (satellite-framed) English, it was expected that the speakers of this language would take more time to make their (P- vs. M-congruency) decisions, as opposed to French, who were expected to focus mostly on Path-congruency and thus take less time to decide which component is most important to them. Despite the general tendency in the similarity judgments to select P-congruency over Manner-congruency across groups (in line with previous observations by Ji, 2017) and a significant task effect (categorization-type factor), such that participants took longer to respond in the verbal task than in the non-verbal task, no language effect was found in the time spent from stimulus onset until categorization. However, depending on the specific Manner or Path information involved in the stimuli, participants spent more or less time processing incoming information before making their categorical choices, but again they did not differ in this respect across language groups. Specific item type analyses revealed that RTs were longer with items involving instruments and one-boundary crossing, as opposed to simple no instrument, default for humans, displacements (*running/jumping* or *walking*), and double-boundary crossing, respectively.

The analysis of the eye-fixations during similarity judgments showed that motion components attracted the visual attention of participants to different degrees and that the number of fixations to specific AoI depended not only on the type of the task (verbal/non-verbal) but also on their language background. Overall participants fixated Path more frequently than Manner variants in the non-verbal task, but not in the verbal one, during which English participants fixated more the Manner-congruent variants than the Path-congruent ones, and this is significantly more than the French viewers. With respect to the global categorization-type factor, the analysis showed a significant interaction between categorization and AoI types on the number of fixations in that there were significantly more Manner and more Path fixations in the verbal task, as opposed to the non-verbal one, suggesting the difficulty the viewers had in this task to make a choice. Further analysis on specific item types showed no significant Manner-type effect either within or across tasks, but a significant effect of Path type in the non-verbal categorization task as well as across tasks. More specifically, Path fixations were more frequent with vertical (up, down) and one-boundary-crossing (into, out-of) events, while Manner fixations were more frequent with default (along) and double-boundary-crossing events, in both tasks. However,

further tests exploring how different groups allocated their attention with different event types in the targets revealed no significant language effects.

With respect to fixation lengths, the durations of the fixations were found to vary significantly depending on the variant involved, both within and across tasks. More specifically, in both tasks, fixations were overall longer for Path-congruent areas than for Manner-congruent areas, especially for the French viewers who spent more time exploring the two video variants, but also remained mainly focused on the P-congruent ones across tasks.

Finally, the gazeplots revealed once again a ‘ballistic’ way of processing visual information by French viewers as opposed to the more linear pattern of English viewers, who tended to follow the figure’s progression step by step. This observation most probably relates to the linguistic properties of the two languages and their implications not only for the expression of spatio-temporal boundaries but also for the visual attention patterns the viewers adopt, more generally. For example, the ballistic eye exploration of French viewers could be explained in part by the optionality of Manner in this language. That is, French participants had to decide for a given motion event whether it was Manner of motion worthwhile to be selected or Path. To make their decision, French viewers first inspected the Path (core component/essential for verbal production) but then went back to check the motion again, even when verbal encoding was not expected in the task.

Apart from the differences in lexicalization patterns, the variation in gaze patterns could also stem from the specificities of these two languages in terms of aspectual features (cf. Aske, 1989; Demagny, 2015; Riegel et al., 1999). For example, in English, the systematic encoding of both Manner (in the verb) and Path (in other devices) invites speakers to focus their visual attention on both spatial components. In addition, the existence of a grammaticalized progressive marker in English (but not in French) enables them to express explicitly ongoing events and thus focus on the progressive development of a displacement. As a result, English participants tend to adopt a linear strategy whereby a motion event is viewed as a sequence of successive points. In comparison, French participants typically focused on Paths, especially the median and final parts of Path (systematically lexicalized in verbs when verbal encoding is at play, downplaying Manner or not expressing this component at all), which explains to some extent their use of ballistic exploration strategies even when verbal encoding is not at play.

Finally, this preference for ‘linear’ versus ‘ballistic’ event exploration may stem from other, psychophysical parameters, such as the general tendency of humans (but also of some nonhuman primates) to explore events following a more or less ‘focal’ versus ‘ambient’ visual strategy, and more specifically to be interested in Endpoints and time-to-contact (TTC) estimations, especially in complex situations of moving objects and agents (for an extended review on the processes involved during TTC estimation and ambient/focal visual exploration strategies, see Bennett et al. (2010), DeLucia and Lidell (1998), Tresilian (1995), Helo et al. (2016), and Pannasch et al. (2011), among others).¹⁵ The qualitative difference observed in the gazeplot data may stem from the fact that French and English viewers are not equally sensitive to the bigger picture/the aim of the event, or at least do not make the same effort to estimate

¹⁵For age and modality differences during visual processing and TTC, see also Keshavarz et al. (2017), Luna et al. (2008), and Helo et al. (2016).

TTC. French viewers have a greater tendency for extrapolation and scan events following an ambient strategy that involves saccades with larger amplitudes and fixations that help them estimate the distance and the velocity of moving figures, from initial (Source) to final (Goal/Endpoint) parts, back and forth several times, most probably driven by the Endpoint/contact- or boundary-crossing point of the event, which seems more salient to them than to English viewers.

4.3. Variability during event processing and theoretical implications

The results only partially support the initial predictions (e.g., those from the linguistic/typological approach). Despite some general language effects, important variations were observed depending on the types of scene sets, event types, and measures. For example, in the production data, language effects were less strong with videos and with events involving intrinsic boundaries. In the on-line data, the frequency of fixations and the gazeplots suggest language effects that stem from the degrees to which participants focus on Path (and its subcomponents) across languages (but not on Manner). Additionally, and contrary to the initial expectations, the results of fixation lengths and RTs during categorization suggest that the timing of cognitive processing during visual perception and decision-making is language-independent.

Several factors could account for the complex patterns observed in participants' verbalizations in comparison with their visual attention and decision-making. First, *general cognitive factors* may account for some of the similarities that were found across the two language groups. In particular, notwithstanding differences in perceptual focus across languages, both groups overall encoded Path in their utterances and paid much visual attention to Path (Pbroad AoI) as compared to areas that also included Manner ($P \pm M$).¹⁶ This pattern was not expected by any version of the typological hypothesis, which predicted generally more visual attention to Path by the French group and either more visual attention to Manner or equal attention to Path and Manner by the English group. This result may not be surprising if we assume that Path is the most basic semantic and cognitive component of motion, determining for example details about the changes of location of protagonists that are essential to construe the event perceptually and to reconstruct it in discourse maintaining the framing event as the core element of encoding (see also Talmy, 2000, about the universal aspects of Path).

Second, *language-specific factors* invited speakers to pay attention to different components of motion, showing overall a clearer focus on Path as compared to Manner in the French group than in the English group. Given the properties of English and French, all versions of the typological hypothesis predict that the most accessible components should differ across languages. Furthermore, at least one version of this hypothesis predicts that such language-specific factors should also (at least partially) affect speakers' attention allocation. This language effect was clearer in the quantitative analyses of fixation counts that showed the different degrees to which viewers allocated their attention to Path. With respect to similarity judgments, the language effect stemmed from the different degree of Manner

¹⁶Recall that the *Pbroad* areas included *P* (intermediate part), *S* (initial part), and *G* (final part) fixations and that the $P \pm M$ areas were mixed so that fixations on these areas involved also some Path fixations even when the main focus was on the Manner in which the limbs moved.

salience, which led the English participants to choose this criterion twice as much as the French ones.

The results of this study also indicate the impact of a third constraint related to *perceptual factors*. In particular, subjects' attention allocation differed to some extent across the two sets of scenes, as well as across item types within each set (e.g., *into/out-of* for videos, *up/down* for cartoons). These differences raise some questions concerning the perceptual features of such visual stimuli. The video set provided highly controlled natural scenes of human voluntary displacements in fixed settings (interior/exterior doors, hill, road), while the cartoon set provided more varied situations of voluntary motion involving different types of figures (humans and animals) in diverse settings (e.g., house, plant, tree, lake, road, river), all contextualized with elements motivating the aim of the motion event (e.g., climbing up a table/tree to get food). Thus, cartoon scenes invited speakers to organize their responses in the form of a short narrative, as opposed to videos that elicited rather short responses merely describing main spatial components (Manner/Path). As a result, when planning their verbalizations but also when attending to specific components in these scenes for the purposes of categorization, speakers/viewers adopted different perceptual strategies: more ambient by the French viewers and more focal by the English viewers – a difference that was more striking in the cartoon set than in the video set (for a similar discussion, see also Soroli et al., 2019).

In summary, the findings show a complex interrelationship between language, perception, and action (see also Spivey, 2023 for a recent discussion). In some cases, language knowledge clearly guided visual processing and consequently – at least in strongly verbal contexts – decision-making (action decisions); in others, language experience only partially influenced cognitive processing. This work suggests that cognitive processing (perception, reasoning, decision-making), including language processing, depends in part on general genetic/universal determinants (as reflected in some systematic language-independent choices – e.g., overall preference for the expression or the selection of the core spatial component (Path), compact TS encodings) and in part on acquired/epigenetic learning mechanisms that arise from our interactions with the world (e.g., experience with specific perceptual features, linguistic properties, action affordances).

Such an interactive perspective is in line with current dynamical system conceptions of language that consider it as a complex and open interactive network in which variability is an inherent property (e.g., De Bot et al., 2007; Hotton & Yoshimi, 2011; Larsen-Freeman & Cameron, 2008). More specifically, in the domain of spatial language and cognition, the present work can contribute to such dynamical accounts by defining the necessary scientific concepts that capture the specific semantic and syntactic properties of the involved meaningful units (e.g., *focus* to specific spatial components), the general constraints of the system in terms of lexical/grammatical distribution (locus), as well as the laws of their combinatorial assemblies (main architectural patterns) in order to better understand the individual interactions at the perceptual (visual processing) and motor action (decision-making) levels. Developing further a formalized dynamical approach to language–thought interactions that takes into account the relative weight of perceptual, linguistic, and action mechanisms involved in spatial cognition may improve current integrated accounts about human behavior.

From the above, it is obvious that future research on the constraints involved in the conceptualization and representation of motion must further examine variation across

and within languages or language types as a function of several other extralinguistic factors. General/cognitive factors as well as typological and perceptual factors (related to event features, scene settings, etc.) all seem to play an important role not only at the formulation level but also at the level of our internal representations.

4.4. *Toward a non-modular interactive account*

During the past decades, the general paradigm of cognitive science has been switching from (linear or cascade) modular conceptions of cognition toward models that integrate body/mind, language, and environment interactions, thus allowing progressively for contextual (top-down) influences to become operative the moment they are made available in thought through the sensory input (e.g., McClelland, 1996; Ryskin & Fang, 2021; Spivey, 2007). In this work, the aim was to focus on voluntary motion event processing as perceived in the form of visual and auditory sensory input, and let speakers/viewers from different language backgrounds (English- and French-speaking) (re)act through active/overt (e.g., button selection, verbal production) and covert inferences (e.g., eye-movements, RTs). By manipulating the visual contexts (the degree of saliency of specific spatial components, the nature of human/animal body affordances, and the naturalism of event execution) and by introducing controlled speech streams for explicit verbal bias, when necessary, the effects of visual perception and language to overt behavior (verbal encodings, similarity decisions) and sometimes even in covert behavior (attention allocation patterns) emerged. The focus was on events that involve representations of spatial relations, e.g., Manner, Path, Goal/Endpoints, Source, etc. through specific semantic focus in construal selection, distributed in meaningful units that take the form of verbs, adverbs, prefixes, particles, prepositional phrases, etc. (locus), and further organized into semantico-syntactic constructions – architectural chains of form, meaning, and function (cf. Talmy, 2005).

The integrated approach supported by the findings, and schematically presented in [Figure 25](#), aims to describe possible language and cognitive interactions in the domain of motion events. It is an illustration inspired by previous situated frameworks (such as the action perception theory by Pulvermüller (2013), the complex and dynamic systems theory by De Bot (2017) among others) that identify processes and active interactions between the perceptual, action, and language domains. Such a framework has the advantage to allow, for example, spatial semantic mechanisms (e.g., spatial component selection, conflation) to be effective within the language strand but also to be influenced, in an interactive way, by other processing components related to perceptual or action generation mechanisms (e.g., the number and complexity of the sensory properties one has to process/extract, the saliency of the involved components, the naturalistic nature of a scene, the demands of the task).

The schematic diagram proposed in [Figure 25](#) illustrates a dynamic account of event processing, which takes into account different levels of variation. In this dynamic framework, the cognitive system is viewed as able to reshape so that it is temporarily dedicated to one type of task or another, depending on the available context. For example, the system can function mostly as a sensorimotor processing system by boosting the coupling of perception and action processes leaving the coupling with the language processor partially or completely inactive (inactivation is marked in gray), for instance, in cases when language is not explicitly involved or

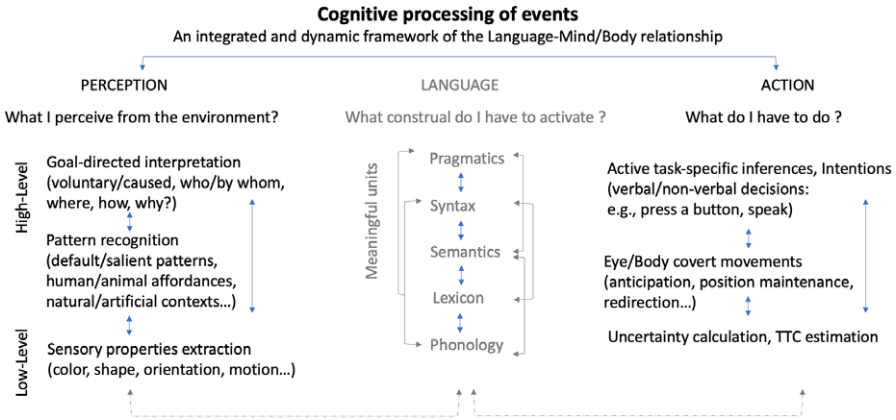


Figure 25. A dynamic framework of the language–body/mind interaction.

necessary (e.g., simple reflex sensorimotor actions such as removing our hand from a heating surface, juggling to grasp a slippery object). In such cases, the system is expected to carry out an action (e.g., juggling to grasp a slippery bar of soap in the shower, as described by Gibbs & Van Orden, 2003). In such contexts, the precise movements are task-specific and active (e.g., bringing the soap back under control), not necessarily anticipated before they are enacted. A covert reaction or a movement (e.g., a rapid saccadic change, juggling) emerges because the situation to be controlled changes. However, in intentional contexts, when the situation changes rapidly and involves complex events that combine core and co-event information as in the case of event categorization tasks described in the present study, language gets back to the game and acts as a support, especially for ambiguous/not sufficiently salient situations when viewers are expected to compare subtle differences, e.g., between default and non-salient Manners of motion without instrument (walking vs. running), and make a decision (similarity judgment). In this sense, a rapidly changing situation, as the one exposed the participants of this study in the categorization tasks, creates the intention to adapt the non-verbal behavior by choosing unique event-specific responses – an adaptation that becomes trickier in complex/non-salient situations and when verbal input is not directly available. In other words, special circumstances (perceptual salience, knowledge of affordances, linguistic knowledge, etc.) vary in their degree of involvement and lead humans to unique inferences, allowing their system to dynamically and flexibly adjust to sensory-, task-, and/or language-specific requirements during action planning.

5. Conclusion

The present paper combined *off-line* and *on-line* measures across two typologically different languages, English and French (satellite- vs. verb-framed). The aim was to explore the language–thought relationship in two ways: (a) to test various hypotheses concerning the presence or absence of language effects on how speakers conceptualize motion events in verbal and non-verbal behavior and (b) to identify how different situations related to event and scene properties contribute to this process.

The findings show clear differences in speakers' linguistic representations that follow directly from the typological properties of their language but also some partial language effects on non-verbal measures (categorization choices, fixation counts) when salient motion-relevant information was involved during event processing. The results, **discussed in light of cognitive psychology and linguistics, support a moderate version of the relativity hypothesis** and highlight the need to formulate precise and subtle views of the relationship between language use and *on-line* visual processing that take into account cognitive and typological factors as well as the perceptual properties of events.

More generally, the findings do not support a modular approach but rather an interactive and integrated (body/mind–language–environment) approach to the Langue-Cognition debate. In such a framework (similar to what has been described recently by other psycholinguists, acquisitionists, and neuroscientists – cf. De Bot, 2017; Pulvermüller, 2018; Spivey, 2023), interactions play a central role – represented here by context effects revealing that contextual information (e.g., the demands of the task, the degree of salience of certain features, speakers' linguistic knowledge) may modulate not only our verbal behavior but also, in some cases, our actions and covert inferences. Context effects are reported in the findings of this study by taking into account different types of salience variables: spatial components involving different core events (events without boundary crossing, with one-boundary crossing, with two-boundary crossings), different co-events (default-Manners, Manners with/without instruments), different scene types (artificial/cartoon-like vs. natural/video depictions), affordances (involving human vs. animal motion features), and language involvement (language involved massively during the production task, partially during the verbal categorization task, or not at all during the non-verbal task).

The interactions reported suggest that the subsystems of perception, identified by physiologists and psychologists in the 80s and 90s, are not as modular as once thought (Fodor, 1983; Karmiloff-Smith, 1992; Marr & Vaina, 1982). Feedback and crosstalk connections with the sensorimotor/action system and the language system are or have to be added to traditional modular 'box-and-arrow' models to accommodate such findings, transforming the classic modular frameworks into something befitting an interactive and integrated network that covers findings from both action perception studies (e.g., Petro et al., 2017; Shebani & Pulvermüller, 2013; Spivey, 2007) and cognitive experimental linguistics (e.g., Boroditsky, 2012; Flecken et al., 2015; Soroli et al., 2019). To conclude, an account of interactions (within and across domains of processing) is necessary in a theory about human functions and event processing. Spatial cognition is no exception. Human cognition, in the spatial/motion domain, should be seen as a highly interactive process that involves several factors of influence including language processing.

Data availability statement. The lists of stimuli, together with the anonymized quantitative summaries of the data reported in this article, can be found at https://osf.io/2hdxg/?view_only=9408ab7e47844ed2b302ef5eedc621b3. The video and cartoon stimulus sets are available upon request.

Acknowledgments. I express my gratitude to Maya Hickmann and Philippe Bonnet for their invaluable assistance and thoughtful discussions regarding earlier versions of this work. My gratitude extends to two anonymous reviewers whose insightful feedback significantly contributed to refining this article.

Funding statement. The research presented herein received partial support from the ANR-DFG Langacross Project under award number ANR-07-FRAL-0007. Lastly, I would like to acknowledge the support generously provided by the Academic Council of the University of Lille, the National Council of Universities,

and the French Ministry of Higher Education, Research & Innovation through a PEDR (2019–2023) and a CRCT (2023–2024) grant.

Competing interest. The author declares none.

References

- Anderson, S. E., Chiu, E., Huette, S., & Spivey, M. J. (2011). On the temporal dynamics of language-mediated vision and vision-mediated language. *Acta Psychologica*, 137(2), 181–189.
- Aske, J. (1989). Path predicates in English and Spanish: A closer look. In K. Hall, M. Meacham, & R. Shapiro (Eds.), *Proceedings of the 15th Annual Meeting of the Berkeley Linguistics Society* (pp. 1–14). Berkeley Linguistics.
- 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.
- Baldo, J. V., Dronkers, N. F., Wilkins, D., Ludy, C., Raskin, P., & Kim, J. (2005). Is problem solving dependent on language? *Brain and Language*, 92(3), 240–250.
- Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59, 617–645.
- Beavers, J., Levin, B., & Tham, S. W. (2010). The typology of motion expression revisited. *Journal of Linguistics*, 46(2), 311–377.
- Bennett, S. J., Bures, R., Hecht, H., & Benguigui, N. (2010). Eye movements influence estimation of time-to-contact in prediction motion. *Experimental Brain Research*, 206(4), 399–407.
- Bock, J. K., Irwin, D. E., & Davidson, D. J. (2004). Putting first things first. In J. M. Henderson, & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world* (pp. 249–278). Psychology Press.
- Bohnemeyer, J., Bowerman, M., & Brown, P. (2001). Cut and break clips. In S. C. Levinson & N. J. Enfield (Eds.), *Field Manual 2001, Language and Cognition Group, Max Planck Institute for Psycholinguistics* (pp. 90–96). MPI.
- Boroditsky, L. (2001). Does language shape thought? Mandarin and English speakers' conception of time. *Cognitive Psychology*, 43, 1–22.
- Boroditsky, L. (2012). How the languages we speak shape the ways we think: The FAQs. In M. J. Spivey, K. McRae, & M. F. Joannisse (Eds.), *The Cambridge handbook of psycholinguistics* (pp. 615–632). Cambridge University Press.
- Bowerman, M., & Levinson, S. (Eds.) (2001). *Language acquisition and conceptual development*. Cambridge University Press.
- Carroll, M., & von Stutterheim, C. (2011). Event representation, event-time relations and clause structure: A cross-linguistic study of English and German. In J. Bohnemeyer, & E. Pederson (Eds.), *Event representation in language and cognition* (pp. 68–83). Cambridge University Press.
- Choi, S. (2006). Influence of language-specific input on spatial cognition: Categories of containment. *First Language*, 26(2), 207–232.
- Choi, S., Goller, F., Hong, U., Ansorge, U., & Yun, H. (2018). Figure and Ground in spatial language: Evidence from German and Korean. *Language and Cognition*, 10, 665–700.
- Chomsky, N. (1977). *Reflections on language* (2nd ed.). Temple Smith (Original work published 1975).
- De Bot, K. (2017). Complexity theory and dynamic systems theory: Same or different? In L. Ortega, & Z. Han (Eds.), *Complexity theory and language development: In celebration of Diane Larsen-Freeman* (pp. 51–58). John Benjamins.
- De Bot, K., Lowie, W., & Verspoor, M. (2007). A dynamic systems theory approach to second language acquisition. *Bilingualism: Language and Cognition*, 10(1), 7–21.
- De Villiers, J., & De Villiers, P. (2000). Linguistic determinism and the understanding of false beliefs. In P. Mitchell, & K. J. Riggs (Eds.), *Children's reasoning and the mind* (pp. 191–228). Psychology Press/Taylor & Francis.
- Deldar, Z., Gevers-Montoro, C., Khatibi, A., & Ghazi-Saidi, L. (2021). The interaction between language and working memory: A systematic review of fMRI studies in the past two decades. *AIMS Neuroscience*, 16(8(1)), 1–32.

- DeLucia, P. R., & Lidell, G. W. (1998). Cognitive motion extrapolation and cognitive clocking process in prediction motion tasks. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 901–914.
- Demagny, A.-C. (2015). Interrelationships between space and time in English and French discourse: Implications for second language acquisition. *Language, Interaction and Acquisition*, 6(2), 202–236.
- Divjak, D., Milin, P., & Medimorec, S. (2020). Construal in language: A visual-world approach to the effects of linguistic alternations on event perception and conception. *Cognitive Linguistics*, 31(1), 37–72.
- Fedorenko, E., & Varley, R. (2016). Language and thought are not the same thing: Evidence from neuroimaging and neurological patients. *Annals of the New York Academy of Sciences*, 1369(1), 132–153.
- Flecken, M., Athanasopoulos, P., Kuipers, J. R., & Thierry, G. (2015). On the road to somewhere: Brain potentials reflect language. Effects on motion event perception. *Cognition*, 141, 41–51.
- Flecken, M., von Stutterheim, C., & Carroll, M. (2014). Grammatical aspect influences motion event perception: Evidence from a cross-linguistic, non-verbal recognition task. *Language and Cognition*, 6(1), 45–78.
- Fodor, J. (1983). *The modularity of mind*. MIT Press.
- Fortis, J.-M. (2010). *Space in language*. Leipzig Summer School – Max Planck Institute. https://www.eva.mpg.de/lingua/conference/2010_summerschool/pdf/course_materials/Fortis_3.MOTION%20EVENTS.pdf
- Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology*, 22(3–4), 455–479.
- Gennari, S. P., Sloman, S., Malt, B., & Fitch, T. (2002). Motion events in language and cognition. *Cognition*, 83, 49–79.
- Gentner, D., & Goldin-Meadow, S. (Eds.) (2003). *Language and mind*. MIT Press.
- Gibbs, R. W., Jr., & Van Orden, G. C. (2003). Are emotional expressions intentional? A self-organizational approach. *Consciousness & Emotion*, 4(1), 1–16.
- Gibson E., Futrell R., Jara-Ettinger J., Mahowald K., Bergen L., Ratnasingam S., Gibson, M., Piantadosi, S. T., & Conway, B. R. (2017). Color naming across languages reflects color use. *Proceedings of the National Academy of Science*, 114(40), 10785–10790.
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton, Mifflin and Company.
- Gleitman, L., January, D., Nappa, R., & Trueswell, J. (2007). On the give and take between event apprehension and utterance formulation. *Journal of Memory and Language*, 57(4), 544–569.
- Goller, F., Choi, S., Hong, U., & Ansorge, U. (2020). Where of one cannot speak: How language and capture of visual attention interact. *Cognition*, 194, 1–14.
- Goller, F., Lee, D., Ansorge, U., & Choi, S. (2017). Effects of language background on gaze behavior: A crosslinguistic comparison between Korean and German speakers. *Advances in Cognitive Psychology*, 13, 267–279.
- Griffin, Z., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, 11, 274–279.
- Griffin, Z. M. (2004). Why look? Reasons for eye movements related to language production. In J. Henderson, & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world* (pp. 213–247). Psychology Press.
- Helo, A., Rämä, P., Pannasch, S., & Meary, D. (2016). Eye movement patterns and visual attention during scene viewing in 3- to 12-month-olds. *Visual Neuroscience*, 33, 1–7.
- Henderson, J., & Ferreira, F. (2004). Scene perception for psycholinguists. In J. M. Henderson, & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world* (pp. 1–58). Psychology Press.
- Hickmann, M. (2006). The relativity of motion in first language acquisition. In M. Hickmann, & S. Robert (Eds.), *Space across languages: Linguistic systems and cognitive categories* (pp. 281–308). John Benjamins.
- Hickmann M., Engemann H., Soroli E., Hendriks H. & Vincent C. (2017). Expressing and categorizing motion in French and English: Verbal and non-verbal cognition across languages. In Ibarretxe-Antuñano, I. (Ed.), *Motion and space across languages and applications* (pp. 61–94). Human Cognitive Processing Series. John Benjamins.
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.
- Hotton, S., & Yoshimi, J. (2011). Extending dynamical systems theory to model embodied cognition. *Cognitive Science*, 35(3), 444–479.

- Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, 60(6), 581–592.
- Ibarretxe-Antuñano, I. (2009). Path salience in motion events. In J. Guo, E. Lieven, N. Budwig, S. Ervin-Tripp, K. Nakamura, & Ş. Özçalışkan (Eds.), *Crosslinguistic approaches to the psychology of language: Research in the tradition of Dan Isaac Slobin* (pp. 403–414). Psychology Press.
- Jackendoff, R. (1990). *Semantic structures*. MIT Press.
- Jackendoff, R. (1996). How language helps us think. *Pragmatics and Cognition*, 4, 1–24.
- Ji, Y. (2017). Motion event similarity judgments in one or two languages: An exploration of monolingual speakers of English and Chinese vs. L2 learners of English. *Frontiers in Psychology*, 8, 909. <https://doi.org/10.3389/fpsyg.2017.00909>
- Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental approach to cognitive science*. MIT Press.
- Keshavarz, B., Campos, J. L., DeLucia, P. R., & Oberfeld, D. (2017). Estimating the relative weights of visual and auditory tau versus heuristic-based cues for time-to-contact judgments in realistic, familiar scenes by older and younger adults. *Attention, Perception & Psychophysics*, 79, 929–944.
- Kiefer, M., & Pulvermüller, F. (2012). Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. *Cortex*, 48, 805–825.
- Kopecka, A., & Vuilleumet, M. (2021). Source-Goal (a)symmetries across languages. *Studies in Language*, 45(1), 2–35.
- Landau, B., & Lakusta, L. (2006). Spatial language and spatial representation: Autonomy and interaction. In M. Hickmann, & S. Roberts (Eds.), *Space in languages: Linguistic systems and cognitive categories. Part of the Typological Studies in Language series* (pp. 309–333). John Benjamins.
- Larsen-Freeman, D., & Cameron, L. (2008). *Complex systems and applied linguistics*. Oxford University Press.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. MIT Press.
- Levin, B. & Rappaport Hovav, M. (1992). The lexical semantics of verbs of motion: The perspective from unaccusativity. In Roca, I. M. (Ed.), *Thematic structure: Its role in grammar* (pp. 247–269). Foris.
- Levinson, S. (1996). Frames of reference and Molyneux's question: Crosslinguistic evidence. In P. Bloom, M. Peterson, L. Nadel, & M. Garrett (Eds.), *Language and space* (pp. 109–170). MIT Press.
- Levinson, S. (2003). *Space in language and cognition: Explorations in cognitive diversity*. Cambridge University Press.
- Lucy, J. (1992). *Grammatical categories and cognition: A case study of the linguistic relativity hypothesis*. Cambridge University Press.
- Luna, B., Velanova, K., & Geier, C. F. (2008). Development of eye-movement control. *Brain and Cognition*, 68(3), 293–308.
- Lupyan, G., Abdel Rahman, R., Boroditsky, L., & Clark, A. (2020). Effects of language on visual perception. *Trends in Cognitive Science*, 24(11), 930–944.
- Lupyan, G., & Spivey, M. J. (2010). Making the invisible visible: Verbal but not visual cues enhance visual detection. *PLoS One*, 5(7), e11452.
- Mahon, B., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology*, 102(1–3), 59–70.
- Majid, A., Bowerman, M., Kita, S., Haun, D., & Levinson, S. (2004). Can language restructure cognition? The case of space. *Trends in cognitive sciences*, 8, 108–114.
- Marr, D., & Vaina, L. (1982). Representation and recognition of the movements of shapes. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 214(1197), 501–524.
- Matsumoto, Y. (2003). Typologies of lexicalization patterns and event integration: Clarifications and reformulations. In S. Chiba, et al. (Eds.), *Empirical and theoretical investigations into language: A Festschrift for Masaru Kajita* (pp. 403–418). Kaitakusha.
- McClelland, J. L. (1996). Integration of information: Reflections on the theme of attention and performance XVI. In T. Inui, & J. L. McClelland (Eds.), *Attention and Performance XVI. Information integration in perception and communication* (pp. 633–656). MIT Press.
- Meyer, A. S., & Lethaus, F. (2004). The use of eye tracking in studies of sentence generation. In J. M. Henderson, & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world* (pp. 191–211). Psychology Press.
- Montero-Melis, G., Eisenbeiss, S., Narasimhan, B., Ibarretxe-Antuñano, I., Kita, S., Kopecka, A., Lüpke, F., Nikitina, T., Trigel, I., Florian Jaeger, T., & Bohnemeyer, J. (2017). Satellite- vs. verb-framing underpredicts nonverbal motion categorization: Insights from a large language sample and simulations. *Cognitive Semantics*, 3(1), 36–61.

- Monti, M. M., Parsons, L. M., & Osherson, D. N. (2012). Thought Beyond Language: Neural Dissociation of Algebra and Natural Language. *Psychological Science*, 23(8), 914–922.
- Munnich, E., Landau, B., & Doshier, B. A. (2001). Spatial language and spatial representation: A cross-linguistic comparison. *Cognition*, 81(3), 171–207.
- Naigles, L., & Terrazas, P. (1998). Motion-verb generalizations in English and Spanish: Influences of language and syntax. *Psychological Science*, 9, 363–369.
- Pannasch, S., Schulz, J., & Velichkovsky, B. M. (2011). On the control of visual fixation durations in free viewing of complex images. *Attention, Perception & Psychophysics*, 73(4), 1120–1132.
- Papafragou, A., Hulbert, J., & Trueswell, J. (2008). Does language guide event perception? Evidence from eye movements. *Cognition*, 108, 155–184.
- Papafragou, A., Massey, C., & Gleitman, L. (2002). Shake, rattle, ‘n’ roll: The representation of motion in thought and language. *Cognition*, 84, 189–219.
- Papafragou, A., Massey, C., & Gleitman, L. (2006). When English proposes what Greek presupposes: The linguistic encoding of motion events. *Cognition*, 98, B75–B87.
- Papafragou, A., & Selimis, S. (2010). Event categorization and language: A cross-linguistic study of motion. *Language and Cognitive Processes*, 25, 224–260.
- Park, H., Lee, S., Lee, M., Chang, M. S., & Kwak, H. W. (2016). Using eye movement data to infer human behavioral intentions. *Computers in Human Behavior*, 63, 796–804.
- Petro, L. S., Paton, A. T., & Muckli, L. (2017). Contextual modulation of primary visual cortex by auditory signals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1714), 20160104.
- Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*. MIT Press.
- Pulvermüller, F. (2013). Semantic embodiment, disembodiment or misembodiment? In search of meaning in modules and neuron circuits. *Brain and Language*, 127(1), 86–103.
- Pulvermüller, F. (2018). Neurobiological mechanisms for semantic feature extraction and conceptual flexibility. *Topics in Cognitive Science*, 10(3), 590–620.
- Pulvermüller, F., & Fadiga, L. (2010). Active perception: Sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*, 11, 351–360.
- Pulvermüller, F., Hauk, O., Nikulin, V. V., & Ilmoniemi, R. J. (2005). Functional links between motor and language systems. *European Journal of Neuroscience*, 21(3), 793–797.
- Richardson, D. C., & Matlock, T. (2007). The integration of figurative language and static depictions: An eye movement study of fictive motion. *Cognition*, 102(1), 129–138.
- Richardson, D. C., Spivey, M. J. & Cheung, J. (2001). Motor representations in memory and mental models: Embodiment in cognition. In *Proceedings of the 23rd Annual Conference of the Cognitive Science Society* (pp. 867–872). Erlbaum.
- Riegel, M., Pellat, J. C., & Rioul, R. (1999). *Grammaire méthodique du français*. PUF.
- Ryskin, R., & Fang, X. (2021). The many timescales of context in language processing. In K. D. Federmeier, & L. Sahakyan (Eds.), *Psychology of learning and motivation* (pp. 201–243). Academic Press.
- Shapiro, L. (2011). *Embodied cognition*. Routledge.
- Shebani, Z., & Pulvermüller, F. (2013). Moving the hands and feet specifically impairs working memory for arm- and leg-related action words. *Cortex*, 49(1), 222–231.
- Slobin, D. I. (1987). Thinking for speaking. In *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society* (pp. 435–445). Linguistic Society of America.
- Slobin, D. I. (2005). Linguistic representations of motion events: What is signifier and what is signified? In C. Maeder, O. Fischer, & W. J. Herlofsky (Eds.), *Iconicity in language and literature: Vol. 4. Outside-in – inside-out* (pp. 307–322). John Benjamins.
- Slobin, D. I. (2006). What makes manner of motion salient? Explorations in linguistic typology, discourse, and cognition. In M. Hickmann, & S. Robert (Eds.), *Space across languages: Linguistic systems and cognitive categories* (pp. 59–81). John Benjamins.
- Slobin, D. I., Ibarretxe-Antuñano, I., Kopecka, A., & Majid, A. (2014). Manners of human gait: A cross-linguistic event-naming study. *Cognitive Linguistics*, 25, 701–741.
- Soroli, E. (2011). *Language and spatial cognition in French and in English: Crosslinguistic perspectives in aphasia* [unpublished doctoral dissertation]. University of Paris 8, France.
- Soroli, E. (2012). Variation in spatial language and cognition: Exploring visuo-spatial thinking and speaking cross-linguistically. *Cognitive Processing*, 13(1), 333–337.

- Soroli, E. (2018). Event processing in agrammatic aphasia: Does language guide visual processing and similarity judgments? *Aphasiology*, 32(1), 219–221.
- Soroli, E., & Hickmann, M. (2010). Language and spatial representations in French and in English: Evidence from eye-movement. In G. Marotta, A. Lenci, L. Meini, & F. Rovai (Eds.), *Space in language* (pp. 581–597). Edizione Testi Scientifici.
- Soroli, E., Hickmann, M., & Hendriks, H. (2019). Casting an eye on motion events: Eye tracking and its implications for typology. In M. Aurnague, & D. Stosic (Eds.), *The semantics of dynamic space in French: Descriptive, experimental and formal studies on motion expression* (pp. 249–288). John Benjamins.
- Soroli, E., & Verkerk, A. (2017). Motion events in Greek. *Cognitextes – Revue de l'Association Française de Linguistique Cognitive*, 15, 1–54. <https://doi.org/10.4000/cognitextes.889>
- Spivey, M. (2007). *The continuity of mind*. Oxford University Press.
- Spivey, M. (2023). Cognitive science progresses toward interactive frameworks. *Topics in Cognitive Science*, 15, 219–254.
- Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Grammatical categories and the lexicon* (Vol. III, pp. 57–149). Cambridge University Press.
- Talmy, L. (2000). *Toward a cognitive semantics. Volume 1: Concept structuring systems. Volume 2: Typology and process in concept structuring*. MIT Press.
- Talmy, L. (2005). The fundamental system of spatial schemas in language. In B. Hampe (Ed.), *From perception to meaning: Image schemas in cognitive linguistics* (pp. 199–234). de Gruyter.
- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.
- Tresilian, J. R. (1995). Perceptual and cognitive processes in time-to-contact estimation: Analysis of prediction motion and relative judgment tasks. *Perception & Psychophysics*, 57, 231–245.
- Tucker, M., & Ellis, R. (1998). On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology: Human Perception and Performance*, 24(3), 830–846.
- Varley R. A., Klessinger N. C., Romanowski C. A. J., & Siegal M. (2005). Agrammatic but numerate. *Proceedings of the National Academy of Sciences, USA*, 102, 3519–3524.
- Whorf, B. L. (1956). Linguistics as an exact science. In J. B. Carroll (Ed.), *Language, thought and reality. Selected writings of Benjamin Lee Whorf* (pp. 220–232). MIT Press.
- Willems, R., Benn, Y., Hagoort, P., Toni, I., & Varley, R. (2011). Communicating without a functioning language system: Implications for the role of language in mentalizing. *Neuropsychologia*, 49, 3130–3135.
- Willems, R. M., & Casasanto, D. (2011). Flexibility in embodied language understanding. *Frontiers in Psychology*, 2, 116.
- Willems, R. M., & Hagoort, P. (2007). Neural evidence for the interplay between language, gesture, and action: A review. *Brain and Language*, 101(3), 278–289.
- Wispinski, N. J., Gallivan, J. P., & Chapman, C. S. (2020). Models, movements, and minds: Bridging the gap between decision making and action. *Annals of the New York Academy of Sciences*, 1464(1), 30–51.
- Yun, H., & Choi, S. (2018). Spatial semantics, cognition, and their interaction: A comparative study of spatial categorization in English and Korean. *Cognitive Science*, 42(6), 1736–1776.
- Zgonnikov, A., Aleni, A., Piironen, P. T., O'Hora, D., & di Bernardo, M. (2017). Decision landscapes: Visualizing mouse-tracking data. *Royal Society Open Science*, 4(11), 170482.

Cite this article: Soroli, E. (2024). How language influences spatial thinking, categorization of motion events, and gaze behavior: a cross-linguistic comparison, *Language and Cognition*, 1–45. <https://doi.org/10.1017/langcog.2023.66>