VERBAL AND NON-VERBAL SPATIAL COGNITION ACROSS LANGUAGES: EVIDENCE FROM ENGLISH AND FRENCH

Efstathia SOROLI* & Maya HICKMANN

Laboratoire Structures Formelles du Langage, UMR 7023, CNRS & Université Paris 8

ABSTRACT

Languages differ in how they lexicalize and grammaticalize spatial information, thereby constraining how speakers organize spatial information to encode motion in discourse. Such cross-linguistic differences raise new questions concerning the relation between language and thought. In this context we examined how speakers of two typologically different languages, English and French, performed several tasks, all of which involved motion events: a production task (describing visual scenes showing motion events), a non-verbal categorization task (grouping visual stimuli of these events), a verbal categorization task (deciding which visual stimulus best corresponds to a sentence describing a motion event). In addition, all three tasks were coupled with an eye-tracking paradigm measuring on-line how participants allocated their visual attention when exploring these events. Subjects’ verbalizations during the production task differed substantially in the two groups as a function of language-specific factors. French speakers focused mostly on Path information (lexicalized in the verb), while English speakers expressed Manner (in the verb) and Path (outside of the verb) equally often, thereby producing denser utterances. Subjects’ preferential choices during the categorization tasks (in both of its verbal and non-verbal versions) were guided by different criteria. However, a more important language effect was observed in the verbal version of this task as compared to its non-verbal version. Finally, although speakers in both groups allocated more attention to Path information overall during their visual exploration of the events, their focus

* Address correspondence to Efstathia Soroli, Laboratoire Structures Formelles du Langage, CNRS & Université de Paris 8, Office 148, 59 rue Pouchet, 75017, Paris, France, e-mail: eva.soroli@sfl.cnrs.fr
of attention also varied across groups, suggesting a greater focus on Path in French as compared to English, as well as a different unfolding of attention during the processing of the visual stimuli. However, the eye-tracking data must be interpreted with care since they also showed variable patterns that seem to be stimulus dependent. In conclusion, we argue that our results support a moderate version of the relativity hypothesis according to which typological language properties have a clear impact on linguistic behaviours, but also on non-linguistic behaviours, although to a lesser extent, in some tasks, and with some stimuli.

KEY WORDS:
Motion events, Language, Space, Categorization, Production, Eye movements

1. INTRODUCTION

Recent psycholinguistic research has begun to examine the cognitive implications of linguistic diversity, thereby reviving the question of the relation between language and thought. Most cognitive and linguistic theories suggest that human perception and cognitive processing are universal. Given the innate character of the human brain maturation and regardless one’s environmental and linguistic experience, language is believed to be guided by universal determinants (Chomsky, 1975; Pinker, 1989, Clark & Clark, 1978; Tomasello, 2003; Gleitman & Papafragou, 2005). The opposite, ‘relativistic’ point of view has been particularly supported by researchers who suggest that important aspects of human cognition may not be universal but at least partially determined by language-specific factors (Whorf, 1956; Lucy, 1992; Pederson et al., 1998; Boroditsky, 2001; Bowerman & Levinson, 2001; Gentner & Goldin-Meadow, 2003; Hickmann, 2006; Hickmann et al., 2010). In the context of this debate and despite the fact that language and thought are considered to be two different
components of higher mental functions in humans, modern psychological and linguistic theories recognize some relationship between language and thought.

Researchers seem to agree that the domain of space provide an empirically rich and a tractable domain to investigate language-thought relations (Levinson, 1996; Boroditsky, 2001) because it is characterized by considerable crosslinguistic variability (Talmy, 1985). Thus, with respect to motion events, this variability includes asymmetries across languages in the types of semantic information that are preferentially encoded (e.g. path, manner), in the types of encoding patterns that are observed (lexical or grammatical), as well as in the density and frequency of encoded information.

Our aim is to investigate whether strong typological differences have an impact on non-linguistic representations of motion events. After a discussion of previous experimental cross-linguistic research on linguistic and non-linguistic representations of these events (Sections 1.1 and 1.2 respectively), we propose a number of hypotheses concerning the relation between the linguistic encoding of motion scenes and the non-linguistic attention allocation and conceptualization of these events (Section 1.3). We then present the experimental framework (Section 2) and results (Section 3) of three experiments that tested these hypotheses in two languages (English and French). Finally, we discuss these results in relation to broader issues raised by the language-thought debate (Section 4) and draw some conclusions concerning their potential implications (Section 5).

1.1. Linguistic representations of motion across languages

Languages differ in how they organize spatial information. There is a more or less universal set of semantic components that are expressed in almost all languages such as Manner (e.g. to jump, to swim), Path (e.g., to arrive, to leave), Figure (e.g., the ball rolled...
across the field) and Ground (e.g., to enter into the house, to run across the street). However, languages provide speakers with a limited number of linguistic means to express these components, thereby acting as ‘filters’ that lead speakers to select different components in their verbalizations of spatial information while ignoring others.

More specifically, according to Talmy (2000), languages either lexicalize or grammaticalize information concerning motion, thereby highlighting some types of information more than others. Talmy proposes a distinction between two types of languages: satellite-framed and verb-framed languages (e.g., Germanic vs. Romance languages). Thus, when expressing motion, a satellite-framed language such as English lexicalizes the Manner of motion in the verb (e.g. to run, to crawl), using satellites (e.g., the particles into, across) to express Path information within one compact structure, as illustrated in examples 1a and 1b:

(1) a. He **is running** into the house.

FIGURE MANNER PATH GROUND

b. The baby **is crawling** across the street.

FIGURE MANNER PATH GROUND

In contrast, a verb-framed language such as French lexicalizes Path in the verb stem (e.g., entrer ‘to enter’), leaving Manner information implicit or placing it at the periphery of the sentence, for instance by means of gerunds (e.g., *en courant* ‘by running’) or of adverbials (e.g., *à quatre pattes* ‘on all fours’), as illustrated in examples 2a and 2b:

(2) a. Il **entre** dans la maison *en courant*.

FIGURE PATH GROUND MANNER

Lit. ‘He is entering in[to] the house by running.’

b. Le bébé **traverse** la rue à quatre pattes.
Thus, although all languages provide different types of path and manner verbs, *satellite-framed* and *verb-framed* languages display preferred lexicalization patterns. Several experimental studies indeed show differences in speakers’ preferences for encoding *Manner* or *Path* components across languages (Choi & Bowerman, 1991; Slobin1996a,b, 2004, 2006; Hickmann, et al. 2009), furthermore suggesting that these differences are already in place during the earliest stages of language acquisition (Berman & Slobin 1994; Naigles et al., 1998; Bowerman & Choi, 2003; Allen et al., 2007; Papafragou et al., 2002; Papafragou & Selimis, 2007).

1.2. *Non-linguistic representations of motion across languages*

Such striking differences across languages are of great relevance for the study of spatial cognition and have led to some debates opposing the two main approaches to the relation between language and cognitive processes. According to a *universalist* hypothesis, spatial cognition is based on universal perceptual and cognitive processes that are independent from language-specific properties and therefore no deep cross-linguistic differences are expected at the level of speakers' mental representations. In contrast, the *relativity* hypothesis puts forth that language-specific factors should affect, at least partially, how speakers construct spatial representations, because languages provide filters that channel incoming information, making different aspects of reality more salient and accessible. In this view, language-specific performances are therefore expected and such differences should affect performance beyond language use.
An example of the latter view can be found in Slobin’s ‘thinking for speaking’ (1996b) hypothesis, which suggests that speakers of different languages do not attend to the same components of events to an equal degree because their language does not make all components equally salient. When communicating, speakers construe situations in terms of those dimensions that are privileged in their own language.

“Thinking for speaking” involves picking those characteristics of objects and events that “(a) fit some conceptualization of the event, and (b) are readily encodable in the language” (Slobin 1996b :76).

Levinson (2002, 2003) furthermore suggests that thinking is shaped by language-specific needs even in non communicative situations and therefore language-specific constraints should manifest themselves even in non-linguistic tasks.

Most studies showing language effects have been based on analyses of language use. In contrast, studies examining non-verbal performances (e.g., memory, categorization tasks, eye-movements) report either no language effect, or effects that are less clear and/or viewed as being relatively superficial (Landau & Lakusta, 2006; Papafragou et al., 2002, 2006). However, several other studies have found language-specific differences in non-linguistic tasks that either were performed immediately after verbal encoding (e.g. categorization: Naigles & Terrazas, 1998; memory and similarity judgments: Slobin, 2005; Gennari et al., 2002) or effects that arise only when linguistic forms are recruited in the same task (eye-movements: Papafragou et al. 2008; categorization: Papafragou & Selimis, 2010). Some recent studies show that language-specific verbalization preferences are also reflected in different patterns of attention distribution to manner and path areas in short cartoons and video clips in the domain of space (Soroli & Hickmann, 2009, in press) and in the domain of aspect (von Stutterheim & Carroll, 2006; Flecken, 2010). It therefore remains to be
investigated further whether typological differences affect speakers’ conceptual representations beyond language itself and to what extent.

1.3. *Experimental prospectus: Hypotheses and predictions*

The debate illustrated in previous sections presents a great challenge in need of an empirical basis that would allow us to distinguish among several hypotheses. Although most approaches (including both the universal and relativity hypotheses) might agree that typological differences might affect speakers’ linguistic performance, the major question to be addressed is to determine the extent to which such language differences go beyond language use. Thus, the absence of language effects on speakers’ performance in non-verbal tasks would either not support the relativity hypothesis or at best support a ‘weak’ version of this hypothesis. In contrast, the finding that differences in linguistic performance are accompanied by differences in non-verbal performance would support a ‘strong’ relativity hypothesis.

In the context of this debate, we conducted a study that elicited verbal and non-verbal responses from the same subjects and about the same motion events. In particular, this study investigated the extent to which the properties of two typologically contrasted languages (French and English) would constrain how speakers performed several tasks that provided different measures of their cognitive processes: how they described motion events (production task); how they categorized these events non-verbally in the absence of any relevant linguistic information (non-verbal categorization); how they categorized these events non-verbally when the stimuli included some linguistic information (verbal categorization) and how they allocated attention when exploring these events during production and categorization tasks (eye-tracking paradigm).

The study addresses two general questions: Does speakers’ linguistic performance in the
production task differ significantly as a function of their language? If an important language effect is found in this task, does it go beyond linguistic production and affect other measures and/or tasks? Furthermore, the design was constructed in such a way as to allow us to examine four hypotheses:

1. A null hypothesis according to which no major language effects should be found in any task, including in the production task where possible differences might be shown merely superficial.

2. A relativity hypothesis which predicts language effects, several versions of which can be distinguished:

   2a. A strong version of this hypothesis predicts language-specific preferences in all tasks, regardless of whether they elicited verbal and non-verbal responses, and regardless of the verbal vs. non-verbal nature of the stimuli.

   2b. A weak version predicts that speakers should behave quite differently in their linguistic performance (production task), but not necessarily in any of their non-verbal responses as measured in the present study by both categorization tasks.

   2c. A moderate version predicts language effects in production, as well as in tasks that require non-verbal responses, but only when relevant linguistic information is simultaneously provided.

2. METHOD

2.1. Participants

The study included 28 native speakers of French and English (14 per language, all right-handed). Inclusion criteria included the following: (1) participants had to be native,
monolingual speakers of English or French, (2) they had to be above 18 years of age, and (3) they had to report no known disorders or deficits (no psychiatric disorders, no seeing or hearing impairment, and no known history of developmental reading/oral language difficulties). All participants were asked to fill out a questionnaire about their language background. They all had been exposed to only one language since birth and had not learned any foreign language before at least age 10 (compulsory teaching at school). None had lived in a foreign country for more than six months. The analysis of the eye-movements included the data of 24 participants (12 per language).  

2.2. Materials

The experimental situation was based on Soroli & Hickmann (in press) and involved three tasks: non-verbal categorization, production, and verbal categorization, all coupled with an eye-tracking paradigm. These tasks involved two sets of stimuli all of which showed voluntary motion. The first set (hereafter video clips) consisted of video films in which an actor (one among three men and three women) performed displacements in different scenes (either outdoors or indoors) that varied along Manner and Path. Six types of Paths were selected (up, down, into, out of, across, along), as well as six types of Manners that either involved the use of an instrument (bicycle, scooter, rollers) or involved no such instrument (run, jump, crawl). The duration of each clip was 4 seconds. The use of an instrument was expected to be most effective in distinctly focusing speakers’ attention on Manner, particularly for the purpose of measuring eye-movements. An additional Manner was included in some items (walk) as it was assumed to be most ‘neutral’ and to attract the least attention on Manner (given that all characters were human). A partially crossed design combined these Manners and Paths across video clips (also see more details below). The

1 Due to insufficient data points during the recording procedure, two subjects had to be excluded from each language group.
second set of stimuli (hereafter *cartoons*) consisted of animated drawings showing characters performing displacements in different *Manners* (e.g., walking, swimming, climbing a tree) along three types of *Paths* (up, down, across). This second set provided more vivid and more varied stimuli (humans and animals, diverse settings). The mean duration of each stimulus in this set was 6.5 seconds.

2.3. *Production task*

Both sets of stimuli (video clips and cartoons) were used in the production task. There was a total of 43 items (2 training, 9 video clips, 18 cartoons, 9 controls, and 5 distractor items) that were presented to all subjects. Stimuli from the two sets were interspersed in a mixed list of items and presented in a pseudo-randomized order. After they saw each stimulus, participants were asked to describe what had happened.

The analysis examined three aspects of their verbal responses: the number of information components they expressed (hereafter *density*), the types of information they selected to express (*focus*), and the means whereby they expressed this information in their responses (*locus*). In the present study, particular attention was placed on whether subjects expressed only *Manner* information (M-responses), only *Path* information (P-responses), or both types of information (MP-responses). In addition, the analysis examined whether they used verbs vs. other devices (particles, prepositions, adverbials) to express each type of information. Our prediction was that speakers should produce structures that are based on the features of their language. More specifically: 1) although they should express *Path* in both languages, they should add *Manner* information more frequently in English than in French; 2) English responses should contain *Manner* verbs and other devices marking *Path*, whereas French responses should mostly contain *Path* verbs and fewer other devices than English responses. As a result, English descriptions should show a higher level of semantic density, since English
speakers should mostly produce MP-responses (two expressed information components, i.e.,
semantic density 2, hereafter SD2), whereas French speakers should mostly produce P-
responses (SD1).

2.4. Non-verbal categorization task

For this task only the video clips were used. Stimuli were digitized at 25 frames/sec, 1008
Bitrate kbps, 384:288 scale size and 4:3 aspect, digitally edited and stored on a computer disk.
On the basis of these films 30 main triad trials were constructed. The procedure was as
follows for each trial. Participants first saw a short target video showing a motion event
performed in a certain Manner and along a certain Path (for example Enter-Bicycle in (3a)).
The target video then disappeared and was followed by two other videos, variants of the target
that differed from it with respect either to Path or to Manner (e.g. Exit-Bicycle and Enter-
Scooter in (3b) and (3c)).

(3) a. Target Video Enter-Bicycle: a woman entering a building on a bicycle
   b. Variant Video 1 Exit-Bicycle: a woman exiting a building on a bicycle
   c. Variant Video 2 Enter-Scooter: a woman entering a building on a scooter

Participants were asked to choose the variant that looked most like the target and to press a
key as fast as they could to indicate their choice. Note that there was no correct answer in any
main trial, so that subjects’ responses reveal their preferences in using Manner or Path as the
criterion for categorization. The task started with three training trials to ensure that
participants understood the task, followed by the main triads, to which 14 control and 7
distractor items were added, resulting in a total of 54 items, presented in a randomized order.\(^2\)

\(^2\) Control items consisted in triads constructed with the same video clips as the main items. The procedure was the
same as for the main triads, except that in this case one of the two variants proposed in the control items
corresponded exactly to the target video displayed at the beginning of each item. Control items allowed us to check
2.5. Verbal categorization task

The verbal categorization task was exactly the same as the non-verbal categorization task, except that the target video was replaced by a target sentence presented auditorily at the same time as the two video variants. Stimuli were digitized at 44 KHz and 16 bits. Sentences were recorded by two female native speakers (of French and of English respectively) and they contained both Path and Manner information, as illustrated by the target sentences shown in (4a) for French and English respectively. Participants were asked to choose which variant video (for example (4b) or (4c)) was best described by the target sentence and to press a key to indicate their choice as fast as they could. Subjects’ responses show their categorical preferences, i.e., their preferred reliance on Manner or on Path as criterion. The question was to determine whether categorical preferences followed language properties. If so, French speakers should choose the Path criterion, whereas English speakers should either choose the Manner criterion or show no preference for one or the other criterion.

(4)

a. Target Sentence - Enter-Bicycle

French: « On voit quelqu’un qui entre en vélo ».

(Lit. ‘One sees someone entering by bike.’)

English: «There is someone cycling in ».

b. Variant Video 1 - Exit-Bicycle :

A woman exiting a building on a bicycle

c. Variant Video 2 - Enter-Scooter

A woman entering a building on a scooter
2.6. Procedure

The three tasks were presented successively in a fixed order. Testing always started with the non-verbal categorization task, followed by the production task, and then by the verbal categorization task. The rationale for this task order was three-fold. First, it was essential for the non-verbal categorization task to be presented first, since this task was meant to elicit non-verbal responses that would involve no linguistic input of any kind. Second, in contrast to the non-verbal categorization task, the verbal version of this task was meant to elicit non-verbal responses that would involve linguistic input (target sentences). Third, it was essential for the production task to occur before the verbal categorization task in order for subjects’ descriptions not to be influenced by the target sentences presented during this task.

Subjects’ eye-movements were recorded during all experimental tasks with a portable eye-tracker while participants were exploring the videos and cartoons in such a way as to measure their attention to various aspects of these events. The question addressed was as follows: Do subjects pay more or less attention to Manner and/or to Path depending on the specific properties of their language? A further question addressed in this research framework is whether language-specific mapping preferences can also affect non-linguistic attention allocation and categorical performance. We present below data concerning subjects’ fixations in order to examine whether they paid attention to the following Areas of Interest (AoI): Path (P), Path±Manner (P±M), Source (S) and Goal (G)\(^3\) for the production task; Path criterion area (p) and Manner criterion area (m) for the categorization task.

\(^3\) The mixed area P±M was created for the analysis, given the difficulty in differentiating Manner from Path areas for some fixations in this task. For our purposes here, only the eye-movements concerning P, P±M, p and m areas are presented and discussed. Ongoing analyses examine eye-movements across all AoI (for further illustrations see Soroli & Hickmann, in press).
3. RESULTS

We were particularly interested in determining whether language effects could be observed in each task with both verbal and non-verbal measures. Recall that our major aim was to examine response patterns across tasks and measures in order to test several hypotheses that differed as follows with respect to their predictions: no language effect in any task (null hypothesis), effects in all tasks (strong relativity), an effect only in the production task (weak relativity), effects in tasks involving verbal input, i.e. the production task and the verbal categorization task (moderate relativity).

3.1. Production task

Verbal responses in the production task

We first focus on the linguistic measures during the production task. Verbal responses were analyzed in several ways in order to determine how much information was expressed (hereafter density) and which particular semantic components were expressed (focus), as well as where it was expressed (locus), with particular attention to two sites: main verbs vs. other devices (particles, prepositions, adverbials). For each language (English, French) and each set of stimuli (Cartoons, Video-clips), Figures 1 and 2 show (in means) the semantic components that were expressed in verbs vs. other devices: Path (P), Manner (M), both (PM), or neither (none).
Regardless of stimuli types, French speakers mostly produced *Path* verbs and provided less information about either P or M in other devices (examples 5a, 5b). In contrast, English speakers used compact and dense structures that systematically expressed *Manner* in verbs and *Path* in other devices (examples 5c, 5d). As a result, the semantic density of subjects’
responses was higher in English than in French. Overall 90% of English responses encoded both *Path* and *Manner* (MP responses, therefore density SD2); in comparison, only 55% of French responses were of density SD2, while 45% only contained one type of information, mostly *Path* (P-responses, SD1). These results are in line with the prediction that language-specific properties affect all aspects in speakers’ verbal responses (density, focus, locus).

(5) a. C’est un ours qui *monte* à l’arbre pour aller chercher du miel. (cartoon)

*FIGURE*  
*PATH GROUND*  
[Path in verb, no *Manner*]

Lit. ‘It’s a bear who is ascending to the tree to go [and] get some honey.’

b. C’est une fille qui *rentre* dans la maison. (video)

*FIGURE*  
*PATH GROUND*  
[Path in verb, no *Manner*]

Lit. ‘It’s a girl who is entering in[to] the house.’

c. There’s a mouse *climbing up* the leg of the table to get some cheese. (cartoon)

*FIGURE*  
*MANNER PATH GROUND*  
[Manner in verb, *Path* in satellite]

d. There’s a girl *jumping up* the hill. (video)

*FIGURE*  
*MANNER PATH GROUND*  
[Manner in verb, *Path* in satellite]

*Eye-movements during the production task*

Table 1 shows for each language group the number and length of fixations on each of the AoI (*P±M, P, S, G*) during the production task with all stimuli (videos and cartoons). The overall results first show no difference in fixation lengths for each area of interest across languages. In both languages, they also show more and longer fixations for *P±M* and *P* areas than for *S* and *G* areas. Third, English speakers’ fixations show no preference for *P±M* vs. *P* areas and French speakers’ fixations only show a slight tendency to focus more on *P* areas than on *P±M* (higher count for *P*-fixations, but no difference in fixation length).
Since the results differed across the video and cartoon stimuli, we present below the data for each set separately. We will subsequently return to a discussion of item effects in Section 4.

**Video Stimuli.** With respect to the eye movements that were elicited with the videos, Table 2 shows the fixation counts and lengths for all AoI with all clips in each language group. These data only show a slight preference for $P\pm M$ areas across the two languages in terms of fixation counts but not in lengths.

<table>
<thead>
<tr>
<th>Fixation Count</th>
<th>$P\pm M$</th>
<th>$P$</th>
<th>$S$</th>
<th>$G$</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>33%</td>
<td>31%</td>
<td>19%</td>
<td>18%</td>
</tr>
<tr>
<td>French</td>
<td>35%</td>
<td>31%</td>
<td>9%</td>
<td>24%</td>
</tr>
<tr>
<td>Fixation Length</td>
<td>$P\pm M$</td>
<td>$P$</td>
<td>$S$</td>
<td>$G$</td>
</tr>
<tr>
<td>English</td>
<td>30%</td>
<td>30%</td>
<td>19%</td>
<td>21%</td>
</tr>
<tr>
<td>French</td>
<td>37%</td>
<td>28%</td>
<td>8%</td>
<td>28%</td>
</tr>
</tbody>
</table>

Table 1. *English and French fixation count and length for all AoI with all stimuli*

<table>
<thead>
<tr>
<th>Fixation Count</th>
<th>$P\pm M$</th>
<th>$P$</th>
<th>$S$</th>
<th>$G$</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>31%</td>
<td>30%</td>
<td>19%</td>
<td>19%</td>
</tr>
<tr>
<td>French</td>
<td>29%</td>
<td>34%</td>
<td>16%</td>
<td>21%</td>
</tr>
<tr>
<td>Fixation Length</td>
<td>$P\pm M$</td>
<td>$P$</td>
<td>$S$</td>
<td>$G$</td>
</tr>
<tr>
<td>English</td>
<td>31%</td>
<td>30%</td>
<td>20%</td>
<td>19%</td>
</tr>
<tr>
<td>French</td>
<td>30%</td>
<td>31%</td>
<td>16%</td>
<td>24%</td>
</tr>
</tbody>
</table>

Table 2. *English and French fixation count and length for all AoI with all Video stimuli*

In addition, the Timeline of fixations for all video items in English (Figure 3) and in French (Figure 4) shows no striking preference for $P$ or $P\pm M$ areas in any language and no difference between the two language groups. English speakers pay as much attention to $P$ and $P\pm M$ areas despite slight variations during processing (slight preference for $P\pm M$ at stimulus onset and at 1500-2500 ms, but for $P$ at 2500-3500 ms). In the French group fixations do not vary much during the entire duration of the stimulus, showing no preference for $P$ or $P\pm M$.
areas, except for a slight tendency to look at $P \pm M$ at around 2000-2500 ms.

However, if we now examine the data for the sub-set of ‘into’ and ‘out of’ items (Table 3), we observe a preference for looking at $P$ areas in both language groups, as measured by both the number and the length of fixations.
<table>
<thead>
<tr>
<th>Fixation Count</th>
<th>P±M</th>
<th>P</th>
<th>S</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>25%</td>
<td>33%</td>
<td>23%</td>
<td>19%</td>
</tr>
<tr>
<td>French</td>
<td>28%</td>
<td>34%</td>
<td>13%</td>
<td>25%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixation Length</th>
<th>P±M</th>
<th>P</th>
<th>S</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>21%</td>
<td>33%</td>
<td>25%</td>
<td>22%</td>
</tr>
<tr>
<td>French</td>
<td>28%</td>
<td>31%</td>
<td>11%</td>
<td>29%</td>
</tr>
</tbody>
</table>

Table 3. English and French fixation count and length for all AoI with all 'into' and 'out of' stimuli

Finally, if we compare the Timeline of fixations for this particular item type in English (Figure 5) vs. French (Figure 6), some differences across language groups can be observed. In the English group, fixations are relatively regular during the entire duration of the stimuli with some variations at different points of processing: no preference for P or P±M at the onset and end of the stimuli, and a punctual preference for P during processing (500-1500 ms and 2500-3500 ms), which accounts for the P-focus observed above in relation to fixation counts and lengths (see Table 3). In contrast, French speakers pay more attention to P areas during the first half of processing, i.e. from the very beginning of processing at 0-500 ms (50 % P and 22% P±M) and until 1500-2000 ms (48% P and 30% P±M), but these speakers show no difference in attention to P or P±M areas thereafter.

Figure 5. Timeline of English fixations during 'into' and 'out of' Video stimuli
Cartoon stimuli. Table 4 shows the length and count of fixations for all cartoon stimuli. The data with these stimuli show a language effect with respect to both measures. In the English group fixations are equally frequent and long for $P$ or $P\pm M$ areas, whereas the French group shows more and longer fixations for $P$ areas.

<table>
<thead>
<tr>
<th>Fixation Count</th>
<th>$P\pm M$</th>
<th>$P$</th>
<th>$S$</th>
<th>$G$</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>29%</td>
<td>31%</td>
<td>20%</td>
<td>20%</td>
</tr>
<tr>
<td>French</td>
<td>26%</td>
<td>36%</td>
<td>19%</td>
<td>20%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixation Length</th>
<th>$P\pm M$</th>
<th>$P$</th>
<th>$S$</th>
<th>$G$</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>31%</td>
<td>30%</td>
<td>20%</td>
<td>19%</td>
</tr>
<tr>
<td>French</td>
<td>26%</td>
<td>33%</td>
<td>19%</td>
<td>22%</td>
</tr>
</tbody>
</table>

Table 4. English and French fixation count and length for all AoI with all Cartoon stimuli

In addition, the Timeline of fixations for all cartoon items in English (Figure 7) and in French (Figure 8) shows a further more qualitative language effect with respect to the on-line unfolding of eye-movements. English speakers show a preference for $P$ areas at the onset of the stimuli (40% $P$ vs. 20% $P\pm M$ at 0-500 ms only), but no preference for $P$ or $P\pm M$ areas can
be observed thereafter and during the entire processing of these stimuli, as was the case with the video stimuli. In contrast, French speakers pay overall more attention to $P$ areas during both early and late processing. In particular, greater attention allocation to $P$ can be observed in this group at stimulus onset, from 0-500 ms (40% $P$ vs. 21% $P\pm M$) and up to 2000 ms (35% $P$ vs. 26% $P\pm M$), as well as again towards the end of the stimuli at 6000-6500 ms (42% $P$ vs. 24% $P\pm M$).

Figure 7. Timeline of English fixations during the cartoon stimuli

Figure 8. Timeline of French fixations during the cartoon stimuli
Table 5 shows the number and length of fixations for the set of items that showed ‘up’ and ‘down’ motion. Overall these data show again more and longer fixations on $P$ areas as compared to $P\pm M$ areas in both language groups. However, this difference is quite marginal in the English group and much more striking among French speakers who clearly look at $P$ more frequently and during more time as compared to $P\pm M$.

<table>
<thead>
<tr>
<th>Fixation Count</th>
<th>P±M</th>
<th>P</th>
<th>S</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>25%</td>
<td>32%</td>
<td>20%</td>
<td>23%</td>
</tr>
<tr>
<td>French</td>
<td>19%</td>
<td>39%</td>
<td>18%</td>
<td>23%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixation Length</th>
<th>P±M</th>
<th>P</th>
<th>S</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>27%</td>
<td>32%</td>
<td>19%</td>
<td>22%</td>
</tr>
<tr>
<td>French</td>
<td>19%</td>
<td>37%</td>
<td>18%</td>
<td>26%</td>
</tr>
</tbody>
</table>

Table 5. English and French fixation count and length for all AoI with ‘up’ and ‘down’ stimuli

Finally, Figures 9 and 10 show a more detailed on-line analysis of the Timeline of fixations from the beginning to the end of all ‘up’ and ‘down’ cartoon items. As was the case with video stimuli, the English group shows a preference for $P$ areas at the onset of these cartoon stimuli (41% $P$ vs. 22% $P\pm M$), but no preference for $P$ or $P\pm M$ thereafter during their entire processing of the stimuli. In sharp contrast, French speakers clearly pay more attention to $P$ areas during the entire duration of the stimuli: attention is focused on $P$ areas from stimulus onset to end with peaks at 1000-1500 ms (47% $P$ vs. 14% $P\pm M$) and at 5000-6500 ms (45% $P$ vs. 22% $P\pm M$).
In summary, the eye-movement data for the video clips show overall no difference in attention allocation between the two languages. Fixations with one item type (‘into’ and ‘out of’) show a preference for $P$ in both language groups and a tendency for French speakers to
focus more on $P$ than English speakers during their early processing of the stimuli. In contrast, the data for the cartoons show a stronger difference between the two language groups. In particular, the French group allocates more attention to $P$ areas during their early and late processing of all stimuli, while the English group shows no striking preference for different areas, despite some $P$-focused fixations at the very onset of the stimuli. With one item type (‘up’ and ‘down’) this language effect is yet more striking, showing a clear preference for $P$ in the French group during the entire processing of these stimuli but again only a slight onset $P$ preference for English.

3.2. Non-verbal categorization task

*Categorical preferences*

Figure 11 shows the percentages of non-verbal responses that were based on *Path* vs. *Manner* criteria in each language in the non-verbal categorization task. Student t-tests showed a main criterion effect for French participants ($t(13)=3.72, p<0.005$) whereby they preferred *Path* choices to *Manner* choices. In contrast, there was no significant criterion effect for English participants ($t(13)=1.05, p>0.3$), nor any significant language effect on *Manner* responses ($t(26)=1.55, p>0.1$). Thus, the comparison of *Manner* responses across languages showed no significant difference.
Eye-movements during the non-verbal categorization task

The results concerning eye-movements during all main triads of the non-verbal categorization task also show no important language difference. With respect to fixation count, 46% of the English fixations were in $m$ areas and 54% in $p$ areas, as compared to 44% and 56% respectively in French. However, a further look at fixation lengths (Figure 12) shows that French speakers’ fixations were longer for $p$ areas whereas English speakers show no difference between $m$ and $p$ with respect to the length of their fixations.
3.3. Verbal categorization task

Categorical preferences

Figure 13 shows the percentages of Path vs. Manner criteria responses in the verbal categorization task in each language. The Student t-tests showed that there was a main criterion effect both within and across languages. English speakers showed a significant preference for Manner choices ($t(13)=3.59$, $p<0.005$), whereas French speakers continued to show a significant preference for Path as in the non-verbal categorization task ($t(13)=2.19$, $p=0.04$). A comparison of Manner choices across languages in this task shows a significant language effect ($t(13)=4.17$, $p<0.001$) with an advantage for English over French. In addition, Manner choices were more frequent in both language groups when categorization was verbal (Figure 13 below) rather than non-verbal (Figure 11 above).
The analysis of the eye-movements during the verbal categorization task showed a difference between the two language groups in terms of both the number and length of the fixations. French speakers paid more attention to $p$ areas (54%) than to $m$ areas (46%) in this task, whereas English speakers continued to show no preference across these two areas of interest (49% and 51% respectively). A further look at the lengths of fixations (Figure 14) shows that English fixations were longer on $m$ areas as compared to $p$ areas, whereas French fixations were longer on $p$ than on $m$ areas.
4. DISCUSSION

On the basis of two different sets of visual stimuli showing voluntary motion events (video clips and animated cartoons), we examined response patterns across languages (French and English) and tasks (verbal and non-verbal) in order to determine whether language effects would occur and whether they would be observed beyond language use. More specifically, we tested the performance of 28 participants (14 per language) in three tasks: a production task (describing motion events), a non-verbal categorization task (categorizing events in the absence of any relevant linguistic information), and a verbal categorization task (categorizing events when linguistic targets were included among the stimuli), all of which were coupled with an eye-tracking paradigm to measure attention allocation.

We tested several hypotheses. The null hypothesis predicts no major language effects in any task, whereas several relativity hypotheses predict such effects to various degrees: in all tasks (strong hypothesis), only in production (weak hypothesis), or also in non-verbal
responses but only when the task requires processing some linguistic information as was the case in our verbal categorization task ([moderate hypothesis]).

The production data showed that participants’ linguistic performance differed depending on their language, as measured by three aspects of the information expressed in their responses: focus, locus, and density. English speakers used compact and dense structures that expressed both Manner and Path information (MP-responses) and in which they systematically encoded Manner in verbs and Path in other devices (mostly particles and prepositions). In contrast, French speakers mostly focused only on Path information (P-responses) that was expressed in the verb, and they either provided less information about Manner in the periphery of the sentence or did not use any other devices in the verbal network to express this information. As a consequence, the density of semantic information in English was systematically higher in comparison to French. These patterns directly follow from language-specific factors that typologically differentiate English and French as Satellite-framed vs. Verb-framed languages respectively.

Turning now to eye-movements, the data showed the following results. Eye-movements based on all videos and animated cartoons during the production task showed no difference across languages with respect to the proportion of time spent looking at the P and P±M areas, with the only exception of a tendency for French fixation counts to be P-focused. Surprisingly, further analyses revealed differences across the two sets of stimuli (videos and cartoons) and across item types within each set. With respect to the data concerning video clips, a first analysis based on all video stimuli showed no significant difference across languages. Nonetheless, one sub-set of video stimuli (‘into’ and ‘out of’ motion events) revealed a difference in how eye-movements unfolded during the task in the two language groups. English speakers paid overall more attention to Path, but their fixations to P and P±M
followed a parallel pattern during the processing timeline. In contrast, French speakers began processing the stimuli by focusing more attentively on *Path* and they continued to do so until the middle of the processing timeline. With respect to the data concerning cartoons, a first analysis based on all of these stimuli revealed that French eye-movements focused on *Path* areas from stimulus onset until the end of the processing timeline. In contrast, English fixations showed no preference *P* or *P±M*, with the exception of a brief *P* preference only at the onset of the stimuli. In addition, this language effect can be observed even more clearly with one sub-set of stimuli (‘up’ and ‘down’ motion) that shows a strong preference for *P* throughout processing in the French group but a *P* preference only at stimulus onset in the English group.

Several factors could account for the complex patterns that were observed in subjects’ allocation of visual attention during their exploration of the scenes. First, universal factors may account for some of the similarities that were found across the two language groups. In particular, notwithstanding differences in degree across languages and across items types (see below), both groups overall paid more attention to *Path* (*P* areas) than to areas that included Manner information (*P±M*).\(^4\) This pattern is not predicted by any version of the relativity hypothesis, which predicts generally more attention to *Path* in the French group and either more attention to Manner or equal attention to *Path* and Manner in 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 concerning the locations and changes of location of protagonists that are essential for the interlocutor to reconstruct the spatial universe described in discourse (see Tamy, 2000).

Second, language-specific factors invited speakers of the two language groups to pay

\(^4\) Recall that the *P±M* areas were mixed so that fixations on these areas could also include some *Path* information even when the focus was on Manner (also see Note 3).
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. This difference was predicted by some of the hypotheses discussed above. Thus, given the typological properties of English and French, all versions of the relativity hypothesis predict that different components should be most accessible in each language. Furthermore, at least two versions of this hypothesis (strong and moderate) predict that such language-specific factors should also (at least partially) affect speakers’ attention allocation. This language effect was the clearest in analyses of the Timeline of fixations that allowed us to follow on-line the unfolding of subjects’ visual processing. Thus, in the French group P-focused fixations were most frequent throughout processing and/or at the beginning and end of processing (particularly for cartoon items, as further discussed below). In contrast, no such pattern was observed in the English group, where subjects either showed a parallel processing of Path and Manner throughout or a greater focus on Path only at stimulus onset.

Our results also indicate the impact of yet a third factor related to our stimuli. In particular, subjects’ attention allocation differed to some extent across the two sets of stimuli in our study (videos vs. cartoons) and across item types within each set (‘into-out of’ videos, ‘up-down’ cartoons). These differences raise some methodological questions concerning the nature of these stimuli. Video stimuli provided highly controlled natural scenes of human voluntary motion events in fixed settings and with short duration (4 seconds in all cases). In comparison, cartoons were longer on the average and varied somewhat in duration (mean 6.5 seconds). Furthermore, they provided more vivid and varied stimuli that showed different types of protagonists (humans and animals) in diverse setting (e.g., tree, house, plant, lake, river, road) and in the context of a little scenario involving a goal on the part of the protagonist (e.g., going up a tree to get a banana). Cartoon stimuli therefore invited participants to organize their response in the form of a real narrative comprising a plot line
with a setting, a goal on the part of the protagonists, main motion events, and an ending, whereas videos elicited shorter and more descriptive responses. As a result, it is likely that, when narrating cartoons, speakers planned their verbalizations with particular attention to the goal and end of the story plot line, whereas they had no such expectations in the case of the video clips.

Analyses in progress further examine the data for particular items in order to determine whether further differences occurred as a function of the specific properties of paths. For example, a preliminary comparison of ‘up’ vs. ‘down’ cartoons suggests that the attention of French speakers was more focused on path information towards the end of ‘up’ items (possibly corresponding to a focus on reaching the goal), while it was more focused on this information at stimulus onset for ‘down’ items (possibly corresponding to a focus on source location). Note that this effect seems to occur particularly in French, perhaps because these speakers are more sensitive to Path information, given the properties of their language.

The non-verbal categorization data showed that French participants preferred Path to Manner as their categorization criterion, while English participants showed no significant preference for Path or Manner. In addition, the data showed no significant language effect with respect to the frequency of Manner responses in French vs. English. These results partially follow from the typological properties of these two languages. The fact that Path information is typically lexicalized in French verbs leads speakers to pay more attention to this information component and thus to show a preference for the Path criterion. In contrast, English participants showed no preference for either Path or Manner criteria, a result which is in line with the fact that English relies on compact and dense structures that systematically contain both types of information, thereby making them equally salient (Manner lexicalized in the verb together with Path information in other devices). However, this language effect did
not result in a significant difference between English vs. French *Manner* responses. Although *Manner* responses tended to be more frequent in English than in French, this difference was not significant.

The eye-movement data during the non-verbal categorization task also showed no important language difference with respect to the number of fixations on the two areas of interest that corresponded to the categorization criteria. However, further analyses examining the time spent looking at the *p* and *m* components showed that fixations were longer for *p* areas in the French group, but did not differ in length for *p* or *m* components in the English group. These findings are in line with the hypothesis of an effect that might result from the typological pattern of each language: French speakers tend to spend more time on Path, which is lexicalized in the verb and therefore salient than Manner, which is expressed only peripherally (if at all). In contrast, English pay as much attention to both components, which are both as salient and as frequently expressed.

Finally, in the verbal categorization task *Manner* choices were more frequent for both language groups as compared to the non-verbal categorization choices. In this verbal version of the categorization task, English speakers also showed a significant preference for *Manner* choices, whereas no significant difference for *Path* vs. *Manner* responses was observed in the non-verbal version of this task. In contrast, French speakers continued to show a significant preference for *Path* in the verbal version, as they did in the non-verbal version. Finally, *Manner* choices showed a significant language effect such that English speakers chose the *Manner* criterion significantly more than French speakers. Thus, regardless of language, the introduction of a linguistic stimulus (target sentence) increased the saliency of *Manner* information in the categorisation task resulting in a higher focus on *Manner* in both language groups. Nonetheless, French speakers continued to choose *Path* significantly more often than
Manner as their categorization criterion, following in this respect the pattern of their language (Path lexicalized in the verb, Manner implicit or peripheral). As for the eye-movement data during the verbal categorization task, they confirmed once again the preference for Path in French (more and longer p fixations). In the English group, despite the fact that fixation count showed no preference for p or m, the length of m fixations was more important as compared to p fixations.

In summary, the overall findings from the production task first suggest a clear language effect in verbal responses, but only a partial language effect in the eye-movements data. Second, the findings from the non-verbal categorization task show no language effect on subjects’ categorical choices or attention allocation during the task. Finally, the findings from the verbal categorization suggest a clear language effect on subjects’ non-verbal preferential choices but not in their allocation of visual attention as measured by eye-tracking.

Let us now return to the set of hypotheses that were tested in the present study. Three main points emerge from our results in this respect. First, our findings do not support the null hypothesis according to which no major language effects should be found in any task. The linguistic responses in the production task show striking cross-linguistic differences in how participants described motion events. These differences are three-fold: speakers do not choose to express the same information when describing motion events (focus on Path in French, but on both Path and Manner in English), resulting in utterances of varying semantic density (higher density in English than French), and they use different structures to organize the selected information at different loci (Path verbs and either peripheral or no Manner in French, Manner verbs and Path satellites in English). Note that languages provide several other grammatical options to describe the same events, but that language-specific factors invite speakers to rely on those structures that are most available in their language because
they constitute the most typical and accessible ways of organizing information.

Second, the other tasks in our study were meant to determine whether such language effects in production could have a deeper impact on event construal beyond language use. In this respect, comparisons across tasks support neither the *strong* nor the *weak relativity* hypotheses, according to which language-specific preferences should occur either in all tasks (*strong version*) or only in verbal tasks (*weak version*). Thus, the results show language effects in some tasks but not in others, and they therefore do not support the strong version of the relatively hypothesis. Furthermore, they show language effects with some non-verbal measures (eye-movements, categorization) and therefore do not support its weak version.

Third, taken together, the results that emerge from this study show a more complex pattern that cannot be explained by all-or-none hypotheses and that must be interpreted with care. They indicate that typological language properties clearly affect how speakers express motion (production task), but also that such language-specific factors can generate different categorical choices in verbal and non-verbal categorization tasks and partially constrain attention allocation during participants’ exploration of some types of visual stimuli. Our results therefore support the *moderate* version of the relativity hypothesis, which predicts language effects on verbal responses, as well as on non-verbal responses when relevant linguistic information is involved during information processing.

5. CONCLUSION

The present paper explored the language-thought interface by means of combined off-line and on-line measures in different verbal and non-verbal tasks and across two typologically different languages, English and French. The aim of this research was to test various hypotheses concerning the presence/absence and strength of language effects on how speakers
construct representations of motion events in speech and beyond language use. The findings show clear differences in speakers’ linguistic representations that follow from the typological status of their language as Verb-framed vs. Satellite-framed, but they also show differences in their non-verbal responses (eye movements, categorization). It is clear that multiple methodologies, as well as data based on more languages and with larger groups of participants will be necessary in future research. The findings that emerge from this research do not support any simplistic hypothesis and indicate the need for the formulation of more precise and subtle views of the relationship between language and cognition.
REFERENCES


Flecken, M. (submitted). What native speaker judgements tell us about the grammaticalization of an aspectual marker of ‘ongoingness’ in Dutch.


Slobin, Dan I. (1996b). From “thought and language” to “thinking for speaking”. In Gumperz & Levinson : 70-96.


