

# Multilingual learning and cognitive restructuring: The role of audiovisual media exposure in Cantonese–English–Japanese multilinguals’ motion event cognition

International Journal of Bilingualism

1–18

© The Author(s) 2022



Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/13670069221085565

[journals.sagepub.com/home/ijb](https://journals.sagepub.com/home/ijb)Yi Wang\*  and Li Wei 

University College London, UK

## Abstract

**Aims and objectives/purpose/research questions:** Can learning additional languages affect what we perceive to be similar events? The current study explores the impact of learning a second language (L2) and a third language (L3) on how motion is categorized in functional Cantonese–English–Japanese multilinguals. Specifically, it examines the extent to which L1 speakers of Cantonese (equipollent-framed) with an L2-English (satellite-framed) and an L3-Japanese (verb-framed) restructure their lexicalization and conceptualization of voluntary motion through audiovisual media exposure to the target language.

**Design/methodology/approach:** A total of 150 participants were recruited and divided into five groups: three groups of monolinguals in Cantonese, English, and Japanese as well as Cantonese–English bilinguals and Cantonese–English–Japanese multilinguals. Participants were given a linguistic encoding task and a non-linguistic similarity judgement task.

**Data and analysis:** Mixed-effects modelling was used to compute participants’ encoding patterns and categorical preferences, as well as the correlation between audiovisual media exposure and the degree of cognitive restructuring.

**Findings/conclusions:** Multilinguals’ L1-based verbalization and categorization showed reverse transfer from both L2 and L3. The degree of cognitive restructuring was modulated by audiovisual media exposure to TV watching in English and Japanese.

**Originality:** This study extends the thinking-for-speaking hypothesis from bilingualism to multilingualism and explores the language–thought interface through reverse cross-linguistic influence. It focuses on how functional Cantonese–English–Japanese multilinguals with partially overlapping language systems encode and gauge similarity of voluntary motion in their L1, which is a rarely studied language combination.

**Significance/implications:** Looking at the cognitive effects of additional language learning can shed light on the mechanism of cognitive restructuring in the thinking-for-speaking perspective, and inform the language learning question of how learners integrate both linguistic

---

\*Yi Wang is also affiliated to School of Psychology, HSE University, Moscow, Russia and School of English, Communication and Philosophy, Cardiff University, Cardiff, UK

## Corresponding author:

Yi Wang, UCL Centre for Applied Linguistics, University College London, 20 Bedford Way, London WC1H 0AL, UK.

Email: [dtvnywa@ucl.ac.uk](mailto:dtvnywa@ucl.ac.uk)

and non-linguistic experience to recalibrate their cognitive dispositions when given sufficient multimodal input.

### **Keywords**

Cognitive restructuring, Cantonese–English–Japanese multilinguals, audiovisual media exposure, thinking-for-speaking, voluntary motion events

## **Introduction**

Does the language we speak influence the way we think? This age-old question has generated extensive debates in past decades, and received renewed interest as a number of new research paradigms have evolved that allow addressing this issue empirically. Experimental evidence shows that language can exert both temporary and immediate, or habitual and durable, effects on various cognitive processes such as perception, recognition memory, visual discrimination, and categorization, in a flexible and context-dependent manner (Filipović, 2018; Papafragou et al., 2008; Slobin, 1996; Wolff & Holmes, 2011). For example, language effects are most likely to arise when linguistic labels are explicitly utilized during thinking, or as a strategy to solve a subsequent cognitive task. This process, termed thinking-for-speaking (Slobin, 1987, 1996), emphasizes the language effects on online thinking when speakers are actively engaged in language-driven activities. There is converging evidence that the thinking-for-speaking effects are language-specific (Flecken et al., 2015; Gennari et al., 2002; Papafragou et al., 2008), especially when the task does not prevent the practice and strategic use of language during task performance (Athanasopoulos, Bylund, et al., 2015; Trueswell & Papafragou, 2010).

Research on the effect of language learning on cognitive processing starts with monolinguals of typologically contrastive languages, although recent studies have extended the scope of interest to bilinguals and various types of L2 learners (Athanasopoulos, Damjanovic, et al., 2015; Brown & Gullberg, 2011; Cook & Bassetti, 2011). Extending the language-and-thought research to the domain of bilingualism offers researchers a unique opportunity to explore how different languages are reconciled within the same mind. Empirical evidence has demonstrated that learning a new language may bring about changes of the entire cognitive outlook, a process called conceptual transfer or cognitive restructuring (Jarvis & Pavlenko, 2008; Pavlenko, 2011). Studies to date have demonstrated that learning an L2 may give rise to the restructuring of L1-based categories, depending on various extra-linguistic factors, such as age of acquisition (Bylund, 2009; Lai et al., 2014), language proficiency (Athanasopoulos, Damjanovic, et al., 2015; Park, 2019), and audiovisual media exposure (Bylund & Athanasopoulos, 2014, 2015).

## **The role of audiovisual input in cognitive restructuring**

The connections between cognitive restructuring and the language-and-thought research are based upon the observation that concepts are multimodal, dynamic and highly intercorrelated across modalities (Athanasopoulos, Damjanovic, et al., 2015; Bylund & Athanasopoulos, 2014; Casasanto, 2008). For example, cross-linguistic differences in construal of motion have been found not only in speech, but also speech-accompanying gestures (McNeill, 2005; Kita & Özyürek, 2003). For instance, speakers of S-languages tend to produce less manner-related gestures compared with V-language speakers (Brown & Chen, 2013). To successfully acquire novel concepts, learners need

to rely on a variety of semiotic resources from the input, be they visual, textual, graphical, or auditory (Athanasopoulos, Damjanovic, et al., 2015; Bylund & Athanasopoulos, 2014).

Recent studies have documented that among different types of exposure, audiovisual input is found to enhance L2 vocabulary and grammar development (Vanderplank, 2010). Research on multimodal L2 input shows that television with subtitles or captions enhances word learning and grammar constructions (Cintrón-Valentín et al., 2019; Montero Perez et al., 2018; Peters et al., 2016; Rodgers & Webb, 2017; Vanderplank, 2010). Unlike pure written or audio exposure (i.e., reading or radio listening), television has a multimodal nature that combines visuals (i.e., dynamic scenes) with sounds, and sometimes on-screen texts (Rodgers & Webb, 2017). Because human cognition utilizes separate channels for processing visual and verbal information, combining visuals with verbal input can generally facilitate information processing (Paivio, 2014). This is well supported by Mayer's (2001) Cognitive Theory of Multimedia Learning, which postulates that 'people learn better from words and pictures than from words alone' (Mayer, 2001, p. 63). Thus, audiovisual input has been regarded as a powerful resource in learners' L2 development (Montero Perez et al., 2018; Peters et al., 2016; Vanderplank, 2010).

Despite the role that audiovisual input plays in grammatical and lexical development, the generalization of such effects to the cognitive domain, such as motion events, hitherto remains an open question. Bylund and Athanasopoulos (2015) took a first step in examining how different types of exposure (i.e., reading, Internet, radio, and TV) affected cognitive restructuring in L1-Swedish learners with English as a foreign language. It is found that Swedish learners exhibited L2-based categorization patterns (i.e., ongoingness) through exposure to English audiovisual media. The results indicate that continuous exposure to the multimodal input not only provides learners with language-specific ways to talk about motion, but also corresponding temporal reviewing frames. Given the multimodal nature of television, English-speaking TV programmes offer learners reduced verbal reference to endpoints when talking about motion. It also affords less visual prominence to endpoints when framing motion. Continuous exposure to these immediate temporal visual frames may lead learners to applying them when categorizing novel events.

## Linguistic encoding of voluntary motion in English, Japanese, and Cantonese

World languages differ typologically in how they select and package information about motion. Based on the semantic distribution of path, Talmy (2000) divided languages into two distinct categories. Satellite-framed languages (S-languages), such as German and English, typically encode path in the satellite (verb particles) whereas manner in the verb root, as shown in (1).

- (1) He walked across the street.

In contrast, for verb-framed languages (V-languages), such as Spanish and Japanese, path is encoded in the main verb whereas manner is optional. For example, Japanese either conflates path in the main verb, without mentioning manner at all (2a), or encodes manner in subordinate forms (the *-te* conjunctive marker) (2b), depending on whether the interlocutor would like to stress manner or not. In addition, Japanese, and V-languages in general, follow a boundary-crossing constraint that manner cannot be conflated via verbs when the target event denotes a categorical change of location (Slobin & Hoiting, 1994).

- (2)a. 彼 は 三階 に 上がった  
 S/he TOP third floor to ascend. PST  
 ‘S/he ascended to the third floor’.
- b. 彼 は 歩いて 道路 を 横切った  
 S/he TOP walking<sub>[Manner]</sub> road ACC cross PST  
 ‘S/he crossed the road walking’.

This S- and V-language dichotomy has been fruitful in analysing most Indo-European languages, but cannot accommodate serial-verb languages, such as Chinese and Thai, where manner and path are encoded in compound forms. Slobin (2004) introduced a third type called equipollent-framed language, where ‘both Manner and Path are expressed by equipollent elements, that is, elements that are equal in formal linguistic terms, and appear to be equal in force or significance’ (Slobin, 2004, p. 226).

Cantonese is a serial-verb language and serial-verb constructions in Cantonese can take up two or more components (Matthews & Yip, 2011). In (3), 翻 (return) and 入 (enter) are path of motion and encoded via verb complements (satellite-framing strategy). However, in (4) and (5), the same elements are encoded via independent verbs (verb-framing strategy). In fact, Cantonese has a rich set of path verbs and both ‘manner + path’ and ‘path-only’ constructions are used frequently in discourse (Wang & Li, 2019; Yiu, 2013, 2014). According to Slobin (2004), the typological classification of different languages in the motion domain is based upon the lexical availability and characteristic expressions of manner and path. In this regard, Cantonese does not pattern with either satellite- or verb-framed languages but rather exhibits the typological properties of both language types.

- (3) 個男仔 行 咗 翻 入 辦公室  
 A boy walk ASP return enter office  
 ‘A boy walked back into the office.’
- (4) 個男仔 翻 咗 辦公室  
 A boy return ASP office  
 ‘A boy returned to the office.’
- (5) 個男仔 入 咗 辦公室  
 A boy enter ASP office  
 ‘A boy entered the office.’

## Cognitive restructuring of motion events in language learning

Conceptual transfer has been a central topic in thinking-for-speaking research. Most studies focus on whether learners can restructure their L1-based thinking-for-speaking in accordance with the target language. Some studies report that learners may still rely on L1-based patterns for speaking during L2 learning (Daller et al., 2011; Hendriks & Hickmann, 2015; Larrañaga et al., 2012), while others show that L2 learners can learn to rethink for speaking (Hickmann & Hendriks, 2010; Ji & Hohenstein, 2014). While the thinking-for-speaking phenomenon has been intensively examined with diverse L2 populations, few studies report evidence for reverse transfer from the L2 to the L1 (Wang & Li, 2019; Aveledo & Athanasopoulos, 2016; Brown & Gullberg, 2008, 2011; Hohenstein et al., 2006). Aveledo

and Athanasopoulos (2016) examined the effect of L2 learning on L1 descriptions of voluntary motion with early Spanish–English bilinguals. Results showed that bilinguals used more manner verbs and fewer path verbs in L1 descriptions compared with Spanish monolinguals, suggesting a reverse transfer from the L2. Similar findings were reported by Wang and Li (2019) that early Cantonese–English bilinguals encoded manner more frequently compared with Cantonese monolinguals in L1 descriptions, indicating a cognitive shift towards L2-based concepts.

Moving beyond language use, it is questioned whether language-specific ways of speaking lead to particular ways of thinking. Some studies suggest that the L1-based thinking is stable and resistant to change (Aveledo & Athanasopoulos, 2016; Filipović, 2018), while others report changes in the conceptualization patterns as a result of bilingual experience (Athanasopoulos, Bylund, et al., 2015; Bylund & Jarvis, 2011; Kersten et al., 2010; Lai et al., 2014; Montero-Melis et al., 2016). For example, Kersten et al. (2010) examined how Spanish–English bilinguals classified novel events with a supervised learning paradigm. Results showed that bilinguals with early exposure to English performed as well as English monolinguals in sorting manner-based events when instructed in an English context than a Spanish context, indicating that thinking after immediate language use is susceptible to recent linguistic experience. Similarly, Montero-Melis et al. (2016) explored how Swedish adult learners of L2 Spanish categorized caused motion through recent linguistic priming. Participants were given L2-based sentences with various degrees of manner and path salience right before making their similarity arrangements. Results suggested that participants in a manner-primed condition were more likely to base their judgements on ‘same-manner’ than those in path-primed conditions, demonstrating a transient thinking-for-speaking effect. The findings echo with Wang and Li (2021a), who explored how Cantonese–English–Japanese multilinguals lexicalized and conceptualized caused motion by manipulating different language contexts. The authors found that multilinguals patterned with Japanese monolinguals in the processing efficiency of caused motion (i.e., reaction time) when their access to L3-Japanese was boosted during decision-making, thus constituting clear evidence for L3-biased cognitive restructuring. However, it still remains unclear as to whether, and to what extent, patterns of cognition acquired through additional languages could bring forth conceptual changes in one’s L1-based patterns for thinking and speaking, a process called reverse or backward transfer. Built on this, the current study takes a further step in investigating conceptual transfer in the reverse direction (i.e., both L2 to L1 and L3 to L1), as well as how language proficiency and audiovisual media input affect this process. The specific research questions are formulated as follows:

1. How do Cantonese–English–Japanese multilinguals encode spatial components of manner and path in their L1 compared with Cantonese–English bilingual and monolingual controls?
2. How do Cantonese–English–Japanese multilinguals represent spatial components of manner and path compared to Cantonese–English bilingual and monolingual controls?
3. To what extent do language proficiency and audiovisual media exposure modulate cognitive restructuring in the L1?

## Method

### Participants

Altogether 150 university students were recruited and divided into five groups ( $N=30$  each group). Monolinguals of Cantonese ( $M_{age}=22.1$ ,  $SD=2.7$ ), English ( $M_{age}=23.7$ ,  $SD=1.9$ ), and Japanese ( $M_{age}=24.6$ ,  $SD=2.3$ ) came from local universities of China, the United Kingdom, and Japan.

**Table 1.** Summary of multilinguals' different types of exposure.

Measure	L1 – Cantonese	L2 – English	L3 – Japanese
Radio listening	0.10 (0.24)	0.03 (0.13)	0.18 (0.33)
Watching TV	1.28 (0.77)	1.68 (1.01)	1.98 (1.35)
Internet/social media	1.62 (1.08)	2.27 (0.99)	2.62 (1.63)
Book/newspaper/magazine	0.82 (0.84)	1.07 (0.77)	2.65 (1.29)
Overall	3.82 (2.38)	5.05 (2.16)	7.43 (3.15)

Note. Exposure to each language was estimated by hours/day within the last 6 months.

Following the common practice in bilingual research (Athanasopoulos, Bylund, et al., 2015; Park, 2019), monolinguals are defined as speakers with minimal exposure and limited proficiency to any foreign or additional language(s). They use their L1 as the dominant language in daily interactions, and regard themselves as monolinguals rather than bilinguals. Cantonese–English bilinguals ( $M_{age} = 20.7$ ,  $SD = 2.1$ ) and Cantonese–English–Japanese multilinguals ( $M_{age} = 21.2$ ,  $SD = 1.8$ ) were from Hong Kong (HK) where both Cantonese and English are the official languages. According to HK language education policy, children should normally start L2-English from an average age of three. This learning continues throughout school years and many attend English-medium schools. They should achieve a high command of English at tertiary level, and many continue learning a third language as either Major or Minor (Table 1).

Following previous studies (Athanasopoulos, Damjanovic, et al., 2015; Montero-Melis et al., 2016), participants' language proficiency was measured by a standardized language background questionnaire (Li et al., 2014). Participants self-reported their current proficiency in all languages they know based on a seven-point scale where 7 is the maximum rating. Given the Common European Framework of Reference for language (Council of Europe, 2011), bilinguals ( $M = 6.23$ ;  $SD = 0.43$ ) and multilinguals ( $M = 6.15$ ;  $SD = 0.41$ ) achieved an advanced level of proficiency in L2-English, and multilinguals reached an upper intermediate level (B2) in L3-Japanese ( $M = 5.71$ ;  $SD = 0.48$ ). Statistical analysis confirmed no difference was found between bi- and multilinguals' proficiency in English,  $t(58) = 0.69$ ,  $p = .492$ .

Following Bylund and Athanasopoulos (2015), multilinguals' language exposure was measured by self-reported hours in doing a list of daily activities, such as radio, television, books/magazines/newspapers, and Internet use. These are out-of-classroom activities and serve as an additional input in the learning process (Rodgers & Webb, 2017).

## Materials

**Task 1: linguistic encoding of voluntary motion.** The linguistic encoding task had a total of 54 dynamic stimuli of voluntary motion, including 36 test items and 18 control items. Each animation was 6 seconds long. Following Hickmann and Hendriks (2010), the test items depicted a boy performing voluntary motion with various types of manner and path, while the control items minimized path but highlighted manner of motion only. A diverse range of manner verbs were included, such as low-manner verbs (walk), general manners (run, jump, crawl, march, hop, swim), and manner with instruments (cycling, skating, surfing and scootering), combined with six different types of path falling into two broad categories: trajectory events (up, down, away from, towards), and events with boundary-crossing (into, cross).



**Task 2: non-linguistic categorization of voluntary motion.** The non-linguistic categorisation covered 18 sets of animated videos, including 12 test triads and 6 filler items, which share the same content with the stimuli in Task 1. For the test items, each triad contains three events: a target event (e.g., A boy runs into a room) and two alternates with manner and path as the contrast of interest. The manner-match alternate shared the same manner with the target but differed in path (e.g., A boy runs out of a room). In contrast, the path-match alternate had the same path with the target but differed in manner (e.g., A boy crawls into a room). To keep manner versus path as the only contrast of interest, other semantic components, such as Figure (i.e., the moving entity), Ground (i.e., the reference point for the moving entity), and Goal (i.e., the endpoint of the moving entity) remain consistent across each triad. To mask the contrast of interest and distract participants strategically following the same patterns, six sets of filler items were used where half of them contrasted manner with Ground, while the other half contrasted path with Ground. All stimuli were horizontal motions and the direction of agent's movement (left-to-right, right-to-left) was counterbalanced across each triad. A whole list of stimuli is presented in the Supplementary File.

## Procedure

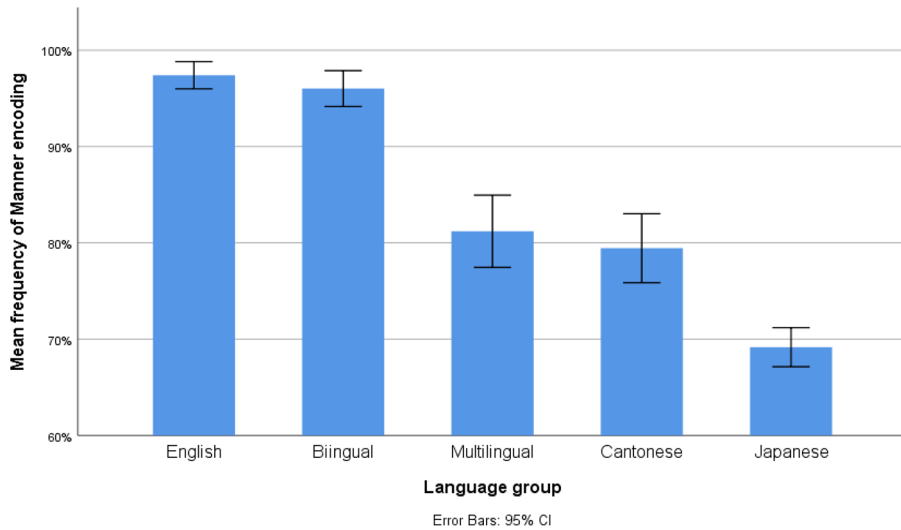
Test was conducted individually in a quiet room at relevant universities. All the stimuli were displayed and run by software SuperLab 5.0. A training session was given at the beginning of the experiment.

In the linguistic encoding task, participants first watched the video stimuli and then described 'what happened' in each video clip right after the viewing. All participants narrated in their respective L1s. The stimuli were presented in a fully randomized order across participants in different language groups.

Following other well-established studies on thinking-for-speaking (Filipović, 2018; Lai et al., 2014; Montero-Melis et al., 2016), participants were instructed to move on to a subsequent categorization task right after the linguistic encoding. This was to maximally boost language involvement during cognitive processing. Participants were informed that the video stimuli were presented in a synchronized order: the target event played first at the bottom of the screen and disappeared when completed. Then its two simultaneous alternating events started playing side by side at the top of the screen. A half-second black screen was placed between the target video and its two alternates within each triad and a one-second black screen was placed between triads. The presentation order of each triad was counterbalanced across participants. The location of manner- and path-match variant on the screen (right- or left-side) was counterbalanced across stimuli in a fixed order. Participants had to decide which variant was the most similar to the target by pressing A and L, respectively, on the keyboard. After the experiments, participants responded to a language background questionnaire.

## Data coding

The linguistic data were transcribed by L1 speakers of each language. Only test items were included for the analysis. Responses were initially segmented into clauses and coded based on the guidelines for English (Hickmann & Hendriks, 2010), Japanese (Brown & Gullberg, 2011), and Cantonese (Ji & Hohenstein, 2014). Descriptions without a specific mentioning of motion elements were removed (e.g., The sky is blue). Most of the responses (96.5%) had only one clause. Within each clause, descriptions were coded from (a) the frequency of manner and path selection; (b) the semantic



**Figure 1.** Mean proportion of manner encoding across participant groups.

distribution of manner and path (i.e., in the main verb or in the satellite); and (c) the framing strategies (satellite- or verb-framed). To establish coding reliability, we asked a second coder to re-code 20% of the data. As indicated by the Kappa Index (Cohen's kappa = .97), a high inter-coder reliability was reached.

## Results

### *Linguistic encoding of voluntary motion*

*Mean frequency of manner and path selection.* Altogether 5,400 linguistic descriptions were included in the statistical analysis. For path encoding, participants across each group reached a ceiling-level frequency (English:  $M=93.89\%$ ,  $SD=5.67\%$ ; Cantonese:  $M=95.65\%$ ,  $SD=4.71\%$ ; Japanese:  $95.55\%$ ,  $SD=4.16\%$ ; bilinguals:  $M=95.00\%$ ,  $SD=5.85\%$ ; multilinguals:  $92.50\%$ ,  $SD=5.89\%$ ), indicating that path is a core element. However, for manner encoding, a hierarchical decrease was observed across participant groups (English:  $M=97.41\%$ ,  $SD=3.79\%$ ; bilinguals:  $M=96.02\%$ ,  $SD=4.98\%$ ; multilinguals:  $M=81.20\%$ ,  $SD=10.05\%$ ; Cantonese:  $M=79.44\%$ ,  $SD=9.58\%$ ; Japanese:  $M=69.17\%$ ,  $SD=5.43\%$ ), as shown in Figure 1.

Statistical analyses were conducted in R (R Core Team, 2018) using the package *lme4* (Bates et al., 2015). To access whether the likelihood of manner or path selection significantly differed as a function of participant groups, we built two separate logistic mixed-effect models.<sup>1,2</sup> In each model, we computed the linguistic encoding of manner or path (*presence*=1; *absence*=0) as the respective binary dependent variable. The fixed effect was participant groups (dummy coded), and the random effects were random intercepts for subject and item. To estimating model variability, likelihood ratio tests were performed to compare models with fixed effects to models with random effects only (null model).

Result showed that in terms of path encoding, including participant groups did not significantly optimize the model compared with the null model,  $\chi^2(4)=7.24$ ,  $p=.124$ , suggesting that language groups were not a main effect. That is, participants across different groups had a similar likelihood to encode path in speech.



**Table 2.** The semantic distribution of manner and path in the main verb (V) or satellite (OTH).

Component	English		Bilinguals in Cantonese		Cantonese		Multilinguals in Cantonese		Japanese	
	V	OTH	V	OTH	V	OTH	V	OTH	V	OTH
Manner	93.2	8.2	74.3	21.3	55.6	27.9	52.1	40.6	14.9	54.4
Path	6.9	86.8	27.8	69.8	44.7	55.5	47.8	48.5	85.0	21.8

Note. The sum of the first two columns within each group does not always add up to 100% as manner and path can be double encoded in V and OTH at the same time.

For manner encoding, results confirmed that language groups were a significant predictor,  $\chi^2(4)=174.01$ ,  $p < .001$ . Results showed that bilinguals encoded manner more frequently than Cantonese monolinguals ( $\beta_{\text{Bilinguals-Cantonese}}=3.25$ ,  $SE=0.39$ , Wald  $z=8.43$ ,  $p < .001$ ), but patterned with English monolinguals ( $\beta_{\text{Bilinguals-English}}=-0.59$ ,  $SE=0.43$ , Wald  $z=-1.38$ ,  $p=.335$ ), indicating a reverse transfer from the L2 to the L1. By comparison, multilinguals encoded manner less frequently than bilinguals ( $\beta_{\text{Multilingual-bilingual}}=-2.99$ ,  $SE=0.38$ , Wald  $z=-7.78$ ,  $p < .001$ ), yet patterned with Cantonese monolinguals ( $\beta_{\text{Multilingual-Cantonese}}=0.26$ ,  $SE=0.33$ , Wald  $z=0.78$ ,  $p=.434$ ), indicating a reverse transfer from L3 to L1. Japanese monolinguals had the lowest frequency ( $\beta_{\text{Japanese-Cantonese}}=-1.42$ ,  $SE=0.33$ , Wald  $z=-4.23$ ,  $p < .001$ ).

**Framing-strategies and event structures.** Regarding the semantic distribution, English encoded manner in the main verb ( $M=93.24\%$ ,  $SD=7.53$ ), whereas path in the satellite ( $M=86.85\%$ ,  $SD=8.78$ ). By comparison, Japanese encoded path in the main verb ( $M=85.00\%$ ,  $SD=6.55$ ), leaving manner unexpressed, or via subordination ( $M=54.35$ ,  $SD=9.05$ ). Located in the middle, Cantonese encoded both manner ( $M=55.65\%$ ,  $SD=12.39$ ) and path ( $M=44.72\%$ ,  $SD=12.37$ ) in verb forms. Bilinguals used more manner verbs than Cantonese monolinguals ( $M=74.35\%$ ,  $SD=10.18$ ), showing an English-oriented pattern. Multilinguals patterned with Cantonese monolinguals in frequently using both manner ( $M=52.13\%$ ,  $SD=17.65$ ) and path verbs ( $M=47.78\%$ ,  $SD=17.71$ ), as shown in Table 2.

Consistent with the semantic distribution of manner and path, English predominantly used satellite-framing for motion event construction ( $M=93.06\%$ ,  $SD=7.43$ ), whereas verb-framing was hardly used, as illustrated in (6) and (7).

(6) Satellite-framing: He ran across the street. (ENG02Vol)

(7) Verb-framing: He crossed the street running. (ENG11Vol)

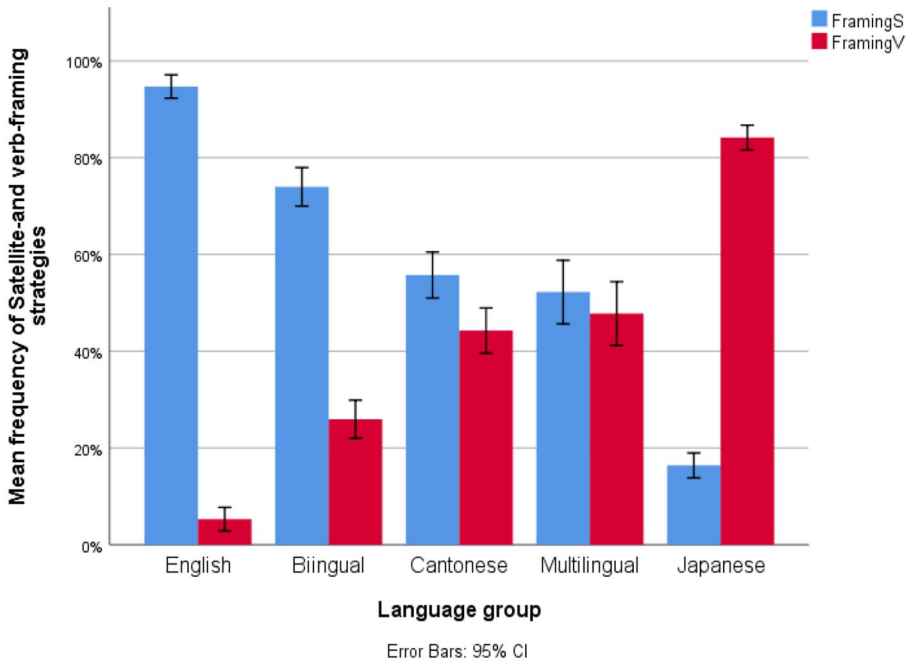
In contrast, Japanese, as a V-language, predominantly used verb-framing ( $M=84.17\%$ ,  $SD=6.85$ ), as illustrated in (8) and (9).

(8) Verb-framing: Path-only construction (by default)

男性 が 階段 を 上りました (JAP12Vol)  
 a boy TOP stairs ACC ascend. PST  
 ‘A boy ascended the stairs’.

(9) Verb-framing: Path verb + Manner subordination (a gerund form ‘-te’)

男性 が 片足で 階段 を 下りました (JAP25Vol)  
 a boy TOP with one foot stairs ACC descend. PST  
 ‘A boy descended the stairs with one foot’.



**Figure 2.** Mean proportion of satellite- and verb-framing across participant groups.

Being an E-language, Cantonese used both satellite-framing ( $M=44.25\%$ ,  $SD=12.51$ ) and verb-framing in speech production, as shown in (10) and (11). The partial overlap across the three languages may account for the reason why motion constructions in Cantonese can be restructured under the influence of learning another satellite-or verb-framed language.

(10) Satellite-framing: Manner + Path construction

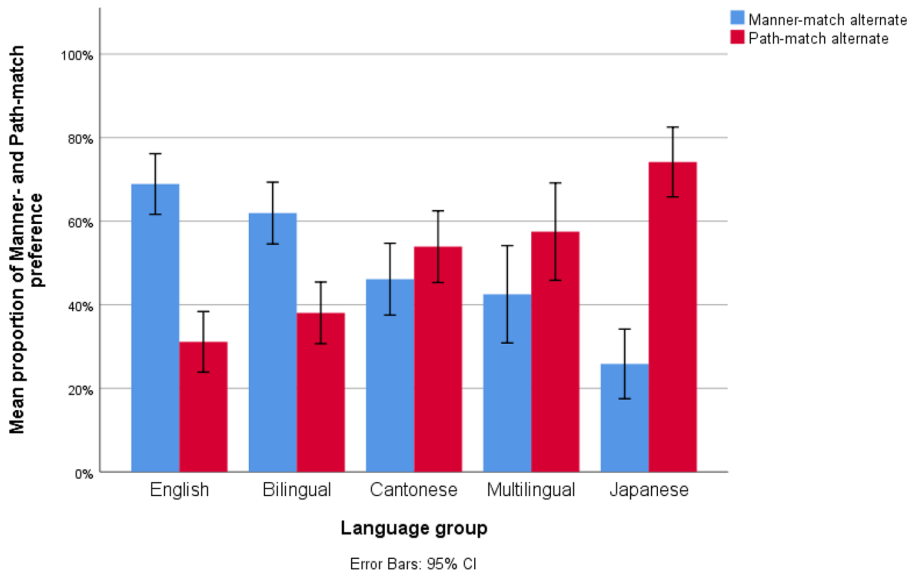
佢 跑 緊 落樓梯 (CAN03Vol)  
 He run ASP descend the stairs  
 ‘A boy is running down the stairs.’

(11) Verb-framing: Path-only construction

佢 落 咗 樓梯 (CAN05Vol)  
 He descend ASP the stairs  
 ‘He descended the stairs.’

Given various degrees of manner- and path-salience in Cantonese, bilinguals’ L1 descriptions patterned with English in using more satellite-framing ( $M=73.98\%$ ,  $SD=10.66$ ) than Cantonese monolinguals, as shown in Figure 2, indicating a possible reverse transfer from L2. By comparison, multilinguals’ L1 descriptions ( $M=52.22\%$ ,  $SD=17.62$ ) patterned with Cantonese in using both satellite- and verb-framing frequently, suggesting that learning another verb-framed language (L3-Japanese) seems to counterbalance the potential impact from the L2-English to L1.

*Non-linguistic similarity judgements of voluntary motion.* To further investigate how participants made their similarity judgements, the proportion of manner-match preferences was compared across groups (English:  $M=68.89\%$ ,  $SD=19.44\%$ ; bilinguals:  $M=61.94\%$ ,  $SD=19.78\%$ ; Cantonese:



**Figure 3.** Mean percentage of manner- and path-match preferences across language groups.

$M=46.11\%$ ,  $SD=23.03\%$ ; multilinguals:  $M=42.50\%$ ,  $SD=31.21\%$ ; Japanese:  $M=25.83\%$ ,  $SD=22.36\%$ ), as presented in Figure 3.

To explore further whether cross-linguistic differences in manner encoding have an impact on participants' categorical preferences for manner, we built a logistic mixed-effects model<sup>3</sup> with participants' categorical choice as a binary dependent variable (*manner-preference*=1; *path-preference*=0), language groups as the fixed effect, and crossed-random intercepts for subject and item as the random effects. Results suggested that in line with speech production, bilinguals selected manner-match variant more frequently than Cantonese ( $\beta_{\text{Bilingual-Cantonese}}=0.84$ ,  $SE=0.33$ , Wald  $z=2.56$ ,  $p=.004$ ), yet patterned with English ( $\beta_{\text{Bilingual-English}}=-0.39$ ,  $SE=0.33$ , Wald  $z=-1.19$ ,  $p=.466$ ), suggesting an L2 influence. By comparison, multilinguals selected manner less frequently than bilinguals ( $\beta_{\text{Multilingual-bilingual}}=-1.02$ ,  $SE=0.33$ , Wald  $z=-3.07$ ,  $p=.01$ ), yet patterned with Cantonese ( $\beta_{\text{Multilingual-Cantonese}}=-0.18$ ,  $SE=0.33$ , Wald  $z=-0.54$ ,  $p=.590$ ), indicating a possible influence from the L3. Japanese monolinguals had the lowest frequency ( $\beta_{\text{Japanese-multilingual}}=-0.95$ ,  $SE=0.34$ , Wald  $z=-2.81$ ,  $p=.01$ ).

*Predictive factors of cognitive restructuring in both linguistic and non-linguistic tasks.* We further explored whether cognitive restructuring in the multilingual mind was modulated by language proficiency and audiovisual exposure. Following Athanasopoulos (2009), we transformed participants' exposure to each language into percentage scores by totalling their overall use of L1, L2, and L3. Based on their self-reports (cf. Table 1), multilinguals' exposure to Cantonese constituted 22.80% ( $SD=9.87$ ), English 32.07% ( $SD=13.47$ ), and Japanese 45.13% ( $SD=15.14$ ) of the time.

We set up two logistic mixed-effects models<sup>4,5</sup> with the frequency of manner encoding and manner-match preference as the respective binary dependent variable. The fixed effects were centred L2 and L3 proficiency and exposure. The random effects were random intercepts for subject and item. Likelihood ratio tests were performed using *anova()* function to evaluate model fits. Collinearity was not an issue for both models. The correlations ( $r$ ) between all fixed effects were

**Table 3.** Fixed effects on the mean frequency of manner in the linguistic encoding.

Fixed effects	Estimate	SE	Wald z	Pr (> z )
L2 proficiency	-0.334	0.547	-0.61	.542
L3 proficiency	0.609	0.472	1.291	.197
L2 exposure	3.995	1.841	2.171	.030*
L3 exposure	-3.412	1.613	-2.116	.034*

Note. SE: standard error.

\* $p < .05$ .

**Table 4.** Fixed effects on the mean frequency of manner-match selection in categorization.

Fixed effects	Estimate	SE	Wald z	Pr (> z )
L2 proficiency	-0.719	0.622	-1.156	.248
L3 proficiency	0.814	0.588	1.384	.167
L2 exposure	5.077	1.964	2.585	.009**
L3 exposure	-6.399	1.950	-3.281	.001**

Note. SE: standard error.

\*\* $p < .01$ .

**Table 5.** Fixed effects on manner encoding and selection in linguistic and non-linguistic tasks.

Fixed effects	Estimate	SE	Wald z	Pr (> z )
Linguistic task				
L2 TV for manner encoding	2.006	0.925	2.168	.030*
L3 TV for manner encoding	-1.988	0.995	-1.999	.040*
Non-linguistic task				
L2 TV for manner choice	4.691	1.715	2.735	.006**
L3 TV for manner choice	-6.833	2.194	-3.114	.002**

Note. SE: standard error.

\* $p < .05$ ; \*\* $p < .01$ .

below .5, and the variance inflation factor VIF were below 1.92. Statistical details are presented in Tables 3 and 4.

The overall results showed that only audiovisual exposure served as a significant contributor in both tasks. That is, participants with greater L2 exposure were more likely to mention manner in language production and use manner-match as a criterion for categorization. In contrast, participants with greater L3 exposure were less likely to mention manner, and instead used path-match as a criterion for categorization.

To better understand the role of different types of exposure in cognitive restructuring, two separate logistic mixed-effects models were set up. The respective dependent variables were the frequency of manner encoding and manner-match selection. The fixed effects were different types of exposure (i.e., radio, TV, Internet, and books/magazines/newspapers). All the fixed effects were centred. Collinearity was not an issue for both models. The correlations ( $r$ ) between all fixed effects were below .5, and the VIF were below 1.98. Results showed that among different types of exposure, only the daily amount of TV watching served as the significant predictor, as shown in Table 5.

## General discussion

The current study examines the extent to which learning additional languages gives rise to cognitive restructuring in how people talk and think about motion. The first research question examined how multilinguals encoded voluntary motion in the L1 compared with their bilingual and monolingual counterparts. The monolingual data confirmed the typological status of each language, that is, participants across each group expressed path with a high-level frequency, indicating that path is a core element (Slobin, 2004; Talmy, 2000). However, the mean frequency of manner encoding certainly contrasted across languages (cf. Table 2). In English, as manner is predominantly encoded in a syntactically dense structure of ‘manner verbs + path satellites’, it is difficult to drop this information from speech. By comparison, Japanese most often conflates path in the main verb, and there is no obligatory syntactic slot for manner (i.e., ‘path-only’ construction). As a result, manner in Japanese can be easily added or dropped from speech depending on whether speakers would like to stress this information or not. Being equipollent-framed, Cantonese exhibits characteristics of both satellite- and verb-framed languages (Wang & Li, 2019; Yiu, 2013, 2014).

Bilinguals’ L1 descriptions displayed L2-based patterns. As Cantonese and English share a satellite-framing verbal strategy, this partial overlap tends to motivate learners to converge to a single lexicalization pattern that works well for both languages (Authors, 2019; Aveledo & Athanasopoulos, 2016; Daller et al., 2011). In a similar vein, multilinguals’ L1 descriptions patterned with Cantonese monolinguals, yet differed from bilinguals in manner selection and event constructions, indicating a reverse impact from the L3. As previously mentioned, Cantonese incorporates characteristics of both S- and V-languages (Yiu, 2013, 2014). The typological similarities between L1 and L2, and L1 and L3 facilitate the simultaneous influence from both languages on speakers’ L1. As L1-Cantonese was the only test language that bi- and multilinguals used for verbalization, we interpret the observed cross-language effects as long-term consequences of additional language learning. That is, to reduce the cognitive load of keeping two or three separate language systems, speakers tend to develop a converged set of concepts as a result of mutual interaction across different languages (Daller et al., 2011; Ji et al., 2011; Ji & Hohenstein, 2014). In the current case, multilinguals are highly proficient L3 users and have a large amount of exposure to Japanese (cf. Table 2). The sufficient exposure and active use of L3 may override L2-based patterns and accelerate the L3-based restructuring process (Wang & Li, 2019; Aveledo & Athanasopoulos, 2016; Bylund & Athanasopoulos, 2014).

The second research question examined how multilinguals categorized voluntary motion in comparison to the bilingual and monolingual controls. Results suggested that consistent with patterns of lexicalization, monolinguals demonstrated a hierarchical decrease in categorization preferences (cf. Figure 3). The consistency in both tasks can be interpreted as a thinking-for-speaking and thinking-with-language effect. That is, the language-specific regularities made available in the linguistic encoding (i.e., preferred verbal strategies) tend to mediate or affect participants’ non-linguistic performance during or immediately after language use (Bylund & Athanasopoulos, 2014; Filipović, 2018; Montero-Melis & Bylund, 2017; Wolff & Holmes, 2011). Given the various degrees of manner salience, speakers of high-manner-salience languages (i.e., English) tend to attach more importance to manner in mental representations due to its high codability, while speakers of low-manner-salience languages (i.e., Japanese) attach less attention when retrieving and utilizing relevant information for categorization. The findings thus indicate that the influence of language on thought is often transient, especially when the task promotes the strategic use of language in cognitive processing (Filipović, 2018; Gennari et al., 2002; Kersten et al., 2010; Montero-Melis & Bylund, 2017; Wolff & Holmes, 2011).

Moving beyond monolinguals, bi- and multilinguals exhibited L2- and L3-based patterns in manner-match selection, indicating that learning additional languages goes beyond the successful

acquisition of target linguistic forms, to entail a new way to think about motion (Wang & Li, 2021a,b; Bylund & Athanasopoulos, 2014). Given the multimodal nature of concepts, bi- and multilinguals can reconstruct their L1-based conceptualization patterns when given sufficient exposure to instances of language-specific patterns (Bylund & Athanasopoulos, 2014; Casasanto, 2008). It is worth mentioning that in the current study we argue for a temporary effect of language on thought within the framework of thinking-for-speaking. In future studies, we intend to adopt a dual task paradigm to further explore whether the impact of L3 on conceptualization mechanisms could go beyond the thinking-for-speaking phenomenon, and lead to changes in the speaker's patterns of thought regardless of which language is being used, if any at all.

The third research question examined the role of proficiency and audiovisual exposure in cognitive restructuring. Results showed that only television exposure served as a significant contributor to the restructuring of both linguistic and conceptualization patterns. Given that concepts are made up by images, scripts, and audiovisual impressions (Athanasopoulos, Damjanovic, et al., 2015), language learning can benefit from multimodal exposure and strategies (Jarvis & Pavlenko, 2008). Television has a multimodal nature, so learners can access a variety of authentic language materials presented across different modalities. These multimodal cues could engage learners to pay special attention to language-specific categories from different dimensions (Cintrón-Valentín et al., 2019; Montero Perez et al., 2018; Vanderplank, 2010). Regarding motion events, television not only exposes learners to different ways of talking about motion, but also visuals and dynamic depictions of motion. Continuous exposure to dynamic motion scenes could help learners figure out the regularities of co-occurrence in categorization and lead to a routinisation of the corresponding conceptual categories (Athanasopoulos, Bylund, et al., 2015; Bylund & Athanasopoulos, 2014). In the current case, participants with greater L3 television exposure are more prone to use L3-based patterns in lexicalization and pay less attention to manner in categorization. The multimodality of television may account for the reason why other types of media exposure, such as the Internet, books, and radio, failed to play a role in cognitive restructuring due to their mono-modal nature. Similar results are reported by Bylund and Athanasopoulos (2015), from a grammatical perspective, that audiovisual exposure could benefit learners in acquiring new temporal viewing frames. The current study extends the positive effects of audiovisual exposure to the motion domain from a lexical perspective of manner and path.

However, the current study failed to document an effect of language proficiency. One plausible explanation is that the effect of proficiency is not linear, and there might be a threshold for it to play a role (Athanasopoulos, Damjanovic, et al., 2015). For example, there might be intervals or ranges where the proficiency effect is most prominent (Bylund & Athanasopoulos, 2014). But once learner's proficiency exceeds a certain point, such effect may level out consequently (Bylund & Athanasopoulos, 2015). In the current case, given that bi- and multilinguals are functional language users, it is possible that their proficiency has already passed a threshold and its effects on cognitive restructuring is no longer visible.

## **Conclusion**

The current study sets out to examine the impact of language learning on Cantonese–English–Japanese multilinguals' cognitive restructuring of voluntary motion. It is found that multilinguals' L1 can be restructured under the influence of both L2 and L3, and the degree of restructuring is modulated by audiovisual media exposure to the target language. The inclusion of L3 in language-and-thought research poses itself how different languages within a single mind affect thought together: whether speakers have a single integrated way of thinking, or they can switch between different thinking patterns depending on which language they are using. The dialogue between multiple language learning and cognitive restructuring allows us to 'theorize the interaction



between multiple languages in the speaker's mind as a natural and ongoing process and to understand why multilinguals may perform differently from monolinguals in all of their languages, including the L1' (Jarvis & Pavlenko, 2008, p. 17).

In sum, the current findings lend support to the thinking-for-speaking hypothesis that the language effect on cognition is flexible and task-dependent and add to the emerging view that audiovisual exposure plays a central role not only in L2 vocabulary development, but also in one's conceptual knowledge. Given the indispensable role of audiovisual exposure in language learning, teachers and learners should make full use of audiovisual tools to facilitate the internalization of new concepts. As Tyler (2012) has pointed out 'visuals meant to provide memorable, meaningful representation for L2 learners' (p. 136), language teachers may choose to use pictorial materials and dynamic videos that combine linguistic resources with visuals or sounds when teaching novel concepts and new ways of perceiving the world.

### Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The article was prepared in the framework of a research grant funded by the Ministry of Science and Higher Education of the Russian Federation (Grant ID: 075-15-2020-928) and the ESRC Postdoctoral Fellowship (ES/V012274/1).

### ORCID iDs

Yi Wang  <https://orcid.org/0000-0003-4175-7467>

Li Wei  <https://orcid.org/0000-0002-2015-7262>

### Supplemental material

Supplemental material for this article is available online.

### Notes

1. Model 1  $\leq$  glmer (Manner encoding  $\sim$  Language group + (1|Subject) + (1|Item), family=binomial, data=data).
2. Model 2  $\leq$  glmer(Path encoding  $\sim$  Language group + (1|Subject) + (1|Item), family=binomial, data=data).
3. Model 3  $\leq$  glmer(Manner-match choice  $\sim$  Language group + (1|Subject) + (1|Item), family=binomial, data=data).
4. Model 4  $\leq$  glmer(Manner encoding  $\sim$  L2 proficiency + L3 proficiency + L2 exposure + L3 exposure + (1|Subject) + (1|Item), family=binomial, data=data).
5. Model 5  $\leq$  glmer(Manner-match choice  $\sim$  L2 proficiency + L3 proficiency + L2 exposure + L3 exposure + (1|Subject) + (1|Item), family=binomial, data=data).

### References

- Athanasopoulos, P. (2009). Cognitive representation of colour in bilinguals: The case of Greek blues. *Bilingualism: Language and Cognition*, 12(1), 83–95.
- Athanasopoulos, P., Bylund, E., Montero-Melis, G., Damjanovic, L., Schartner, A., Kibbe, A., & Thierry, G. (2015). Two languages, two minds: Flexible cognitive processing driven by language of operation. *Psychological Science*, 26(4), 518–526.

- Athanasopoulos, P., Damjanovic, L., Burnand, J., & Bylund, E. (2015). Learning to think in a second language: Effects of proficiency and length of exposure in English learners of German. *Modern Language Journal, 99*, 138–153.
- Aveledo, F., & Athanasopoulos, P. (2016). Second language influence on first language motion event encoding and categorization in Spanish-speaking children learning L2 English. *International Journal of Bilingualism, 20*(4), 403–420.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using Ime4. *Journal of Statistical Software, 67*, 1–48.
- Brown, A., & Chen, J. (2013). Construal of Manner in speech and gesture in Mandarin, English, and Japanese. *Cognitive Linguistics, 24*(4), 605–631.
- Brown, A., & Gullberg, M. (2008). Bidirectional cross-linguistic influence in L1-L2 encoding of manner in speech and gesture: A study of Japanese speakers of English. *Studies in Second Language Acquisition, 30*(2), 225–251.
- Brown, A., & Gullberg, M. (2011). Bidirectional cross-linguistic influence in event conceptualization? Expressions of path among Japanese learners of English. *Bilingualism: Language and Cognition, 14*(1), 79–94.
- Bylund, E. (2009). Effects of age of L2 acquisition on L1 event conceptualization patterns. *Bilingualism: Language and Cognition, 12*(3), 305–322.
- Bylund, E., & Athanasopoulos, P. (2014). Linguistic relativity in SLA: Toward a new research program. *Language Learning, 64*(4), 952–985.
- Bylund, E., & Athanasopoulos, P. (2015). Televised Whorf: Cognitive restructuring in advanced foreign language learners as a function of audiovisual media exposure. *Modern Language Journal, 99*, 123–137.
- Bylund, E., & Jarvis, S. (2011). L2 effects on L1 event conceptualization. *Bilingualism: Language and Cognition, 14*(1), 47–59.
- Casasanto, D. (2008). Who's afraid of the big bad Whorf? Crosslinguistic differences in temporal language and thought. *Language Learning, 58*, 63–79.
- Cintrón-Valentín, M., García-Amaya, L., & Ellis, N. C. (2019). Captioning and grammar learning in the L2 Spanish classroom. *The Language Learning Journal, 47*, 439–459.
- Cook, V., & Bassetti, B. (2011). *Language and bilingual cognition*. Psychology Press.
- Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Press Syndicate of the University of Cambridge.
- Daller, M. H., Treffers-Daller, J., & Furman, R. (2011). Transfer of conceptualization patterns in bilinguals: The construal of motion events in Turkish and German. *Bilingualism: Language and Cognition, 14*(1), 95–119.
- Filipović, L. (2018). Speaking in a second language but thinking in the first language: Language-specific effects on memory for causation events in English and Spanish. *International Journal of Bilingualism, 22*(2), 180–198.
- Flecken, M., Carroll, M., Weimar, K., & Von Stutterheim, C. (2015). Driving along the road or heading for the village? Conceptual differences underlying motion event encoding in French, German, and French-German L2 users. *Modern Language Journal, 99*(Suppl. 1), 100–122.
- Gennari, S. P., Sloman, S. A., Malt, B. C., & Fitch, W. T. (2002). Motion events in language and cognition. *Cognition, 83*(1), 49–79.
- Hendriks, H., & Hickmann, M. (2015). Finding one's path into another language: On the expression of boundary crossing by English learners of French. *Modern Language Journal, 99*(Suppl. 1), 14–31.
- Hickmann, M., & Hendriks, H. (2010). Typological constraints on the acquisition of spatial language in French and English. *Cognitive Linguistics, 21*(2), 189–215.
- Hohenstein, J., Eisenberg, A., & Naigles, L. (2006). Is he floating across or crossing afloat? Cross-influence of L1 and L2 in Spanish-English bilingual adults. *Bilingualism: Language and Cognition, 9*(3), 249–261.
- Jarvis, S., & Pavlenko, A. (2008). *Crosslinguistic influence in language and cognition*. Routledge.
- Ji, Y., Hendriks, H., & Hickmann, M. (2011). The expression of caused motion events in Chinese and in English: Some typological issues. *Linguistics, 49*(5), 1041–1077.
- Ji, Y., & Hohenstein, J. (2014). The syntactic packaging of caused motion components in a second language: English learners of Chinese. *Lingua, 140*, 100–116.

- Kersten, A. W., Meissner, C. A., Lechuga, J., Schwartz, B. L., Albrechtsen, J. S., & Iglesias, A. (2010). English speakers attend more strongly than Spanish speakers to manner of motion when classifying novel objects and events. *Journal of Experimental Psychology: General*, *139*(4), 638–653.
- Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, *48*(1), 16–32.
- Lai, V. T., Rodriguez, G. G., & Narasimhan, B. (2014). Thinking-for-speaking in early and late bilinguals. *Bilingualism: Language and Cognition*, *17*(1), 139–152.
- Larrañaga, P., Treffers-Daller, J., Tidball, F., & Ortega, M.-C. G. (2012). L1 transfer in the acquisition of manner and path in Spanish by native speakers of English. *International Journal of Bilingualism*, *16*(1), 117–138.
- Li, P., Zhang, F., Tsai, E., & Puls, B. (2014). Language history questionnaire (LHQ 2.0): A new dynamic web-based research tool. *Bilingualism: Language and Cognition*, *17*(3), 673–680.
- Matthews, S., & Yip, V. (2011). *Cantonese: A comprehensive grammar*. Routledge.
- Mayer, R. E. (2001). *Multimedia learning*. Cambridge University Press.
- McNeill, D. (2005). *Gesture and thought*. The University of Chicago Press.
- Montero-Melis, G., & Bylund, E. (2017). Getting the ball rolling: The cross-linguistic conceptualization of caused motion. *Language and Cognition*, *9*(3), 446–472.
- Montero-Melis, G., Jaeger, T. F., & Bylund, E. (2016). Thinking is modulated by recent linguistic experience: Second language priming affects perceived event similarity. *Language Learning*, *66*(3), 636–665.
- Montero Perez, M., Peters, E., & Desmet, P. (2018). Vocabulary learning through viewing video: The effect of two enhancement techniques. *Computer Assisted Language Learning*, *31*, 1–26.
- Paivio, A. (2014). *Mind and its evolution: A dual coding theoretical approach*. Psychology Press.
- Papafragou, A., Hulbert, J., & Trueswell, J. (2008). Does language guide event perception? Evidence from eye movements. *Cognition*, *108*(1), 155–184.
- Park, H. I. (2019). How do Korean–English bilinguals speak and think about motion events? Evidence from verbal and non-verbal tasks. *Bilingualism: Language and Cognition*, *23*(3), 483–499.
- Pavlenko, A. (2011). Thinking and speaking in two languages: Overview of the field. In A. Pavlenko (Ed.), *Thinking and speaking in two languages* (pp. 237–257). Multilingual Matters.
- Peters, E., Heynen, E., & Puimège, E. (2016). Learning vocabulary through audiovisual input: The differential effect of L1 subtitles and captions. *System*, *63*, 134–148.
- R Core Team (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rodgers, M. P. H., & Webb, S. (2017). The effects of captions on EFL learners' comprehension of English–Language television programs. *CALICO Journal*, *34*, 20–38.
- Slobin, D. I. (1987). Thinking for speaking. *Proceedings of the Annual Meeting of the Berkeley Linguistic Society*, *13*, 435–445.
- Slobin, D. I. (1996). From 'thought and language' to 'thinking for speaking'. In J. Gumperz & S. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70–96). Cambridge University Press.
- Slobin, D. I. (2004). The many ways to search for a frog: Linguistic typology and the expression of motion events. In S. Strömquist & L. Verhoeven (Eds.), *Relating events in narrative: Typological and contextual perspectives* (pp. 219–257). Erlbaum.
- Slobin, D. I., & Hoiting, N. (1994). Reference to movement in spoken and signed languages: Typological considerations. *Proceedings of the Twentieth Annual Meeting of the Berkeley Linguistics Society*, *20*(1), 487–505.
- Talmy, L. (2000). *Toward a cognitive semantics*. MIT Press.
- Trueswell, J. C., & Papafragou, A. (2010). Perceiving and remembering event cross-linguistically: Evidence from dual-task paradigms. *Journal of Memory and Language*, *63*(1), 64–82.
- Tyler, A. (2012). *Cognitive linguistics and second language learning: Theoretical basics and experimental evidence*. Routledge.
- Vanderplank, R. (2010). Déjà vu? A decade of research on language laboratories, television and video in language learning. *Language Teaching*, *43*(1), 1–37.

- Wang, Y., & Li, W. (2019). Cognitive restructuring in the bilingual mind: Motion event construal in early Cantonese-English bilinguals. *Language and Cognition, 11*(4), 527–554.
- Wang, Y., & Li, W. (2021a). Cognitive restructuring in the multilingual mind: Language-specific effects on processing efficiency of caused motion events in Cantonese–English–Japanese speakers. *Bilingualism: Language and Cognition, 24*(4), 730–745.
- Wang, Y., & Li, W. (2021b). Two languages, one mind: The effects of language learning on motion event processing in early Cantonese-English bilinguals. *Proceedings of the Annual Meeting of the Cognitive Science Society, 43*(43), 2169–2175.
- Wolff, P., & Holmes, K. J. (2011). Linguistic relativity. *Wiley Interdisciplinary Reviews: Cognitive Science, 2*(3), 253–265.
- Yiu, C. (2013). Directional verbs in Cantonese: A typological and historical study. *Language and Linguistics, 14*(3), 511–569.
- Yiu, C. (2014). Directional verbs in modern Cantonese: A typological perspective. In C. Yiu (Ed.), *The typology of motion events: An empirical study of Chinese dialects* (pp. 33–135). De Gruyter Mouton.

### Author biographies

Dr. Yi Wang is a Lecturer of English Language and Linguistics at Cardiff University. She received her Ph.D. from University College London. Her research expertise lies in language and cognition, motion event processing, and experimental psychology.

Professor Li Wei is Director and Dean of the UCL Institute of Education at University College London. He works on different aspects of bilingualism and multilingualism. He is currently the Editor of the International Journal of Bilingual Education and Bilingualism and Applied Linguistics Review.