*Studies in Second Language Acquisition* (2025), **47**, 53–79
doi:10.1017/S0272263124000433

RESEARCH ARTICLE

# Modulating motion event categorization through brief training: *Meaning-focused versus form-focused instructional conditions*

Yuyan Xue and John N. Williams

University of Cambridge, UK
**Corresponding author:** Yuyan Xue; Email: yx324@cam.ac.uk

(Received 13 June 2023; Revised 21 April 2024; Accepted 13 May 2024)

## Abstract

There is evidence that learning a second language (L2) can shift cognition toward that predicted for the L2 and that this effect might vary with L2 proficiency, age of acquisition, length of immersion, etc. Here we explore the previously neglected variable of language instructional conditions. Participants categorized motion events in a triads-matching task after being trained on two novel linguistic labels highlighting (in)transitivity through one of three instructional conditions. Participants who learned the relevant knowledge under a meaning-focused instructional condition (memorizing meanings of exemplar sentences) showed a higher likelihood of categorizing based on motion (in)transitivity immediately after training than a control group; those who learned under a required rule search instructional condition showed this effect only after additional practice; while those who learned through another type of form-focused instructional condition (direct metalinguistic explanation) did not show this effect even after such practice. These differences were obtained despite the fact that the three groups were matched on awareness of the target system at the level of understanding and near-perfect performance on a grammaticality judgment task. The findings are discussed in terms of the depth of processing in instructed SLA and models of language–cognition interactions.

**Keywords:** Instructional condition; rule search; meaning-focused; metalinguistic explanation

## Introduction

Cognitive and psycholinguistic research has increasingly focused on whether language learning can modify our cognition beyond language itself, that is, during tasks without overt language use (Montero-Melis et al., 2016). A growing body of evidence suggests that learning an additional language that encodes a familiar concept differently from our first language (L1) and/or further highlights a familiar concept compared to our L1 can indeed shift our cognition beyond language per se in a variety of domains including time

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0272263124000433 Published online by Cambridge University Press

(e.g., Athanasopoulos & Bylund, 2023), motion (e.g., Athanasopoulos, Burnand et al., 2015), and number/amount of objects/substance (e.g., Athanasopoulos, 2006), etc[1]. Moreover, such cognitive shifts appear to be modulated by a list of factors including current language contexts (Athanasopoulos, Bylund et al., 2015), second language (L2) proficiency, age of L2 acquisition, length of immersion in an L2 setting, amount of daily exposure to the foreign language (FL), etc. (see Bylund & Athanasopoulos, 2014a).

Here we explore the role played by instructional conditions, a factor attracting decades of research in the second language acquisition (SLA) literature, but which has not been systematically investigated in the light of language–cognition interactions. The three instructional conditions we compared were: one meaning-focused (memorizing meanings of exemplar sentences) and two form-focused (required rule search and direct metalinguistic explanation).

To systematically control for and manipulate instructional conditions, we adopted a lab training paradigm, where participants were trained on an artificial linguistic system prior to performing a task measuring cognitive shifts that might have occurred as a result of this training. It has been argued that training studies can control for various confounds suffered by cross-linguistic comparisons while at the same time being sufficiently ecologically valid to be comparable to and inform, real-world language–cognition interactions in learners of a natural L2/FL (Casasanto, 2008; Montero-Melis et al., 2016, see below).

We trained Mandarin L1 – English L2 participants on an artificial linguistic system involving two novel grammatical morphemes highlighting the concept of (in)transitivity (*gi* was obligatorily used in intransitive sentences, *ro* in transitive sentences), a concept that is familiar but not obligatorily marked in either Mandarin or English (see Everett, 2013). Besides a grammaticality judgment test (GJT) and a debriefing to assess whether they had achieved awareness at the level of understanding (Schmidt, 1990) of this target system, our critical dependent variable was the likelihood of categorizing motion events on the basis of motion (in)transitivity in a triads-matching task without overt language use. Our research question was whether participants who had learned this target system (defined by achieving near-perfect GJT accuracy and awareness at the level of understanding of the target system) would attend more to motion (in)transitivity (compared with other dimensions such as motion direction) in such a triads-matching task compared with a control group naïve to this target system and if yes whether this cognitive shift effect would vary with instructional conditions.

By probing this research question, we also aimed to extend the focus of comparisons between form- and meaning-focused instructional conditions from language attainment (e.g., accuracy, fluency, nativelikeness of event-related potential (ERP) responses, etc.) to cognitive shifts.

## Effects of (in)transitivity markers on motion event categorization

According to many cognitive linguistic and psycholinguistic theories, human cognition is essentially categorization based on similarity assessment (Hahn, 2014; Ameel et al., 2005). A classical task used to explore this is the triads-matching task, whereby participants are asked to indicate which one of two alternative stimuli is more similar

[1]Another important strand of research focuses on cognitive restructuring induced by learning an additional language that establishes novel categories (e.g., color, shape, odor, tactile, and gender categories) absent in one's L1. But as this strand is not directly pertinent to our study, it is not covered in the Introduction.

https://doi.org/10.1017/S0272263124000433 Published online by Cambridge University Press

to a target stimulus (Athanasopoulos, Bylund et al., 2015). A substantial body of research on language–cognition interactions has employed such triads-matching tasks, where neither the stimuli nor the task involved overt language use.

For example, Everett (2013), which our study is modeled on, compared motion event triads-matching preferences between native speakers of English and Karitiâna, an Amazonian language that obligatorily marks (in)transitivity. In this language, *i-* is obligatorily prefixed to intransitive verbs, while *naka-* to transitive verbs. In a triads-matching task without overt language use, Everett found that Karitiâna native speakers were more likely to match motion events based on (in)transitivity, while English native speakers, who do not obligatorily mark (in)transitivity in their language, were biased toward other dimensions such as the shape of the entities involved in a motion event. Note that we replaced *i-* with *gi-* and *naka-* with *ro-* in this study to match the length and consonant–vowel structure of the two labels.

## Motion event categorization in L2/FL learners

Beyond comparisons of native speakers, several studies have also investigated L2/FL learners' triads-matching preferences of motion events. For example, Bylund and Athanasopoulos (2015) asked English monolinguals and L1 Swedish learners of FL English to match in silence a motion event of intermediate endpoint and ongoingness orientation (e.g., a person walking toward an outdoor market) to either a motion event of high endpoint and low ongoingness orientation ([+ endpoint], e.g., a person walking and entering a shop) or vice versa ([– endpoint], e.g., a person walking in a parking lot). The results showed that on the group average level, the FL English learners chose the (+ endpoint) video clip significantly more often and thus exhibited a stronger endpoint preference than English monolinguals. Crucially, the FL learners' endpoint preference was negatively correlated with their daily exposure to FL audiovisual media (television [TV] watching), such that the more they watched TV in English the more they showed an ongoingness preference. The authors attributed such cognitive difference during a task without overt language use to a grammatical difference between the two languages: English is an aspect language obligatorily marking progressive aspect (the morpheme *-ing*) on the main verb to express ongoingness, while Swedish is a non-aspect language where ongoingness is optionally expressed via lexical means (e.g., adverbials) outside the main verb. Consequently, when describing endpoint-oriented motion events, aspect language speakers preferred to exclude the endpoint while obligatorily marking ongoingness; whereas nonaspect language speakers tended to encode the endpoint while omitting the expression of ongoingness. Both ongoingness and endpoint are familiar concepts that can be expressed, through whichever means, in both the FL learners' L1 and FL, but learning an FL that further highlights ongoingness appeared to have shifted their cognition during triads-matching without overt language use.

An intriguing finding of the above study is that only daily exposure to audiovisual media (TV) but no other primarily monomodal types of media (e.g., radio and reading) in English predicted the degree of cognitive shift in the FL learners. This leads to the possibility that it might be how motion events are visually depicted on TV (e.g., the visual scene itself emphasizing motion ongoingness over endpoint), instead of language per se, that induced a cognitive shift. Relatedly, other studies reporting similar cognitive shifts on motion event categorization all involved L2 learners who had lived in an L2 setting or had L2-mediated instruction at school and had L2 exposure in various contexts daily (e.g., Athanasopoulos, Burnand et al., 2015; Athanasopoulos, Bylund

https://doi.org/10.1017/S0272263124000433 Published online by Cambridge University Press

et al., 2015; Bylund & Athanasopoulos, 2014b; Bylund, Athanasopoulos & Oostendorp, 2013). This further leads to the possibility that the acculturation process, instead of language itself, might have given rise to the cognitive shift (Montero-Melis et al., 2016).

## Perceptual training studies on language–cognition interactions

The possibility of acculturation confounds has motivated a strand of research (Montero-Melis et al., 2016; Casasanto, 2008; Dolscheid, Shayan, Majid & Casasanto, 2013) on perceptual training, whereby participants from a single background are trained on the target linguistic feature through condensed exposure in the lab, followed by a cognitive task without overt language use that measures cognitive shift. By comparing with a control group naïve to the target linguistic feature, this truly experimental design not only controls for the acculturation process but also the amount of input each participant receives regarding the target linguistic feature, which is very difficult, if not impossible, to control in a quasiexperimental, cross-linguistic study (Montero-Melis et al., 2016; Casasanto, 2008; Dolscheid et al., 2013).

Regarding ecological validity, though one can always question whether there is a giant leap from cross-linguistic studies to lab-based training studies, authors of the latter strand usually situate their research within the former strand, arguing that the miniature lab training paradigm can provide valuable insight into bilingual cognition in general because linguistic exposure in lab-based training provides "a very condensed version of what might be going on in the long process of learning a new language that carves up reality in a different way than our L1" (Montero-Melis et al., 2016: 657) and that our experience with natural language may shift cognition "in much the same way" as a lab training task (Casasanto, 2008: 75).

Montero-Melis et al. (2016) reported such a lab training study on another domain of motion event cognition. According to Talmy's (2000) typology, satellite-framed languages (e.g., English and Swedish) tend to encode manner information of motion events in the main verb (e.g., The boy *walks/jumps/runs* up the stairs) and path information in a verb satellite (The boy walks *up/down* the stairs); whereas path-framed languages (e.g., Spanish) prefer to encode path in the main verb (e.g., El chico subió/bajó las escaleras) and omit manner information, though manner can also potentially be expressed in a gerund (e.g., El chico subió las escaleras caminando/corriendo) or the main verb (e.g., El chico corrió hacia la casa). Therefore, it has been theorized that the manner of motion events is cognitively more salient for speakers of satellite-framed than path-framed languages. Montero-Melis et al. (2016) made use of the aforementioned flexibility in Spanish to prime L1 Swedish–L2 Spanish speakers to describe 32 videos of motion events in Spanish either in the path-primed way (path highlighted in the main verb) or the manner-primed way (manner highlighted in the main verb). After this, participants performed a similarity assessment task on motion events without overt language use. Results showed that manner-primed participants based their similarity judgments on manner more than path-primed participants. There was also a trend toward path-primed participants basing their judgments on path more than manner-primed participants. In other words, the brief priming manipulation highlighted the familiar concept of either manner or path.

## The label-feedback and structural-feedback hypotheses

The label-feedback hypothesis (Lupyan, 2012), following interactive activation principles, may offer a working hypothesis on the mechanism underpinning the

https://doi.org/10.1017/S0272263124000433 Published online by Cambridge University Press

aforementioned cognitive shift. To elaborate, long-term or repetitive linguistic exposure may give rise to a strong loop between a linguistic label (a word or a grammatical morpheme, etc.) and the perceptual features diagnostic of the category that the label refers to (see Montero-Melis et al., 2016: 638). During a cognitive task without overt language use, linguistic labels can nonetheless be drawn on covertly and online, passing top-down activation to the perceptual features with which they are associated, thereby shifting cognition (e.g., triads-matching preferences) via the loop.

Montero-Melis et al. (2016), for instance, suggested that though their similarity assessment task did not involve overt language use, the linguistic label (word) of manner and/or path can nevertheless be recruited covertly online to redirect participants' attention to different perceptual features of the motion events. For manner-primed participants, for instance, the label of manner (e.g., "empuja") arguably had a higher activation level than the label of the path (e.g., "sube") due to residual activation induced by priming, thus redirecting participants' attention to manner more than path during similarity assessment.

The label feedback account is further supported by findings that language effects on cognitive shifts appear to be disrupted by online verbal interference (e.g., repeating digits aloud during the cognitive task), which might inhibit online and covert recruitment of linguistic labels (e.g., Vanek, 2020; Athanasopoulos, Bylund et al., 2015). In fact, the disruption from verbal interference is one of the major rationales for the label-feedback hypothesis. Therefore, this hypothesis predicts cognitive shift effects of language when language is not overtly involved nor blocked in a task.

More recently, Sato and Athanasopoulos (2018) further argued that such feedback is not limited to whole labels as Lupyan proposed, but grammatical properties can also produce such feedback, which they refer to as "structural feedback." For example, grammatical gender properties of inanimate objects can be recruited online covertly, automatically, and inevitably to shift sex-related judgments. Seeing two daily objects with feminine grammatical gender in French primed French but not English native speakers to judge a subsequent sex-ambiguous face as being more feminine (Experiment 2, ibid.).

How might label and/or structural-feedback be applied to (in)transitivity? While the notion of transitivity may be represented as a core functional category within syntactic theories (Bowers, 2002), or may be analyzed within semantic theories in terms of multiple semantic facets such as agency and the affectedness of the object (Hopper & Thompson, 1980), here we just make the minimal working assumption that there is some form of abstract linguistic representation that indicates, for each verb, whether it can partake in the transitive and/or intransitive construction. Within the syntactic priming literature such representations have been posited in the form of "combinatorial nodes"—all intransitive verbs are linked to a common intransitive combinatorial node, and all transitive verbs to a common transitive node (van Gompel et al., 2012).[2] Within this framework, our experimental group participants will learn the connections between the novel labels (*gi* and *ro*) and the intransitive and transitive combinatorial nodes (much like in German the masculine article *der* would connect to a "masculine" gender combinatorial node which is linked to the lemmas for specific nouns, Salamoura & Williams, 2008). According to the label- and structural-feedback hypotheses, these

[2]In general, "combinatorial nodes" represent how words combine with other grammatical elements. They are attached to the representations of words at the lemma level, and are the assumed locus of relatively long-lived structural, as opposed to conceptual, priming effects (Pickering & Branigan, 1998).

https://doi.org/10.1017/S0272263124000433 Published online by Cambridge University Press

novel linguistic connections might lead to a cognitive shift in triads-matching preferences in the following way: for both experimental and control groups, a transitive motion event stimulus might covertly activate the combinatorial node for transitivity even without overt language use (analogous to Sato & Athanasopoulos, 2018 above). For experimental group participants who have learned the relevant connection between this combinatorial node and the novel label *ro*, a feedback loop will be established whereby activation passes between the two representations, increasing the activation of this combinatorial node, and thereby increasing activation of the perceptual correlates of motion transitivity, leading to a bias toward transitivity in triads-matching (compared with other dimensions such as shape). No such feedback loop can be activated for control participants who have not learned the relevant connections. In short, feedback from the novel label to the relevant combinatorial node and then to the perceptual correlates of motion (in)transitivity might give rise to a cognitive shift.

An alternative possibility is that the activity of learning the novel labels temporarily heightens the activation of the combinatorial nodes for (in)transitivity. This residual activation (i.e., priming of the combinatorial nodes) may produce a cognitive shift through feedback from the combinatorial node to the perceptual correlates of motion (in)transitivity, without online activation of the novel labels during triads-matching. In other words, it might be the feedback from the combinatorial nodes of (in)transitivity, instead of the label feedback from *gi/ro*, that gives rise to the cognitive shifts (this may be analogous to what Sato & Athanasopoulos (2018) termed "structural feedback"). Crucially, to explain any difference between experimental and control groups it must be assumed that it is the process of learning the mappings between the novel labels and the combinatorial nodes that leads to enhanced priming of the combinatorial nodes. Hence, the second possibility can be seen as an extension of the first, given that they both link cognitive shifts to learning of the target system.

## Factors affecting cognitive shifts

Previous studies on the broad topic of language–cognition interactions, be they cross-linguistic or miniature lab training, cognitive shifts of familiar conceptual categories, or establishment of novel ones via language, have investigated a range of factors including L2 proficiency (Li et al., 2018), age of L2 acquisition (Kersten et al., 2010), length of immersion in an L2 setting (Park, 2020), amount of daily exposure to the FL (Athanasopoulos, Burnand, et al., 2015), the language of operational context (Athanasopoulos, Bylund et al., 2015), (un)availability of online language involvement (Vanek, 2020), and (in)consistency of the form-meaning mapping (Vanek et al., 2021), but yielding inconsistent results.

A factor that has been more or less neglected is the language learning/acquisitional context and/or instructional conditions. These variables have been studied for decades in the SLA literature, but have not been systematically investigated in the context of language–cognition interactions. Our study therefore takes the first step toward filling this gap by throwing light on one of these variables, namely instructional conditions.

## Instructional conditions in SLA

Instructed SLA (ISLA) researchers often categorize instructional conditions as form-focused versus meaning-focused. The latter refers to situations where learners receive no metalinguistic explanation nor directions to attend to any specific linguistic form(s). Instead, the learner is exposed to the language in the context of a meaning-focused task. In

https://doi.org/10.1017/S0272263124000433 Published online by Cambridge University Press

a form-focused instructional condition, on the other hand, learners receive either direct metalinguistic explanation or direction to search for rules, i.e., there is a focus on form over meaning. Of course, labeling learning tasks as either form- or meaning-focused can only refer to the instructional condition as provided, and do not necessarily correspond to the mental processes actually engaged, e.g., learners under meaning-focused instructional conditions may spontaneously search for rules (e.g., Dekeyser, 1995).

Among lab-based training studies where task requirements and exposure are controlled (which are therefore most pertinent to the present study), many studies have compared instructional conditions involving metalinguistic explanation with inductive ones ranging from very form-focused (required rule search, e.g., the "rule-search" condition in Robinson, 1996; and the "[-formal instruction (FI), + rule search (RS)]" condition in Rosa & O'Neill, 1999) to very meaning-focused (e.g., the "incidental" condition in Robinson 1996; the "implicit-inductive" condition in DeKeyser 1995; the "random" condition in N. Ellis, 1993; the "implicit" condition in Lichtman, 2021). Some studies have employed an inductive instructional condition arguably between these two extremes (encouraged rule search, e.g., the "guided induction" condition in Leow et al., 2019, Cerezo et al., 2016, Zhuang, 2019, and Martin et al., 2019; the "learner-centered" condition in Hsieh et al., 2016; the "implicit" condition in Morgan-Short, Steinhauer et al., 2012; and de Graaff, 1997). This is because though participants were engaged in a meaning-focused task without the *overt* direction to search for rules, they were nonetheless provided with feedback regarding the grammatical system in question and/or guided questions facilitating hypothesis and rule formulation. These designs arguably prompted rule search more than a purely meaning-focused task without such feedback[3].

Regarding the dependent variable of the above studies, most of them have compared the efficacy of different instructional conditions as measured by accuracy at the condition–average level in various production and judgment tests, yielding inconsistent results. It is beyond the scope of the present study to elucidate the methodological details potentially giving rise to such inconsistencies. What is most pertinent to the current study, however, is the fact that we know very little from these studies whether the resultant linguistic knowledge will still differ in certain tacit aspects under different instructional conditions among participants who have all achieved awareness at the level of understanding (Schmidt, 1990) of the target knowledge and can thus perform various production and judgment tests at near-perfect behavioral accuracy.

There is indeed some suggestive evidence of such tacit differences. In Morgan-Short, Steinhauer et al. (2012), one group of participants learned an artificial language under an inductive instructional condition arguably between the very meaning-focused and very form-focused extremes (see above), while another group were first given metalinguistic grammatical explanations, followed by the same amount of game practice using the language. Though both groups showed equivalent near-perfect accuracy in a GJT on the group-average level, only participants from the former condition showed native-like ERP responses to syntactic (word category) violations, which were suggestive of rule-governed syntactic processing. Moreover, this edge was retained over time (Morgan-Short, Finger et al., 2012). Although Morgan-Short and colleagues did not

[3]We avoided referring to the instructional conditions investigated in the present study as "implicit" or "explicit" because we are aware of inconsistencies in the semantics of these terms in the literature.

https://doi.org/10.1017/S0272263124000433 Published online by Cambridge University Press

report debriefing data, the fact that all participants could achieve accuracy above 80% in the GJT strongly suggested that awareness at the level of understanding was reached.

What might be the tacit difference driving the differential ERP patterns in Morgan-Short and colleagues' studies? One hypothesis comes from Leow's (2015) model on the depth of processing (DoP) in instructed SLA (ISLA) which suggests that linguistic knowledge can still differ in degrees of internalization depending on the DoP of the training materials even among participants with awareness at the level of understanding and near-perfect GJT accuracy. DoP is defined as the relative amount of cognitive or mental effort, attention, or time cognitively spent on processing information (Hsieh et al., 2016) and/or the level of analysis and elaboration of the information (Leow, 2015). DoP may play a crucial role in all stages of ISLA whereby higher DoP of the training materials tends to contribute to more internalized knowledge than lower DoP. In the present context, we interpret "internalization" as referring to the establishment of the aforementioned connections between the novel labels and combinatorial nodes of (in)transitivity; the degree of internalization thus refers to the strength of such connections (this may be what Leow (2015) called "robustness" of the knowledge). These are consistent with Leow's (2015) implications that internalization refers to the establishment of certain *systems* (as opposed to rote items) and robustness as the retainability of certain internalized knowledge in memory (see for example p. 111 and 125, *ibid.*). In other words, although awareness at the level of understanding and near-perfect GJT accuracy are achievable under various instructional conditions, the underlying linguistic knowledge may still vary in degree of internalization, arguably induced by different DoP of the training material.

Though Morgan-Short et al.'s studies did not measure DoP online, Leow et al. (2019) suggested that rule search through practice may prompt higher DoP of the training materials than metalinguistic explanation plus practice. Specifically, their thinking aloud data suggested that while the predominant strategy of the latter condition was repeating and executing the propositional knowledge of the rules provided without much deep processing, participants under the former condition processed the training materials at a high DoP as instantiated by rule and hypothesis formulation, activation of (recent) prior knowledge (e.g., training sentences encountered shortly before), and metacognitive processes. This suggests that in Morgan-Short et al.'s studies, the inductive instructional condition (which arguably prompted rule search, see above) might also yield higher DoP than the other instructional condition, which may have given rise to different degrees of internalization of linguistic knowledge, as shown by the two instructional conditions' difference in rule-based (i.e., system-based, as opposed to item-based) linguistic processing. However, it should be noted that no such difference was observed between the same two instructional conditions when gender agreement errors were targeted (Morgan-Short et al., 2010). Hence, there is some evidence suggesting that among participants with awareness at the level of understanding and near-perfect GJT accuracy, rule search (as the actual learning process engaged) may be superior or at least equivalent to metalinguistic explanation in terms of developing more internalized linguistic knowledge.

However, participants can engage in rule search as the actual learning process under both a very form-focused instructional condition, where rule search is required, and a very meaning-focused one, where any rule search is spontaneous and incidental (e.g., DeKeyser, 1995). We do not know how these two inductive instructional conditions compare in terms of the degree of internalization of the resultant linguistic knowledge when only participants who attain equivalent awareness at the level of understanding and near-perfect GJT accuracy are compared.

https://doi.org/10.1017/S0272263124000433 Published online by Cambridge University Press

There is reason to predict that a meaning-focused instructional condition, when tested in this way, will have an edge over a form-focused instructional condition. Calderón (2013) suggested that the DoP of upcoming training materials logically decreased once awareness at the level of understanding had been achieved. We reason that this might be more likely under a required rule search instructional condition, as participants might logically decrease their DoP once they believe that they have satisfied the task requirement. Under a meaning-focused instructional condition, however, participants might continue to engage in high DoP to complete the meaning-focused task (which in the present case was to remember the meanings of all training sentences).

Athanasopoulos, Burnand et al. (2015) suggested that cognitive shifts arise from internalized linguistic knowledge. Besides, for label- and structural-feedback to occur, logically the relevant connections between the novel labels (gi/ro) and the combinatorial nodes of (in)transitivity would need to be sufficiently strong (i.e., the linguistic knowledge would need to be sufficiently internalized). If so, then cognitive shift may serve as another dependent variable to compare different instructional conditions in terms of the degrees of internalization of the knowledge gained.

## The current study

Our research question was whether cognitive shifts in the domain of motion (in)transitivity could be induced after brief training on a target system modeled on Karitiâna highlighting (in)transitivity in Mandarin L1 – English L2 participants (neither Mandarin nor English obligatorily marks [in]transitivity as in Karitiâna) and if so, whether such a cognitive shift might vary with instructional conditions. Our study will therefore provide a fuller picture of the malleability of human cognition by extending the line of inquiry to the factors influencing language–cognition interactions.

To the best of our knowledge, previous training studies on language–cognition interactions have not investigated whether brief training on linguistic labels beyond whole words, such as grammatical morphemes, can also induce cognitive shifts. Besides being an underinvestigated domain of motion event cognition, our study on morphemes highlighting motion (in)transitivity can also extend this line of inquiry.

In Experiment 1 we chose to compare two inductive instructional conditions, testing our prediction above for the comparison between meaning-focused and form-focused (required rule search) instructional conditions. That is, when subgroups of participants with equivalent awareness at the level of understanding and near-perfect GJT performance are considered, a meaning-focused instructional condition will be more likely to lead to a cognitive shifting effect than a required rule search instructional condition (Hypothesis 1).

To preview the results of Experiment 1, Hypothesis 1 was borne out. Based on this, we conducted Experiment 2 to see if additional GJT practice would benefit cognitive shifting in the required rule search instructional condition because the GJT could arguably help to further internalize the knowledge (in the sense outlined above).

Experiment 3 went on to test the prediction based on Morgan-Short and colleagues' studies (see above); that is, an inductive instructional condition might be more effective than metalinguistic explanation in terms of inducing cognitive shifts, again when subgroups of participants with equivalent awareness at the level of understanding and near-perfect GJT are considered (Hypothesis 2). More specifically, the inductive instructional condition in Experiment 3 was required rule search because it does not make sense to test for cognitive shifts immediately after a metalinguistic explanation without any form-focused practice (e.g., GJT). Therefore, ideally, the two groups should

https://doi.org/10.1017/S0272263124000433 Published online by Cambridge University Press

be matched on such form-focused GJT practice. This precludes the meaning-focused instructional condition because it will no longer be a purely meaning-focused intervention with the addition of such form-focused GJT practice.

## Experiment 1

All experiments were conducted online and built using the Gorilla Experiment Builder (Anwyl-Irvine et al., 2020). All materials are available at: https://osf.io/t3vx5/?view_only=89c2a2cdae05449ab125d7757880dca9.

Experiment 1 investigates whether the effects of learning the novel system on triads-matching will differ between meaning-focused (memorizing meanings of exemplar sentences) and form-focused (required rule search) instructional conditions. See Figure 1 for an overview of the design.

### *Participants*

One hundred and forty-three Mandarin L1 English L2 speakers were divided randomly into four groups: the form-focused experimental group ($N = 39$, $M_{age} = 22.10$, 17 males); the form-focused control group ($N = 28$, $M_{age} = 21.32$, 8 males); the meaning-focused experimental group ($N = 51$, $M_{age} = 21.75$, 16 males); and the meaning-focused control group ($N = 25$, $M_{age} = 21.24$, 8 males). Due to the Chinese national curriculum, all participants had learned English for more than 12 years and had passed the English test in the national university entrance exam. No participants reported knowing any third language. Different numbers of participants and different gender ratios in each group were because different instructional conditions led to different proportions of participants that qualified for the analysis.

### *Methods*

#### *Step 1: Training*

The training materials for the two experimental groups were 128 Mandarin sentences with *ro* and *gi* prefixed to transitive and intransitive verbs respectively (same as the training materials used in Xue & Williams (2024)). Half of the sentences were

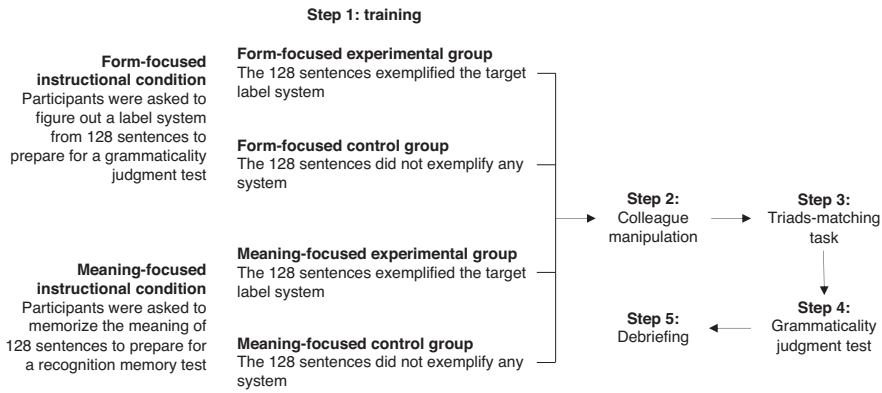



**Figure 1.** Overview of Experiment 1

https://doi.org/10.1017/S0272263124000433 Published online by Cambridge University Press

**Table 1.** Examples of training and training materials (le-PST = past tense marker; ACC = accusative case marker *ba*[4])

| Experimental groups | Control groups |
|---|---|
| **Transitive** | |
| 死寂的 废墟 上, 官兵 把 伤者<br>Quiet ruin on, soldier(s) ACC- the injured<br>ro 拯救 了。<br>person (people) ro rescue le-PST.<br>On the quiet ruin, the soldier(s) ro rescued the injured person (people). | 死寂的 废墟 上, 官兵 把 伤者<br>Quiet ruin on, soldier(s) ACC- the injured<br>ro 拯救 了。<br>person (people) ro rescue le-PST.<br>On the quiet ruin, the soldier(s) ro rescued the injured person (people). |
| 胶囊 把 症状 ro 改善 了。<br>Capsule(s) ACC-symptom(s) ro relieve le-PST.<br>The capsule(s) ro relieved the symptom(s). | 胶囊 把 症状 gi 改善 了。<br>Capsule(s) ACC-symptom(s) gi relieve le-PST.<br>The capsule(s) gi relieved the symptom(s). |
| **Intransitive** | |
| 前一夜, 狡诈的 囚犯 gi 逃窜 了。<br>Last night, cunning prisoner(s) gi escape le-PST.<br>Last night, the cunning prisoner(s) gi escaped. | 前一夜, 狡诈的 囚犯 gi 逃窜 了。<br>Last night, cunning prisoner(s) gi escape le-PST.<br>Last night, the cunning prisoner(s) gi escaped. |
| 本 季度, 股价 gi 上涨 了。<br>This quarter, stock price(s) gi increase le-PST.<br>This quarter, the stock price(s) gi increased. | 本 季度, 股价 ro 上涨 了。<br>This quarter, stock price(s) ro increase le-PST.<br>This quarter, the stock price(s) ro increased. |

intransitive while the other half were transitive. To ensure that only (in)transitivity but not sentence length could predict *gi/ro* use, we manipulated the existence and length of prepositional phrases and modifiers so that intransitive and transitive sentences could have 6/9/12/15 and 9/12/15/18 characters respectively. This manipulation also prevented participants from predicting the use of *gi/ro* based on the number of heads or noun phrases (NP) in a sentence. Additionally, because Mandarin does not have obligatory plurality markers, in most sentences the plurality of the subject and object NPs remained ambiguous. A group of 20 Mandarin native speakers who were not involved in the main study and were unaware of its purpose judged that only 11.00 (*se* = 0.32) of the 64 intransitive sentences had an unambiguously singular subject NP, and that only 10.20 (*se* = 0.39) of the 64 transitive sentences had subject and object NPs that were unambiguously singular. As a result, we can predict the use of *gi/ro* based on the number of argument(s) but not the raw number of entities in a sentence. This corresponds with the conventional definition of intransitive predicates as having only one argument and transitive predicates as having two arguments (Bowers, 2002).

To rule out that mere exposure to the 128 sentences without learning the *gi/ro* rule could already produce an effect on triads-matching, the training materials for the two

[4]In Mandarin, *ba* is an optional accusative case marker (Cheung, 1973). There is no obligatory case or (in) transitivity marker in Mandarin. No accusative case marker is allowed in the canonical SVO word order of Mandarin. However, this canonical word order (i.e., "S ro-V O" and "S gi-V") may lead participants to induce other systems that also correctly predict *gi/ro* usage but are unrelated to (in)transitivity, e.g., sentences with *gi* and *ro* end with a noun and a verb respectively. In our pilot studies, using artificial word order failed to resolve these issues, because it hindered the learning of the target label system. Therefore, we adopted a noncanonical but grammatical word order to solve these problems: "S *ba*-O V" and "S V," where *ba* denotes the accusative case of the following noun. Crucially, the experimental and control groups were equated on every aspect except the distribution of *gi/ro*. Therefore, any effect on triads-matching should result from this label system per se.

https://doi.org/10.1017/S0272263124000433 Published online by Cambridge University Press

**Table 2.** Examples of the stimuli for the triads-matching task

| | Target | Alternative 1 | Alternative 2 |
|---|---|---|---|
| Triad 1 |  |  (in)transitivity-matched |  colour-matched |
| Triad 2 |  |  direction-matched |  (in)transitivity-matched |
| Triad 3 |  |  size-matched |  (in)transitivity-matched |
| Triad 4 |  |  shape-matched |  (in)transitivity-matched |
| Triad 5 (filler) |  |  colour-matched |  shape-matched |

control groups were also the 128 sentences. But in these sentences, *gi* appeared in half of both transitive and intransitive sentences, while *ro* appeared in the other half of transitive and intransitive sentences, i.e., no system governed the use of *gi* and *ro* at all. See Table 1 for examples.

The form-focused groups were asked to figure out a rule governing *gi/ro* usage from the training sentences by themselves to prepare for a GJT. The meaning-focused groups were asked to memorize the meaning of the training sentences to prepare for a recognition memory test on the events described. They were told nothing about *gi/ro* nor anything related to the novel system. The sentences were presented both auditorily and visually.

*Step 2: "colleague manipulation"*

All participants were informed that the subsequent parts were a separate and unrelated experiment by our colleague and that we were only helping to recruit participants. Once they agreed to participate, they received a link to the subsequent parts. This "colleague manipulation" was implemented to hide the connection between the preceding part and the triads-matching task. We anticipated that it would reduce the use of strategies by participants to perform triads-matching in a way that they believed would satisfy the experimenter. This was done to address a concern with previous training studies, where participants were likely to have been aware of the relationship between the

https://doi.org/10.1017/S0272263124000433 Published online by Cambridge University Press

language training and the subsequent measurement of cognitive shifts, making it possible that any effects observed were at least partially due to strategic behavior. The "colleague manipulation" was approved by the departmental research ethics committee.

*Step 3: triads-matching task*

Loosely following Everett (2013), this task contained 16 triads of pictures denoting motion events. In each triad, the first picture was the Target. Half of the Targets depicted one figure hitting another figure along the direction of an arrow (transitive motion, e.g., Triads 1 and 2 in Table 2), while the other half depicted one figure moving alone along the arrow (intransitive motion, e.g., Triads 3 and 4 in Table 2). The following two pictures in a triad were Alternatives presented side-by-side that each differed from the Target in only one of the dimensions among shape, color, size, (in)transitivity, and motion direction (shown by arrow direction).

Participants received the following task instruction (here translated from Mandarin): "Next you will see 16 groups of pictures describing motion events. Each group consists of three pictures. When there is only one figure in the picture, it describes the figure moving alone along the direction of the arrow. When there are two figures in the picture, it describes one figure hitting the other figure along the direction of the arrow. These figures have different shapes, colors, and sizes. In each group, you will first see one picture representing a motion event, followed by two pictures representing motion events that are placed side-by-side. Between these two, please select the one that you think is more similar to the first motion event. Please select quickly based on your intuition." All choices were self-paced following triads-matching tasks in previous studies (see Introduction).

For example, in Triad 1 in Table 2, the Target is one big black circle hitting another from left to right. Alternative 1 only differs from the Target in color, while Alternative 2 only differs from the Target in (in)transitivity. Therefore, selecting Alternative 1 would indicate categorizing based on (in)transitivity, whereas selecting Alternative 2 would indicate categorizing based on color.

Among the 16 triads, 8 triads had a (in)transitivity-matched Alternative. The critical dependent variable is the likelihood of choosing the (in)transitivity-matched Alternative for these triads, i.e., the likelihood of categorization based on (in)transitivity in these triads. The other eight triads were fillers in which both Alternatives were (in)transitivity-matched (e.g., Triad 5 in Table 2). Alternatives matching the Target in each distractor dimension (color, direction, size, shape) were counterbalanced.

Pictures were used instead of videos because videos of different (in)transitivity and figure sizes introduced artifacts including different motion speeds, length of movement, or width of frame. In our pilot studies, some participants reported to have based their categorization on these artifactual dimensions. We acknowledge that using static pictures might increase cognitive demands and might thus introduce a confound. Yet this would be the same for both experimental and control participants. Therefore, it seems unlikely that this would affect the critical group difference. To further validate this task, we ran a test in which 22 participants who were from the same pool but did not participate in the main experiments first performed the triads-matching task without any training, followed by a verbal description task where they described all pictures used and how each Alternative was different from the Target. Next, they rated on a 7-point Likert scale the perceptibility of each dimension and the ease of seeing pictures as motions according to the task instruction. The results of this pretest confirmed

https://doi.org/10.1017/S0272263124000433 Published online by Cambridge University Press

that static pictures could be easily seen as motions according to the task instructions ($M$ = 6.50, $se$ = 0.11); all dimensions were equally highly perceptible ($M$ > 6 for all dimensions); all transitive stimuli were verbally described using transitive constructions and intransitive stimuli intransitive constructions; in all cases, participants correctly identified the dimension in which each Alternative differed from the Target; crucially, in all cases where this dimension was (in)transitivity, participants' verbalization included not only the number of entities but also (in)transitivity (i.e., the interaction between the entities, e.g., "the target and one alternative involve one entity hitting the other, whereas the other alternative doesn't").

*Step 4: grammaticality judgment test (GJT)*

All participants performed the self-paced GJT consisting of 32 new sentences structured in the same way as the training sentences. Half were grammatical based on the use of *gi/ro*, while the other half were ungrammatical. To ensure that specific sentence wordings could not influence GJT results, we randomly allocated participants to one of two sets of GJT sentences, each comprising 32 sentences but with reverse grammaticality (sentences bearing the grammatical label in the first set had the ungrammatical label in the second set and vice versa).

The form-focused groups were asked to perform the GJT according to the rule they figured out before. The meaning-focused groups were told that actually there was a rule governing *gi/ro* usage and that they would perform the GJT according to their intuition formed about this rule. After each judgment, all participants were asked to indicate how confident (very confident vs. not very confident) they were in their judgment.

*Step 5: Debriefing*

All participants were debriefed on three questions:

a. If they found the grammar of *gi* and *ro*, what it is;
b. If they found the grammar, they were asked to indicate when they found it by dragging a slider bar that schematized the whole experimental procedure;
c. If they thought the tasks before and after "the colleague manipulation" were related and if yes, how?

*Participant exclusion criteria*

Since our study investigates the effect of learning novel labels on subsequent triads-matching, logically experimental group participants who showed no learning of the labels were excluded from further analysis. We operationalized the notion of "having learned the labels" with the following three criteria: (1) making 22 or more correct choices out of 32 GJT trials (following a binomial test showing that a participant's GJT performance was better than chance when they made 22 or more correct choices out of 32 total trials, $p$ < .05, one-tailed); *and* (2) displaying awareness of the target system at the level of understanding (see Schmidt, 1990) defined as mentioning in the debriefing the relationship between *gi/ro* with at least one of the following:

a. the (in)transitivity of the actions/sentences;
b. whether the action/sentence involves one participating body doing something alone or it is something happening between two participating bodies;
c. whether the action/sentence only involves a subject or both a subject and object;

https://doi.org/10.1017/S0272263124000433 Published online by Cambridge University Press

d. the existence of the accusative case marker *ba*;
e. the activeness and passiveness of the participating bodies.

And (3) not mentioning any incorrect system irrelevant to the target system such as the usage of *gi/ro* being governed by the animacy of the agent and/or patient of the sentence.

The rationale for these exclusion criteria is based on Cleeremans's (2008) proposal that the development from unaware to aware representation of a linguistic system at the level of understanding might be a gradual enhancing process. In the present case, if we only applied the GJT criterion (criterion 1), we would include participants both with and without awareness of the target system at the level of understanding. For those without awareness, it is difficult to quantify and thus enter into statistical models their stage in the gradual process of the development of conscious awareness. Some participants may also exceed the GJT threshold by applying rules based on different criteria such as animacy, given the intrinsic relationship between animacy and subject/object in natural languages. We would not expect criteria such as animacy to induce an increase in the likelihood of (in)transitivity-based categorizations. On the other hand, if we only applied the awareness criterion (criterion 2), we would include participants who only became aware during or after the GJT, for example through reflection during the debriefing (see *the information criterion* in Shanks and St. John, 1994). Therefore, we would again include participants in varying stages of the development of awareness of the target label system when performing triads-matching.

Our aim is to investigate the role of instructional conditions per se in triads-matching. The degree to which subsequent cognition might be affected by the stage in the gradual development of conscious awareness of the relevant linguistic system remains largely unknown. Therefore, we attempt to maximally control for this tacit variable that is difficult to quantify and enter into statistical models. We thus only included experimental group participants who met *all three* criteria above.

In addition, to rule out potentially strategic participants, we excluded any participants aware of the relationship between the 128 sentences and triads-matching, defined as mentioning in the debriefing that:

a. Number of figures or the activeness/passiveness of the figures in the triads-matching task *and*
b. The analogy to the sentences/labels *and*
c. They had this feeling *during* triads-matching, i.e., this feeling was not a result of reflection during debriefing.

Even though at first sight the participant exclusion criteria seem radical, we believe it is the logical way to address our research question.

## Results

*GJT and verbal report:*
*form-focused experimental group.* Fourteen out of 39 form-focused experimental group participants were excluded for not having met the aforementioned learning criteria. They made on average 16.79 out of 32 correct GJT judgments ($SD$ = 4.41), which was not significantly different from chance, $t$ (13) = .667, $p$ = .516. One of the remaining 25 participants was excluded for being potentially strategic. Thus 24 participants (eight males) from the form-focused experimental group entered further analysis, making on average 30.79 correct GJT judgments ($SD$ = 1.91).

https://doi.org/10.1017/S0272263124000433 Published online by Cambridge University Press

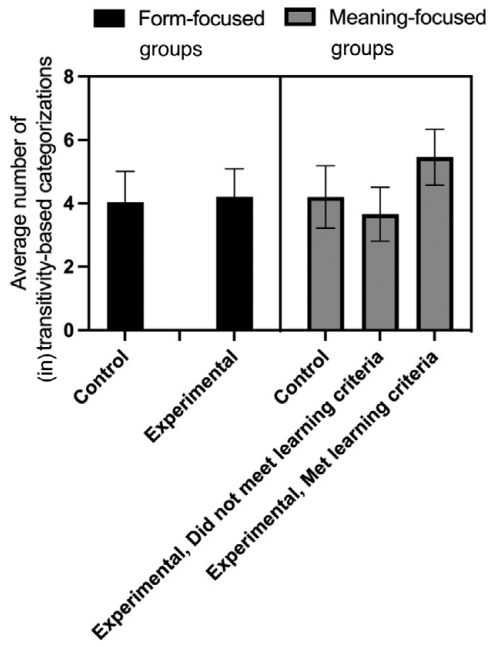



**Figure 2.** Average number of (in)transitivity-based categorizations made by each group during triads-matching in Experiment 1 (error bars indicate 95% CI)

*Form-focused control group.* Two participants were excluded for being potentially strategic. Thus 26 form-focused control participants entered the analysis on triads-matching (eight males), making on average 14.65 correct GJT judgments (*SD* = 3.93).

*Meaning-focused experimental group.* Twenty-seven (11 males) out of 51 meaning-focused experimental group participants failed to meet the learning criteria. This rate is as expected since the learning is incidental. They made on average 17.08 correct GJT judgments (*SD* = 5.53), which was not significantly different from chance, $t$ (25) = .993, $p$ = .330. One of them reported being aware of the relationship between the training and triads-matching tasks. Crucially, since this subgroup was large enough, we included it (26 participants) in the analysis on triads-matching as the "meaning-focused, experimental, did not meet learning criteria" subgroup.

Twenty-four meaning-focused experimental group participants met the learning criteria. Two of them were excluded for being potentially strategic. Thus 22 participants entered further analysis as the "meaning-focused, experimental, met learning criteria" subgroup (5 males), making on average 30.32 correct GJT judgments (*SD* = 1.81).

*Meaning-focused control group.* The 25 participants (8 males) made on average 14.68 correct GJT judgments (*SD* = 4.88).

*Triads-matching task.* Figure 2 shows the average number of (in)transitivity-based categorizations made by each group. We fit mixed logit models with the lme4 package (Bates et al., 2015) in R (version 4.2.1; R Core Team 2022) to predict participants' likelihood of making (in)transitivity-based categorizations during triads-matching. Specifically, for the dependent variable Outcome, a participant categorizing based on (in)transitivity (choosing the [in]transitivity-matched Alternative) in a triad was coded as 1, while categorization based on other dimensions was coded as 0. Mixed

https://doi.org/10.1017/S0272263124000433 Published online by Cambridge University Press

**Table 3.** The optimal model comparing the form-focused control and experimental groups in Experiment 1 (Outcome ~ Group + [1|Participants] + [1|Trials])

| Effect | Estimate | Standard error | l–95% CI | u–95% CI | *z* value | *p* |
|---|---|---|---|---|---|---|
| (Intercept) | 0.02 | 0.41 | –0.77 | 0.82 | 0.06 | 0.953 |
| group (compared with the control group) | 0.14 | 0.44 | –0.72 | 0.99 | 0.32 | 0.752 |

**Table 4.** The optimal model comparing the meaning-focused control and experimental groups in Experiment 1 (Outcome ~ Group + [1|Participants] + [1|Trials])

| Effect | Estimate | Standard error | l–95% CI | u–95% CI | *z* value | *p* |
|---|---|---|---|---|---|---|
| (Intercept) | 0.16 | 0.41 | –0.65 | 0.97 | 0.38 | 0.704 |
| [did not meet learning criteria] subgroup | –0.40 | 0.43 | –1.23 | 0.44 | –0.93 | 0.353 |
| [met learning criteria] subgroup | 0.94 | 0.45 | 0.05 | 1.82 | 2.07 | 0.038 |

logit models need at least fifteen observations per regression coefficient (Levshina, 2015). All models in this study exceed this criterion.

We first report the mixed logit model comparing the form-focused experimental and control groups. This model had Group as the fixed effect variable, and Participants and the trials of the triads-matching task (abbreviated as "Trials," note that "trial" is the same as "triad") as random effects. Following Wigdorowitz et al., (2023), for every model in this study, to obtain the optimal random structure, we ran analysis of variance tests (ANOVAs) to compare the models with all possible random structures. The converging model with the lowest AIC and BIC values was chosen. In all cases, this was the model with Participants and Trials as random intercepts without any random slope. Here, the selected model (Table 3) showed no significant effect of learning the target system under form-focused instructional conditions on the likelihood of making (in)transitivity-based categorizations.

Similarly, we next report the optimal mixed logit model comparing the meaning-focused control group, the [meaning-focused, experimental, did not meet learning criteria] subgroup and the [meaning-focused, experimental, met learning criteria] subgroup (Table 4). This model also had Group as the fixed effect, and Participants and Trials as random intercepts. We report the odds ratio exponentiated from the coefficient for better interpretability. This model showed that the [meaning-focused, experimental, met the learning criteria] subgroup was more likely to make (in) transitivity-based categorizations than the control group by an odds ratio of 2.55 (l–95% CI = 1.05; u–95% CI = 6.20). However, participants who did not meet the learning criteria did not show this tendency.

To directly illustrate the effect of instructional conditions, we also ran several mixed logit models comparing the form-focused experimental group with the [meaning-focused, experimental, met learning criteria] subgroup. These two groups received the same amount of input and both showed near-perfect GJT accuracy and awareness of the target system at the level of understanding. These models had Group and number of correct judgments in the GJT (abbreviated as GJT) as fixed effects and Participants and Trials as random effects. Following Wigdorowitz et al. (2023), we first determined the optimal random structure by running ANOVAs between the models with all possible random structures while retaining the maximal fixed structure. After obtaining

https://doi.org/10.1017/S0272263124000433 Published online by Cambridge University Press

**Table 5.** The optimal model comparing the form-focused experimental group and the [meaning-focused, experimental, met learning criteria] subgroup in Experiment 1 (Outcome ~ Group + [1|Participants] + [1|Trials])

| Effect | Estimate | Standard error | l–95% CI | u–95% CI | *z* value | *p* |
|---|---|---|---|---|---|---|
| (Intercept) | 0.16 | 0.42 | –0.67 | 0.98 | 0.37 | 0.713 |
| group (compared with the form-focused experimental group) | 0.93 | 0.44 | 0.07 | 1.79 | 2.12 | 0.034 |

the optimal random structure (i.e., Participants and Trials as random intercepts without any random slope), we compared three models with different fixed structures:

Model 1: Outcome ~ Group + GJT + Group: GJT + (1|Participants) + (1|Trials)
Model 2: Outcome ~ Group + GJT + (1|Participants) + (1|Trials)
Model 3: Outcome ~ Group + (1|Participants) + (1|Trials)

No model showed any significant effects of GJT or interaction between GJT and Group. Model 3 was favored for having the lowest AIC and BIC values (Table 5). Model 3 showed that the [meaning-focused, experimental, met learning criteria] subgroup was more likely to make (in)transitivity-based categorizations than the form-focused experimental group by an odds ratio of 2.54 (l–95% CI = 1.07; u–95% CI = 6.01).

*Progress bar.* This task may be subject to memory bias and has not been validated as a psychometric test. Therefore, we only performed an exploratory analysis which suggested that the form-focused experimental group and the [meaning-focused, experimental, met learning criteria] subgroup became aware of the label system at the level of understanding after 57.75 (*se* = 6.98) and 79.45 (*se* = 9.02) sentences on average. The group difference was marginally significant, $t$ (44) = -1.92, $p$ = .061.

An exploratory two-way mixed ANOVA on the time spent per training sentence in experimental group participants who met the learning criteria revealed a significant interaction between group (form- vs. meaning-focused) and stage (pre- vs. post- the point of becoming aware of the target system at the level of understanding), $F$ (1, 38) = 7.36, $p$ = .010, $\eta_p^2$ = .16. Planned comparisons showed that after the awareness point, the form-focused group spent significantly shorter time per training sentence than the meaning-focused group, $F$ (1, 38) = 6.25, $p$ = .017, $\eta_p^2$ = .14. The form-focused group showed a significant drop in the time spent per training sentence after the awareness point, $F$ (1, 22) = 27.86, $p$ < .001, $\eta_p^2$ = .56. No other significant effects were found.

## Experiment 2

Experiment 2 investigated whether participants who learned the target system under a required rule search instructional condition and then practiced using it in the GJT before the triads-matching task would show a cognitive shifting effect.

### *Participants*

Another 56 Mandarin L1–English L2 speakers from the same participant pool but who had not participated in Experiment 1 were randomly allocated to two groups: the form-focused experimental group ($N$ = 31, $M_{age}$ = 21.77, 11 males) and the form-focused control group ($N$ = 25, $M_{age}$ = 21.24, 8 males).

https://doi.org/10.1017/S0272263124000433 Published online by Cambridge University Press

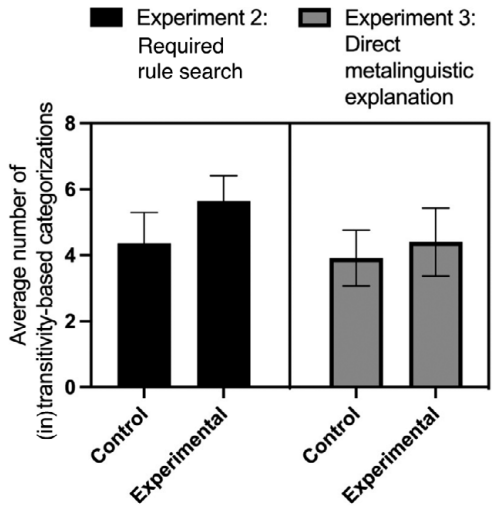



**Figure 3.** Average number of (in)transitivity-based categorizations made by each group in Experiments 2 and 3 (error bars indicate 95% CI)

**Table 6.** The optimal model comparing the experimental and control groups in Experiment 2 (Outcome ~ Group + [1|Participants] + [1|Trials])

| Effect | Estimate | Standard error | l–95% CI | u–95% CI | *z* value | *p* |
|---|---|---|---|---|---|---|
| (Intercept) | 0.28 | 0.37 | –0.45 | 1.00 | 0.75 | 0.452 |
| group (compared with the control group) | 0.89 | 0.41 | 0.08 | 1.70 | 2.15 | 0.032 |

## *Methods*

Procedures for the form-focused experimental and control groups of Experiment 2 had the following two major changes from those of Experiment 1: First, Experiment 2 had the following procedure: Step 1: form-focused training → Step 2: GJT → Step 3: "colleague manipulation" → Step 4: triads-matching task → Step 5: debriefing. Second, the progress bar was removed.

## *Results*

### *GJT and verbal reports*

Five experimental group participants were excluded for not having met the learning criteria. They made on average 15.50 correct GJT judgments (*SD* = 2.12). Another three experimental group participants were excluded for being potentially strategic. The remaining 23 experimental group participants entered further analysis (8 males). They made on average 31.00 correct GJT judgments (*SD* = 1.93). The 25 control participants made on average 14.00 correct GJT judgments (*SD* = 5.27).

### *Triads-matching task*

Figure 3 shows the average number of (in)transitivity-based categorizations made by each group. Identical to previous analyses, we report the optimal mixed logit model

https://doi.org/10.1017/S0272263124000433 Published online by Cambridge University Press

**Table 7.** The optimal model comparing the experimental and control groups in Experiment 3 (Outcome ~ Group + [1|Participants] + [1|Trials])

| Effect | Estimate | Standard error | l–95% CI | u–95% CI | *z* value | p |
|---|---|---|---|---|---|---|
| (Intercept) | –0.03 | 0.41 | –0.84 | 0.78 | –0.07 | 0.942 |
| group (compared with the control group) | 0.32 | 0.44 | –0.55 | 1.19 | 0.72 | 0.472 |

with Group as the fixed effect and Participants and Trials as the random intercepts to predict participants' likelihood of making (in)transitivity-based categorizations (Table 6). This model showed that, when the GJT preceded triads-matching, participants under the form-focused instructional condition (required rule search) showed an effect on triads-matching: the experimental group was more likely to make (in)transitivity-based categorization than the control group by an odds ratio of 2.43 (l–95% CI = 1.08; u–95% CI = 5.48).

## Experiment 3

Experiment 2 found that participants under a required rule search instructional condition showed an effect on triads-matching when the GJT preceded triads-matching. Experiment 3 investigates whether this effect persists if we employ another type of form-focused instructional condition: direct metalinguistic explanation.

### *PARTICIPANTS*

Another 50 Mandarin L1–English L2 speakers from the same participant pool but who had not participated in Experiment 1 or 2 were randomly allocated to two groups: the experimental group ($N$ = 26, $M_{age}$ = 21.81, 8 males) and the control group ($N$ = 24, $M_{age}$ = 21.71, 10 males).

The experimental group received the following metalinguistic explanation (here translated from Mandarin): "Next you will learn a new grammar: When there are two participating parties in the action described in a sentence (i.e., the sentence is transitive), add *ro* before the verb. When there is only one participating party in the action described in a sentence (i.e., the sentence is intransitive), add *gi* before the verb." They also read 4 example sentences, two containing *gi* and two containing *ro*. The control group did not receive any training. The concept of (in)transitivity is covered in China's compulsory education curriculum standard (PRC Ministry of Education, 2013), so should be familiar to the participants.

Next, the experimental group was asked to perform the GJT based on the label system. The control group was asked to judge whether the *gi*/*ro* in each sentence was used correctly solely based on intuition. The other procedures and materials were identical to Experiment 2.

### *Results*

#### *GJT and verbal report*

One experimental group participant was excluded for only having made 20 out of 32 correct GJT judgments, leading to 25 participants entering further analysis (8 males). They made on average 30.48 correct GJT judgments ($SD$ = 1.73).

https://doi.org/10.1017/S0272263124000433 Published online by Cambridge University Press

**Table 8.** The optimal model directly comparing the effect of required rule search vs. direct metalinguistic explanation instructional condition on the likelihood of making (in)transitivity-based categorization (Outcome ~ Group + [1|Participants[ + [1|Trials])

| Effect | Estimate | Standard error | l–95% CI | u–95% CI | *z* value | *p* |
|---|---|---|---|---|---|---|
| (Intercept) | 0.30 | 0.45 | –0.58 | 1.18 | 0.67 | 0.505 |
| group (compared with direct metalinguistic explanation) | 0.97 | 0.48 | 0.03 | 1.92 | 2.02 | 0.044 |

One control participant was excluded for not understanding the requirements. The 23 remaining control participants (9 males) made on average 14.61 correct GJT judgments (*SD* = 4.09).

### *Triads-matching task*

Figure 3 shows the average number of (in)transitivity-based categorizations made by each group. Identical to previous analyses, we report the optimal mixed logit model with Group (experimental vs. control) as the fixed effect and Participants and Trials as random intercepts to see if participants receiving direct metalinguistic explanation also showed an effect on triads-matching (Table 7). However, no significant effect on triads-matching was found.

### *Cross-experimental comparison*

To directly illustrate the effect of the two types of form-focused instructional conditions (required rule search in Experiment 2 vs. direct metalinguistic explanation in Experiment 3) on triads-matching, we also ran mixed logit models directly comparing the two experimental groups of Experiment 2 and 3. These models had Group and GJT as fixed effects and Participants and Trials as random effects. To obtain the best random and fixed structure, the aforementioned procedure was applied. No model showed any significant effects of GJT or any interaction between GJT and Group. The optimal model (Table 8) showed that, with the GJT preceding triads-matching, participants who learned the labels through required rule search were more likely to make (in)transitivity-based categorizations than those who learned through direct metalinguistic explanation by an odds ratio of 2.65 (l–95% CI = 1.03; u–95% CI = 6.83).

The likelihood of making (in)transitivity-based categorizations was at a chance level for all control groups (the passive control group without any intervention and the control group(s) of each experiment) and no difference was found among them, as assessed by a mixed logit model with Group as the fixed effect variable and Participants and Trials as the random intercepts, all $p > .05$ and the 95% CI of the estimated marginal mean of each control group containing 0. For each model in this study yielding a null result, a Bayesian mixed logit model with the same variable structure and default priors in JASP (JASP Team, 2022) showed qualitatively similar results. All control groups who performed a GJT performed at chance level, all $p > .05$.

## General discussion

In this study, we investigated whether learning two novel labels highlighting the familiar concept of (in)transitivity would induce cognitive shifts on subsequent motion

https://doi.org/10.1017/S0272263124000433 Published online by Cambridge University Press

event triads-matching, and if yes, whether this effect would differ between instructional conditions. Among participants who met the learning criteria, Experiment 1 found cognitive shifts on subsequent triads-matching only in those under a meaning-focused but not in those under a form-focused (required rule search) instructional condition (Hypothesis 1 was borne out). However, Experiment 2, building on Experiment 1, found that participants under a required rule search instructional condition showed this effect after performing additional GJT practice. Experiment 3, building on Experiment 2, found that this effect was significantly reduced to a nonsignificant level in participants under another type of form-focused instructional condition - direct metalinguistic explanation (Hypothesis 2 was borne out).

Due to the "colleague manipulation," it is highly unlikely that participants would verbalize the stimuli during triads-matching using the *gi/ro* labels. Of course, we cannot rule out online involvement of general verbalization. However, the degree of such verbalization should be the same for both experimental and control participants and thus cannot account for the group differences in triads-matching.

With respect to previous SLA studies on language–cognition interactions, in which lab-based training studies are usually situated, we expanded the independent variables to instructional conditions. Additionally, we showed that not only whole labels but also grammatical morphemes could induce cognitive shifts through brief training.

With respect to previous studies on motion event cognition, we extended Everett's (2013) comparison of native speakers of Karitiâna and English, showing that grammatical morphemes highlighting (in)transitivity could indeed bias motion event categorization, even after controlling for multiple confounds suffered by cross-linguistic comparisons by using the lab training paradigm. More generally speaking, while previous studies focused on two domains of motion event cognition (endpoint vs. ongoingness and manner vs. path), we provided evidence of language–cognition interactions in a very underinvestigated domain of motion event cognition - motion (in)transitivity as opposed to other dimensions.

With respect to studies comparing meaning- and form-focused instructional conditions, we expanded the focus from language attainment to cognitive shifts. This is among the few studies to suggest enhanced internalization of novel linguistic knowledge following meaning-focused compared with form-focused instructional conditions, especially direct metalinguistic explanation. Most importantly, this advantage lies in a tacit aspect indexed by cognitive shifts even among participants who have all achieved awareness of the target knowledge at the level of understanding and near-perfect GJT accuracy.

Given the conventional explanations of cognitive shifts as label- or structural-feedback (see Introduction), how do we explain the existence and absence of such an effect under different instructional conditions in our study? Following Calderón (2013) (see Introduction), we hypothesized that in Experiment 1 it could be that after achieving awareness at the level of understanding, those under the meaning-focused instructional condition may have continued with the deep processing of upcoming training stimuli to satisfy the task requirement of memorizing sentence meaning, but those under the form-focused instructional condition may have resorted to shallow processing, as they believed they had satisfied the task requirement. This speculation is indeed corroborated by the analysis of the average time spent on each training trial (note that the time spent cognitively on processing incoming information is part of the aforementioned definition of DoP): a significant decrease in the average time spent per training trial after achieving awareness at the level of understanding was observed in participants under a form-focused but not meaning-focused instructional condition. Based on

https://doi.org/10.1017/S0272263124000433 Published online by Cambridge University Press

Leow's (2015) model, overall higher DoP in participants under our meaning-focused than form-focused instructional condition may have contributed to overall more internalized knowledge of the target system in the former than the latter group, which, through label- and/or structural-feedback, may have given rise to the observed difference in triads-matching. This speculation echoes previous studies suggesting that higher DoP was associated with not only superior performance on immediate posttests (e.g., Godfroid & Schmidtke, 2013; Godfroid et al., 2013; Thinglum et al., 2019; Martin et al., 2019; Zhuang, 2019) but also better retention of L2 knowledge over time (e.g., Li, 2019; Leow et al., 2019; Cerezo et al., 2016; Hsieh et al., 2016; Zhuang, 2019; Leow, 1998).

Experiment 2, building on Experiment 1, showed that participants under the same form-focused instructional condition but after performing additional GJT practice showed the effect on triads-matching, presumably because the GJT increased the activation of the combinatorial nodes of (in)transitivity and/or strengthened the relevant connections just before triads-matching.

Finally, regarding Experiment 3, previous studies suggested that the overall DoP tended to be higher in learners who arrived at the target system through rule search and practice than those who received direct metalinguistic explanations followed by practice (Leow et al., 2019; Martin et al., 2019). While the dominant strategy of the latter group was to shallowly repeat and execute the rules provided, the former group has been found to engage in deep processing including hypothesis testing, rule formation, activation of recent prior knowledge, and metacognitive processes (Leow et al., 2019). Therefore, the overall DoP may have been lower in the experimental group in Experiment 3 than in Experiment 2, which might have given rise to the reduction of cognitive shifts to a nonsignificant level in Experiment 3.

Overall, our meaning-focused and required rule search instructional conditions seem analogous to the learner-centered instructional condition in ISLA literature, and our direct metalinguistic explanation analogous to the teacher-centered instructional condition in the literature. Our findings echoed with previous ISLA literature comparing learner- and teacher-centered instructional conditions, where the former have been found to produce not only superior immediate learning outcomes but also better retention of L2 knowledge (e.g., Hsieh et al., 2016; Cerezo et al., 2016; Leow, 1998). According to Leow's (2015) model, the learner-centered groups, who logically had to engage in deeper processing than the teacher-centered group, may have developed more internalized knowledge of the target system than the teacher-centered group. This difference in the internal system of knowledge, which the GJT may not be sensitive to, might have nonetheless been reflected in triads-matching. Therefore, pedagogically speaking, our findings echo recent calls for the involvement of more learner-centered tasks in ISLA that encourage high DoP (e.g., Cerezo et al., 2016; Leow et al., 2019), which may, in turn, promote not only awareness at the level of understanding and internalization of the knowledge but also cognitive shifts.

### *Limitations and future directions*

The speculative explanations we suggested above require future testing. This study, by suggesting that the effect of linguistic systems on cognition may vary with instructional conditions, may be the first step in a line of future research that probes the underlying mechanism of this phenomenon. First, online measurements of participants' cognitive processes during training (e.g., thinking aloud and eye-tracking) are necessary to verify

https://doi.org/10.1017/S0272263124000433 Published online by Cambridge University Press

our explanations, especially in relation to the identification of when awareness at the level of understanding was reached during training and the monitoring of changes in DoP thereafter. Second, it remains unclear what effect the Mandarin accusative case marker had in the present study. As case markers are common in world languages, future studies can recruit speakers of other languages with different case markers to investigate the effect of prior linguistic knowledge and how it interacts with novel linguistic knowledge in this domain.

Note that though we made the minimal working assumption that (in)transitivity is represented by abstract combinatorial nodes to provide one possible rationale and explanation of our study, we by no means preclude other possible representations of (in)transitivity as suggested by various schools of linguistic theory (e.g., Bowers, 2002; Hopper & Thompson, 1980, see Introduction). From here, another issue raised by this study is whether any specific kind of linguistic representation of (in)transitivity can modulate subsequent triads-matching to a greater extent than other kinds of representation. It is indeed difficult to measure individual differences in such tacit, underlying representations. Though we used verbal debriefing in this study, it may be incomplete and insensitive because it may not be able to satisfy the sensitivity criterion and the information criterion (Shanks & St. John, 1994). Therefore, more complete and sensitive measurements of the content of the underlying representation of (in)transitivity are needed in future studies to address this issue.

## Conclusion

To misquote a song from the 1930s, "Tain't what you know it's the way that you know it, that's what gets results." In the present study, all participants who entered the analyses on triads-matching "knew" the target system, in the sense of having awareness of it at the level of understanding and being able to apply such knowledge to achieve equivalently near-perfect GJT accuracy. However, by another index of "knowing" the system-(in)transitivity biases during triads-matching—outcomes varied with instructional conditions (i.e., the way of knowing). Hence, our study adds to the range of factors that might influence language–cognition interactions, and to the range of learning products to be compared between instructional conditions.

## References

Ameel, E., Storms, G., Malt, B. C., & Sloman, S. A. (2005). How bilinguals solve the naming problem. *Journal of Memory and Language*, *53*, 60–80.

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407.

Athanasopoulos, P. (2006). Effects of the grammatical representation of number on cognition in bilinguals. *Bilingualism: Language and Cognition*, *9*(1), 89–96.

Athanasopoulos, P., Burnand, J., Damjanovic, L., & Bylund, E. (2015). Learning to think in a second language: effects of proficiency and length of exposure in English learners of German. *The Modern Language Journal*, *99*(S), 138–153.

Athanasopoulos, P., & Bylund, E. (2023). Cognitive restructuring: Psychophysical measurement of time perception in bilinguals. *Bilingualism: Language and Cognition*, 26(4), 809-818.

Athanasopoulos, P., Bylund, E., Montero-Melis, G., Damjanovic, L., Schartner, A., Kibbe, A., Riches, N., & Thierry, G. (2015). Two languages, two minds: Flexible cognitive processing driven by language of operation. *Psychological Science*, *26*(4), 518–526.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

https://doi.org/10.1017/S0272263124000433 Published online by Cambridge University Press

Bowers, J. (2002). Transitivity. *Linguistic Inquiry*, *33*(2), 183–224.
Bylund, E., & Athanasopoulos, P. (2014a). Linguistic relativity in SLA: Towards a new research programme. *Language Learning*, *64*(4), 952–985.
Bylund, E., & Athanasopoulos, P. (2014b). Language and thought in a multilingual context: The case of isiXhosa. *Bilingualism: Language and Cognition*, *17*, 431–443.
Bylund, E., & Athanasopoulos, P. (2015). Televised Whorf: Cognitive restructuring in advanced foreign language learners as a function of audiovisual media exposure. *The Modern Language Journal*, *99*(S), 123–137.
Bylund, E., Athanasopoulos, P., & Oostendorp, M. (2013). Motion event cognition and grammatical aspect: Evidence from Afrikaans. *Linguistics*, *51*(5), 929–955.
Calderón, A. M. (2013). The effects of L2 learner proficiency on depth of processing, levels of awareness, and intake. In J. M. Bergsleithner, S. N. Frota, & J. K. Yoshioka (Eds.), *Noticing and second language acquisition: Studies in honor in Richard Schmidt* (pp. 103–121). Honolulu, HI: University of Hawai'i, National Foreign Language Resource Center.
Casasanto, D. (2008). Who's afraid of the big bad Whorf? Crosslinguistic differences in temporal language and thought. *Language Learning*, *58*, 63–79.
Cerezo, L., Caras, A., & Leow, R. P. (2016). Effectiveness of guided induction versus deductive instruction on the development of complex Spanish "gustar" structures: An analysis of learning outcomes and processes. *Studies in Second Language Acquisition*, *38*, 265–291.
Cheung, H. S. (1973). A comparative study in Chinese grammars: The ba-construction. *Journal of Chinese Linguistics*, *1*(3), 343–382.
Cleeremans, A. (2008). Consciousness: The radical plasticity thesis. In R. Banerjee & B. K. Chakrabarti (Eds.), *Progress in brain research* (pp. 19–33). Elsevier.
de Graaff, R. (1997). The eXperanto eXperiment: Effects of explicit instruction on L2 acquisition. *Studies in Second Language Acquisition*, *19*, 249–76.
DeKeyser, R. M. (1995). Learning second language grammar rules: An experiment with a miniature linguistic system. *Studies in Second Language Acquisition*, *17*, 379–410.
Dolscheid, S., Shayan, S., Majid, A., & Casasanto, D. (2013). The thickness of musical pitch: Psychophysical evidence for linguistic relativity. *Psychological Science*, *24*(5), 613–621.
Ellis, N. (1993). Rules and instances in foreign language learning: Interactions of explicit and implicit knowledge. *European Journal of Cognitive Psychology*, *5*, 289–318.
Everett, C. (2013). *Linguistic relativity: Evidence across languages and cognitive domains*. Göttingen: De Gruyter.
Godfroid, A., & Schmidtke, J. (2013). What do eye movements tell us about awareness? A traingulation of eye-movement data, verbal reports and vocabulary learning scores. In J. M. Bergsleithner, S. N. Frota, & J. K. Yoshioka (Eds.), *Noticing and second language acquisition: Studies in honor in Richard Schmidt* (pp. 183–205). Honolulu, HI: University of Hawai'i, National Foreign Language Resource Center.
Godfroid, A., Boers, F., & Housen, A. (2013). An eye for words: Gauging the role of attention in L2 vocabulary acquisition by means of eye tracking. *Studies in Second Language Acquisition*, *35*, 483–517.
Hahn, U. (2014). Similarity. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*, 271–280.
Hopper, P. J., & Thompson, S. A. (1980). Transitivity in grammar and discourse. *Language*, *56*(2), 251–299.
Hsieh, H-C., Moreno, N., & Leow, R. P. (2016). Awareness, type of medium, and L2 development: Revisiting Hsieh (*2008*). In R. P. Leow, L. Cerezo, & M. Baralt (Eds.), *A psycholinguistic approach to technology and language learning* (pp. 131–150). De Gruyter Mouton.
JASP Team. (2022). JASP (Version 0.16.3) [Computer software].
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*, 638–653.
Leow, R. P. (1998). The effects of amount and type of exposure on adult learners' L2 development in SLA. *The Modern Language Journal*, *82*, 49–68.
Leow, R. P. (2015). *Explicit learning in the L2 classroom: A student-centered approach*. Routledge.
Leow, R. P., Cerezo, L., Caras, A., & Cruz, G. (2019). CALL in ISLA: Promoting depth of processing of complex L2 Spanish "Para/Por" prepositions. In R. DeKeyser & G. Prieto Botana (Eds.), *SLA research with implications for the classroom: Reconciling methodological demands and pedagogical applicability* (pp. 155–78). Amsterdam, Netherlands: John Benjamins.
Levshina, N. (2015). *How to do linguistics with R: Data exploration and statistical analysis*. John Benjamins.

https://doi.org/10.1017/S0272263124000433 Published online by Cambridge University Press

Li, F. (2019). Explicit instruction, prior knowledge, depth of processing, and grammatical knowledge development of Advanced EFL learners: The case of the English subjunctive mood. In R. P. Leow (Ed.), *The Routledge handbook of second language research in classroom learning* (pp. 347–360). Routledge.

Li, J., He, H., Wu, B., Hou, Y., Cao, K., & A., Ruhan (2018). Behavioral and ERP study of color categorical perception in proficient and nonproficient bilinguals. *Acta Psychologica Sinica*, *50*(11), 1259–1268.

Lichtman, K. (2021). What about fluency? Implicit vs. explicit training affects artificial mini-language production. *Applied Linguistics*, *42*(4), 668–691.

Lupyan, G. (2012). Linguistically modulated perception and cognition: The label-feedback hypothesis. *Frontiers in Psychology*, *3*, 1–13.

Martin, A., Niu, M., & Leow, R. P. (2019). Processing instruction, guided induction, and L2 development. In R. P. Leow (Ed.), *The Routledge handbook of second language research in classroom learning* (pp. 375–390). Routledge.

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*, 636–665.

Morgan-Short, K., Finger, I., Grey, S., & Ullman, M. T. (2012). Second language processing shows increased native-like neural responses after months of no exposure. *PLoS ONE*, *7*(3), e32974.

Morgan-Short, K., Sanz, C., Steinhauer, K., & Ullman, M. T. (2010). Second language acquisition of gender agreement in explicit and implicit training conditions: An event-related potential study. *Language Learning*, *60*(1), 154–193.

Morgan-Short, K., Steinhauer, K., Sanz, C. & Ullman, M. (2012). Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *Journal of Cognitive Neuroscience*, *24*, 933–47.

Park, H. I. (2020). How do Korean–English bilinguals speak and think about motion events? Evidence from verbal and non-verbal tasks. *Bilingualism: Language and Cognition*, *23*, 483–499.

Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language*, *39*, 633–651.

PRC Ministry of Education. (2013). *Curriculum standard for compulsory education.* http://www.moe.gov.cn/srcsite/A26/s8001/200303/W020200401347863199102.pdf

R Core Team. (2022). R: A language and environment for statistical computing (Version 4.2.1) [Computer software]. http://www.R-project.org

Robinson, P. (1996). Learning simple and complex second language rules under implicit, incidental, rule-search and instructed conditions. *Studies in Second Language Acquisition*, *18*, 27–67.

Rosa, E., & O'Neill, M. (1999). Explicitness, intake, and the issue of awareness: Another piece to the puzzle. *Studies in Second Language Acquisition*, *21*, 511–556.

Salamoura, A., & Williams, J. N. (2008). The representation of grammatical gender in the bilingual lexicon: Evidence from Greek and German. *Bilingualism: Language and Cognition*, *10*, 257–275.

Sato, S., & Athanasopoulos, P. (2018). Grammatical gender affects gender perception: Evidence for the structural-feedback hypothesis. *Cognition*, *176*, 220–231.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*, 129–158.

Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning system. *Behavioral and Brain Sciences*, *17*, 367–447.

Talmy, L. (2000). *Toward a cognitive semantics: Typology and process in concept structuring*. Cambridge, MA: MIT Press.

Thinglum, A., Serafini, E. J., & Leow, R. (2019). Exploring the relationships between lexical prior knowledge and depth of processing during the intake processing stage: An online investigation of L2 vocabulary learning. In R. P. Leow (Ed.), *The Routledge handbook of second language research in classroom learning* (pp. 375–390). Routledge.

van Gompel, R. P. G., Arai, M., & Pearson, J. (2012). The representation of mono- and intransitive structures. *Journal of Memory and Language*, *66*, 384–406.

Vanek, N. (2020). Changing event categorization in second language users through perceptual learning. *Language Learning*, *70*(2), 309–348.

Vanek, N., Sóskuthy, M., & Majid, A. (2021). Consistent verbal labels promote odor category learning. *Cognition*, *206*, 104485.

https://doi.org/10.1017/S0272263124000433 Published online by Cambridge University Press

Wigdorowitz, M., Pérez, A. I., & Tsimpli, I. M. (2023). High-level listening comprehension in advanced English as a second language: Effects of the first language and inhibitory control. *Bilingualism: Language and Cognition*, *26*(5), 865–879.

Xue, Y., & Williams, J. (2024). Inducing shifts in attentional and preattentive visual processing through brief training on novel grammatical morphemes: an event-related potential study. *Language Learning*, *74*(S1), 185–223.

Zhuang, J. (2019). Computer-assisted guided induction and deductive instruction on the development of complex Chinese ba structures: Extending Cerezo et al. (2016). In R. P. Leow (Ed.), *The Routledge handbook of second language research in classroom learning* (pp. 375–390). Routledge.

**Cite this article:** Xue, Y., & Williams, J. N. (2025). Modulating motion event categorization through brief training: *Meaning-focused versus form-focused instructional conditions*. *Studies in Second Language Acquisition,* 47: 53–79. https://doi.org/10.1017/S0272263124000433

https://doi.org/10.1017/S0272263124000433 Published online by Cambridge University Press