Manner, result, and intention: implications for event typology from a cognitive account of verb semantics based on fulfilment types

Xinyan Kou [ID] and Jill Hohenstein [ID]

School of Education, Communication & Society, King's College London, London, UK
Corresponding author: Xinyan Kou; Email: xinyan.1.kou@kcl.ac.uk

Abstract

Verb semantics has been widely approached as a dichotomy of manner and result. However, from a cognitive perspective, manner and result are often linked by intention, as captured by the 'fulfilment type' property formulated in the Realisation event domain in Talmy's event integration theory. The four 'fulfilment types' (intrinsic-, moot-, implied-, and attained-fulfilment) indicate different degrees of result certainty in verbs. This study investigates whether manner/result complementarity is cognitively less dichotomous and more nuanced, as the four fulfilment types in verbs could indicate more than two mental representations of verbs. Through two psycholinguistic experiments, we examine whether fulfilment types influence the cognitive salience of manner and result in novel verb meaning interpretation (Experiment 1) and the semantic relatedness between English verbs with different fulfilment types (Experiment 2). Our results demonstrate that manner and result in the mental lexicon act less like a dichotomy but more like a cline. This blur between manner and result verb statuses has consequences for a language's typological stance in the Realisation domain and implications for how Talmyan event research should be extended beyond well-studied Motion.

Keywords: cognitive semantics; event integration; fulfilment types; intention; manner/result complementarity; Realisation events; Talmy

1. Introduction: Realisation events, fulfilment types, and manner/result complementarity

According to Talmy's (2000) event integration theory, the human mind often packs events into 'macro-events' by condensing multiple simplex events into the same clause. In the Motion domain, for example, 'I ran into the room' is a combination of the manner of motion (running) and the path traversed by the agent (into). In this macro-event, entering is the 'framing event' outlining the conceptual frame for how

this event unfolds, and running is the 'co-event' fleshing out the frame with extra information.

While myriad studies have been devoted to Motion, little academic attention has been paid to other types of macro-events (Li, 2013). For instance, *I hunted the fugitive down* is a combination between an intention (to capture) and its realisation (successful capture). This type of macro-event is a **Realisation event** (Talmy, 2000, pp. 261–278), with intention being the co-event and its realisation being the framing event (Kou & Hohenstein, 2022).

Within the Realisation event framework, verbs differ in their 'fulfilment type' (Talmy, 2000, pp. 262–268), a property indicating how likely an intended goal denoted by the verb is realised (Kou & Hohenstein, 2022). This property divides verbs into four types: 1) **intrinsic-fulfilment** verbs do not indicate specific intended results without context (e.g., *kick*); 2) **moot-fulfilment** verbs independently encode intended results but leave moot whether they are realised (e.g., *hunt*); 3) **implied-fulfilment** verbs encode intended results and imply their realisation (e.g., *wash*); 4) **attained-fulfilment** verbs encode intended results and entail their realisation (e.g., *kill*). Figure 1 is a schematic illustration of fulfilment types arranged in terms of their certainty of intention realisation. A question mark next to 'intrinsic-fulfilment' shows that these verbs encode no specific intentions.

The concept of fulfilment types boils down to three keywords: **manner, result**,[1] and **intention**.[2] The first two elements constitute a widely-attested semantic rule for verbs: **manner/result complementarity** (Rappaport Hovav & Levin, 2010), namely, a verb (root) can only lexicalise either manner or result, not both. While manner and result are fundamental to verb meaning, intention could link manner and result and thus blur the manner/result dichotomy in the mind. It is argued that semantic manner/result verb statuses are mainly 'extensional or referential definitions that characterize a verb in terms of the set of events it can label […] and make no specific claims about […] the *mental representations*[3] of these verbs' meanings' (Behrend, 1990, p. 685, note 1, our emphasis). Thus, in this study, we take a cognitive approach to verb meaning from the perspective of fulfilment types and investigate whether verbs mentally fall into dichotomous manner/result verb mental representations.

In Section 2, we briefly review relevant literature on manner, result, and intention, discuss how they interact in verb meaning, and specify the research question. Next, we empirically examine how verbs of different fulfilment types cluster and vary in
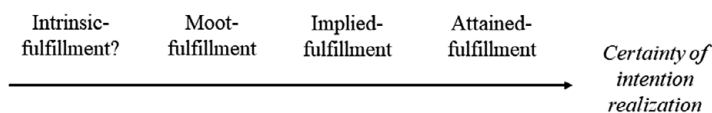

Figure 1. Schematic cline of certainty of intention realisation in verbs with different fulfilment types

---

[1]Although 'result' is generally received to be fundamental to event conceptualisation (cf. Zacks & Tversky, 2001; see also Gerwien & von Stutterheim, 2018), there is no consensus as to what an event result is (Ji & Papafragou, 2020; Santin et al., 2021). In this study, 'result' refers to the outcome of any intended caused change.

[2]The term 'intention' refers to the change (of state, of location, etc.) inherent to the relevant verb's denotation; the mere completion of an action is not considered as intention.

[3]The 'mental/cognitive representation' of verbs refers to how verb meaning is represented in the mental lexicon; it concerns not only semantics but also world knowledge and conventional language use, etc.

different task situations (Section 3), before discussing the blur between manner/result verb representations in the mind and its implications on Talmyan event typology in the Realisation domain (Section 4).

2. Manner, result, intention, and their interplay in verb meaning

Although manner and result are argued to be mutually exclusive in verb semantics, they are often intertwined in cognition. Manner verbs (e.g., *hit*) often denote intended results and sometimes imply their realisation (Levin & Rappaport Hovav, 2014, p. 339; Talmy, 2000, pp. 262–267), and result verbs can often be associated with certain manners (e.g., *clean* is often achieved by wiping/scrubbing/washing/etc., Levin & Rappaport Hovav, 2014). The conceptual link between manner and result might lead to the co-existence of manner and result in a verb's denotation in the mind, though presumably with different cognitive salience. Of note, we are not arguing for *semantic co-existence* of manner and result in a verb; rather, our discussion pertains to the *cognitive representation* of these elements.

Salience of event components is well-discussed in Motion events, focussing particularly on manner and path. Cognitively more salient components tend to be mentioned more frequently in speech (e.g., Gennari et al., 2002; Hohenstein et al., 2006; Ji et al., 2011), allocated more visual attention (e.g., Ji & Hohenstein, 2017; Soroli & Hickmann, 2011), processed faster (e.g., Ünal et al., 2021), considered dominant in similarity judgments (e.g., Hickmann et al., 2017; Pourcel, 2003) and novel verb interpretation (e.g., Hohenstein et al., 2004), and remembered better (e.g., Billman et al., 2000; Filipović, 2011, 2022; Flecken et al., 2014; but see Engemann et al., 2015 and Papafragou et al., 2002 for opposite findings). A general bias towards path over manner is often reported (e.g., see Santin et al., 2021 and Ünal & Papafragou, 2019 for reviews). Of note, the relative salience of manner and path in these studies has been generally discussed as *a property of the language* (i.e., if participants mention manner more in language A than in language B, then manner is more salient in language A than in language B), rather than *a property of the verb* (i.e., how salient manner and path are in a verb's meaning).

Relatively few studies have explored the salience of different event components as a property of the verb (e.g., Behrend, 1990; Forbes & Farrar, 1993). For example, in Behrend (1990), native English speakers viewed videos depicting novel scenes labelled with novel verbs and alternate videos with manner, result, agent, or instrument altered, judging whether the alternate videos could be described by the same novel verbs. It was found that adults tended to interpret novel verbs as encoding result rather than manner – or as result verbs in the view of manner/result complementarity.

However, although the studies above were targeted at verb meaning, it is difficult to pin down the factors underlying verb meaning interpretation. Were the novel verbs interpreted in such a way because that is how the verbs should be? Or could it stem from other factors? Indeed, Behrend (1995) pointed out three groups of factors influencing children's interpretation of novel verbs: child-driven factors (e.g., perceptual and cognitive abilities), language-driven factors (e.g., syntactic framing, verb inflection), and environment-driven factors (e.g., adult verbal input). Although this remark was specifically made for children's verb acquisition, it nevertheless showcases the diversity of factors that can contribute to one's interpretation of verb

meaning. Thus, although previous research does illuminate the relative salience of manner and result in verb meaning, it only offers 'snapshots' of how verbs are interpreted in specific experimental settings with specific combinations of factors. To ensure that the salience of manner and result is indeed examined as *a property of the verb*, one needs to examine how manner/result salience varies as a response to variation *generated by verbs*. Fulfilment types provide an ideal testbed for this. Given that verbs of the four fulfilment types encode intention and intention realisation to different degrees of specificity and certainty, if change in fulfilment types leads to change in manner/result salience, we can be confident that manner/result are examined as a verb property.

Apart from manner and result, the fulfilment type concept introduces intention into verb meaning representation. The idea of intention influencing verb semantics is present in Behrend and Wittek (2003) and Behrend and Scofield (2006), who showed that when intention is unknown (i.e., unfamiliar event scenes), the act of labelling an action with a (novel) verb encourages the belief that the action is intended (although in reality, the action might be a failed attempt to fulfil other goals), which in turn shapes novel verb interpretation. For example, in one scene the actor tried to lift a pretend piece of bread with a magnet on top using a spatula with a magnet on its underside; the attempt failed as the bread stayed put. When the actor's action was not labelled with a novel verb, children tended to see through the failed attempt and imitate the real intention (i.e., to lift the bread). However, when the same activity came with a novel verb (e.g., 'Look, I'm *meeking*!'), children would interpret the intention as *not* to lift the bread. Behrend and colleagues were primarily concerned with intention inference through language when intention is unclear. In the context of fulfilment types, we wish to explore how clear intention would shape the mental representation of verb meaning. The lack of literature in this regard calls for a new evaluation of the contribution of intention in the mental representation of verbs.

Though not as directly visible as manner and result, intention is present in almost all human behaviour (Pourcel, 2004; Zwaan & Radvansky, 1998) and has 'causal effects in the physical world' (Kim, 2000, p. 31). Understanding intention facilitates event apprehension by activating the knowledge of how relevant events typically develop (Zwaan & Radvansky, 1998), and the fulfilment of intention marks event boundaries (e.g., Carroll & von Stutterheim, 2010, p. 69; Gerwien & von Stutterheim, 2018). Importantly, intention can boost the perceived smoothness in motion, making fragmented actions with short pauses appear more continuous (Peng et al., 2020) and shortening the perceived temporal lapse between the causing and the caused events (Haggard et al., 2002; Moore et al., 2013). Focussing on French motion verbs, Nakamura et al. (2021) noted that when manner-of-motion verbs such as *courir* ('run') co-occur with goal/purpose information (e.g., *à sa chambre*, 'to his/her room'), they show characteristics of path verbs (e.g., no longer compatible with low-velocity adverbs such as *doucement*, 'mildly'); this also suggests that intention can bring manner(-of-motion) verbs closer to result (path) verbs. The above findings suggest that intention can shorten the cognitive distance between action and result. Would intention also draw manner and result closer in the mental lexicon?

In light of manner/result complementarity (Rappaport Hovav & Levin, 2010), intrinsic-/moot-/implied-fulfilment verbs would fall into the category of manner verbs (as they do not semantically entail intention realisation); attained-fulfilment verbs would be result verbs (as they semantically entail intention realisation). However, given the potential influence of intention-binding manner and result,

would interpretation of verbs in the Realisation domain show nuanced verb representations beyond the manner/result split?

The idea of more nuanced categories within manner/result verbs is not new (e.g., Aurnague, 2011; Sarda, 2019; Stosic, 2009). For motion verbs, manner and path have been conceptualised as sets of independent features (e.g., directionality, intentional opposition to a force, and so forth for path, Aurnague, 2011; speed, force, means, and so forth for manner, Stosic, 2009, p. 111). These features are organised in family resemblance, and a verb can denote one or more of these features (Sarda, 2019); the more relevant factors a verb contains, the more typical manner/result verb representation it has. Thus, motion verbs tend to simultaneously contain elements of both manner and path, contrary to manner/result complementarity (e.g., Rappaport Hovav & Levin, 2010). However, these studies differ from the current discussion in two aspects. First, previous research concerns verb semantics, while the present discussion concerns the cognitive representation of verbs. Second, previous studies deconstructed manner/result into independent features; motion verbs fall into subclasses because they contain different (numbers of) semantic components, so the fine-grained classification is multi-dimensional in nature. In the present study, manner and result are conceptualised as a unified cline, and variation among verbs is regarded as differences in a single dimension.

Specifically, we expect intrinsic-fulfilment verbs to be the closest to having a manner verb representation, while attained-fulfilment verbs are the closest to having a result verb representation, with moot- and implied-fulfilment verbs located increasingly further away from manner interpretation and closer to result interpretation.

3. The present study

This study contains two experiments. Experiment 1 is a novel verb learning task exploring how fulfilment types would influence the salience of manner and result in verb meaning interpretation. Experiment 2 measures the mental relatedness between intrinsic-, moot-, implied-, and attained-fulfilment verbs and their semantically corresponding result (i.e., attained-fulfilment) verbs in English, in order to examine whether verbs with the four fulfilment types differ in cognitive proximity to result verb counterparts. Together, the experiments investigate whether verbs of different degrees of fulfilment are represented in a binary (as the manner/result split) or in a more nuanced way.

Participants in both experiments were adult monolingual British English speakers (all British nationals resident in the UK) with normal or corrected-to-normal vision and without neurological or psychiatric disorders or literacy difficulties recruited via Prolific (www.prolific.co). The data in this study was collected from 2021 to Spring 2022, and thus participation was remote via the online experiment builder Gorilla (www.gorilla.sc) due to COVID-19 restrictions. Each participant received a small monetary reward upon full completion. All statistical analyses were performed on R (version 4.1.2, R Core Team, 2021) with the initial alpha set to 0.05 and Bonferroni-adjusted p values in post hoc analyses. The experiments were piloted to ensure the validity of materials, experiment design, and procedure. All data, R codes, pilot studies, stimuli, additional analyses, and extra tables are available as appendices at https://osf.io/jpaqu/?view_only= 8880dc4deac84e9988070e455e1c56aa.

3.1 Experiment 1: Novel verb learning task

Given the imbalanced cognitive salience of manner and result (Section 2), it is unlikely that these features carry equal weight in verbs with different fulfilment types. We hypothesised that the four fulfilment types are represented along a cline between manner and result verbs. We tested this hypothesis through a novel verb learning task. Specifically, novel verbs with intrinsic-, moot-, implied-, and attained-fulfilment properties should yield progressively higher result salience and lower manner salience in verb meaning interpretation.

3.1.1 Forbes and Farrar (1993) and the challenge of the present experiment
Before explaining the current experiment, we first review a particularly relevant study. Forbes and Farrar (1993) examined how manipulation of the salience of event components would influence novel verb interpretation. Focussing on manner, result ('outcome' in their terminology), agent, and instrument, they trained each novel verb with three videos that were said to exemplify the verb. The authors manipulated the frequency of the training videos displaying the relevant event components: in the *same* condition, the event components remained the same across the training videos; in the *different* condition, one event component changed in each video while all other components held constant (thereby informing participants that the changing component was unimportant for the novel verb's meaning); in the *mixed* condition, one event component remained constant while all other components varied in each video. After training for each novel verb, participants were shown a test video featuring a changed event component and judged whether it was an acceptable instance of the novel verb. Adult participants were reluctant to generalise novel verbs to videos containing manner and result changes compared to agent and instrument changes, suggesting that manner and result were considered as more salient components of verb meaning, but the salience of manner and result did not significantly differ. Importantly, the frequency of adults' extensions for each event component did not significantly vary across the *same/different/mixed* training conditions (Forbes & Farrar, 1993, p. 12), indicating stable perceptions of the salience of the event components.

Having reviewed Forbes and Farrar (1993), let us now discuss its relevance to the present study. A challenge in our experiment is to conjure up fulfilment type readings for novel verbs. It might be useful to generate fulfilment type readings in novel verbs using frequency-based methods (akin to Forbes & Farrar, 1993). For example, showing several teaching videos varying in results to conjure up an intrinsic-fulfilment verb reading; showing teaching videos with identical results to encourage an implied-fulfilment verb reading. There are two reasons why this paradigm might not suit our current purpose. First, as mentioned above, the three training conditions in Forbes and Farrar (1993) did not elicit significantly different degrees of salience associated with the event components (for adults), suggesting that frequency-based training might not effectively assign different fulfilment type readings to novel verbs. Second, even if the frequency-based strategy was effective in distinguishing verbs in different training conditions, it would be difficult to decide *what frequency of result match would generate which fulfilment type reading* – how many different-result teaching videos would be necessary to generate intrinsic-/moot-/implied-/attained-fulfilment readings? Thus, we chose to alternatively generate fulfilment type readings for novel verbs with definitions (Section 3.1.3.2).

3.1.2 Participants

A priori calculation on G*Power (version 3.1.9.4) for this experiment yielded a sample size of 100. In total, 172 participants took part in Experiment 1. Among them, 51 (29.65%) dropped out, 2 (1.16%) were timed out by Prolific, and 16 (9.30%) participants were rejected due to lack of attention. The remaining 103 participants were monolingual English speakers (mean age: 36.30 years; age range: 20–56 years; 49 females; ethnicity: 2 Asian, 2 Black, 4 mixed, 95 White; student status: 68 not student, 5 students, 30 unknown; employment status: 46 full-time, 15 part-time, 9 unemployed, 33 unknown[4]).

3.1.3 Materials

3.1.3.1 Video stimuli. This experiment used 10 sets of silent videos showing novel activities 3 to 8 s long. In each set of 4 videos, the teaching video depicted a manner and a result. One alternate video was identical to the teaching video except that the result was not successfully fulfilled ('Same-Manner-Unfulfilled-Result', or SMUR). Another alternate video also had the same manner as the teaching video but showed a different result ('Same-Manner-Different-Result', or SMDR). A third alternate video depicted the same result as the teaching video but showed a different manner ('Different-Manner-Same-Result', or DMSR). While the SMDR and DMSR videos are respectively analogous to the alternate same-manner-different-path and same-path-different-manner videos used in classic similarity judgment tasks in Motion research, the present experiment further involved SMUR videos. This setup allows a more thorough representation of event scenarios. While in SMDR videos manner obligatorily comes with some result, SMUR videos acknowledge that manner can exist without (conventionally received) results obtaining, thereby allowing a more nuanced examination of how one interprets the verb. To illustrate, in the 'Walnut' video set, the teaching video depicts a situation in which the actor cracks a walnut by stapling it with a stapler. In the SMDR video, the actor staples a walnut and the walnut slips off. In the SMUR video, the actor staples the walnut, but the walnut stays still and remains closed. In the DMSR video, the actor cracks a walnut by hitting it with a stapler.

Videos within each set had the same duration. To ensure novelty, the objects involved in the teaching videos were not used in ways that they would typically be used to reduce 'object affordances'[5] (i.e., common uses for objects, Huang et al., 2002). The teaching video of each video set was rated in terms of familiarity with a different group of 19 native British English speakers recruited through institution platforms. Participants were asked to rate how familiar the activity in each video looked to them on a Likert scale from 1 (very uncommon) to 5 (very common). All teaching videos were rated between 1.6 and 1.9, suggesting that the scenes were fairly

---

[4]Here as in Experiment 2 (Section 3.2.1), we did not set up sampling criteria for ethnicity, student status, or employment status – information in these aspects was collected by the recruitment platform by default. As all participants are monolingual English speakers resident in the UK, we assume homogeneity across different ethnicities, student groups, and employment groups for the present study. As the sampling technique was not intended to be representative of all ethnic/student/employment statuses, the small numbers of participants in some groups make it difficult to conduct meaningful statistics with these demographic factors as independent variables. Thus, demographic factors are not included in the following analyses.

[5]To illustrate, in the 'Walnut' scene, a stapler is used to staple a walnut to crack it (rather than for binding things).

novel (video descriptions and ratings available in Appendix 1). The videos were designed in such a way that the actions were highlighted, showing only the actor's hands, to preclude potential stereotypes about the actor's appearance, gender, ethnicity, and so forth.

3.1.3.2 Novel verbs and their definitions. Each set of videos was paired with a novel verb, making a total of 10 novel verbs. The novel verbs were mono-syllabic 3–5 letter long words borrowed from previous studies involving English speakers to ensure that they satisfy phonotactic restrictions of English. The novel verbs were: *dotch* (Chan et al., 2011), *krad*, *zub* (Hohenstein, 2005), *smick* (Buccola et al., 2018), *jav*, *glorp* (Imai et al., 2008), *bock*, *stife* (Behrend, 1990), *spog* (Kersten et al., 2010), and *hirsh* (Maguire et al., 2010).

The novel verbs in this experiment were defined to take on different fulfilment types. Having shown that a frequency-based approach might not effectively serve this purpose (Section 3.1.1), we present here a semantically-based approach. In the intrinsic-fulfilment condition, novel verbs were given manner-only definitions due to their lack of goal information. For example, in one scene, someone crushed a walnut placed in a stapler by stapling it using their right hand; the novel verb for this scene, *hirsh*, was defined as 'staple using right hand' in the intrinsic-fulfilment condition. In the moot-fulfilment condition, the uncertainty in intention realisation was highlighted through the *try [result] by [manner]* construction, and *hirsh* here was defined as 'try crushing walnut by stapling using right hand'. In the implied-fulfilment condition, the infinitive was used to create the implication of intention realisation and to downplay the sense of uncertainty with the formula of *[manner] in order to [result]*, and *hirsh* was defined as 'staple using the right hand to crush walnut'. The attained-fulfilment condition had result-only definitions, as attained-fulfilment verbs semantically entail result and no manner; *hirsh* was thus defined as 'crush walnut'. Apart from these defined conditions, a no-definition condition served as control.

Considering that the novel verbs were defined, a possible criticism is that the acceptability judgments would simply reflect participants following these definitions, rather than participants' own biases for manner and result. However, let us clarify that Experiment 1 is not a conventional novel verb learning task – this paradigm is only adopted to legitimately present the definitions. Our aim is *not* to explore the tendency to attend to manner or result – this can only happen when participants *freely* interpret the novel verbs. *When a verb has a specific fulfilment type, its meaning is no longer subject to free interpretation in terms of manner and result.* Rather, we are only concerned with manner and result salience under the influence of the definitions. Thus, we would need participants to follow the definitions instead of showing their own manner/result biases.

During the display of teaching videos, the novel verbs were introduced in written form in the transitive and the progressive (e.g., *Someone was hirshing the walnut*), each with a definition underneath ('[novel verb] = [definition]'), except for the no-definition condition. Motion research shows that syntactic framing could influence novel verb interpretation (e.g., Naigles & Terrazas, 1998). The perfective and the transitive tend to trigger path verb (and possibly result verb) interpretations, while the progressive and the intransitive encourage manner-of-motion verb (and potentially manner verb) readings. For the present experiment, however, syntactic framing

might not be much of a concern, because the novel verbs were explicitly defined apart from the no-definition condition. Nevertheless, effort was made to neutralise the impact of syntax. As Realisation events depict agentive events (Jia & Li, 2015), the transitive frame is necessary. To downplay the influence of the transitive frame, the progressive aspect was used, as these two frames have opposing influences on meaning interpretation and might thus offset one another.

3.1.4 Procedure

Participants were randomly assigned to one of the five fulfilment type conditions. In the novel verb learning phase, participants viewed 10 sets of novel videos, randomised for order. For each set, the teaching video was shown first with a sentence introducing the novel verb, with or without a definition depending on the condition. Next, participants viewed three alternate videos one at a time randomised for order. Each video was played manually by clicking on the video, and participants could view each video up to two times by clicking on it again after the first display (clicks during the video display were invalid). Following each alternate video, participants were asked the following questions: 1) Is this [novel verb]-ing? (e.g., *Is this hirshing?*) 2) Does the actor have the same manner as in the teaching video? 3) Does the actor have the same result as in the teaching video? Answers were recorded by Yes/No buttons under each question. Prior to the task, there was a practice trial to familiarise participants with the procedure. The entire task lasted 10 to 15 min.

As the classic novel verb learning task often only involves acceptability questions, it is necessary to clarify why manner and result judgments were added. Manner and result judgments were conducted to ensure that participants perceived manner and result in the way the videos were configured (e.g., in DMSR videos, participants should see different manners and identical results to the teaching videos). Our data suggested that this was not always the case, so we only analysed the observations containing the expected manner and result judgments to avoid contamination of data.

This setup not only examines whether novel verbs in each fulfilment type condition are interpreted as manner/result verbs, but also reveals whether the verb meaning interpretations across conditions fall into two or more groups. If only two clusters emerge, the manner/result complementarity would be cognitively confirmed; if more groups emerge, manner/result complementarity could be questioned as the only representation of verbs in the mental lexicon.

3.1.5 Data analysis

For each of the three judgment tasks, 3090 observations were collected. These included 9 (0.29%)[6] blank data points where no responses were given (8 acceptability judgments, 1 result judgment), 2 (0.06%) acceptability judgments made before the relevant teaching videos were displayed, and 3 (0.10%) impossibly fast responses made within 200 ms (1 acceptability judgment, 1 manner judgment, 1 result judgment). To maintain completeness of observation, the problematic data were removed along with the corresponding data in the other judgments. The remaining 3076 (99.55%) observations were then sifted according to whether they

---

[6]Here and below, the percents are of the total number of responses in the raw data.

yielded the expected interpretation of manner and result in the videos.[7] This left 2522 (81.62%) acceptability judgments. Positive and negative acceptability judgments were coded as '1' and '0', respectively.

To examine how novel verbs in different fulfilment type conditions vary in terms of manner and result salience, we performed a binomial mixed-effect logistic regression (glmerlme4 package, Bates et al., 2015), which predicted acceptability judgments (yes/no) based on fulfilment type condition (intrinsic/moot/implied/attained/no-definition) and video type (SMUR/SMDR/DMSR). A random intercept of participants was added to minimise non-independence issues (e.g., Winter & Grice, 2021) and enhance the generalisability of findings (e.g., Roettger, 2021; Yarkoni, 2020). Post hoc estimated marginal squares analysis (emmeansemmeans package, Lenth, 2022) was conducted with Bonferroni-adjusted p values and odds ratios (ORs)[8] for effect sizes.

3.1.6 Results

Table 1 presents the raw counts and percentages of positive acceptability judgments. It is clear that the intrinsic-/moot-/implied-fulfilment and no-definition conditions all elicited over 60% acceptability on the manner-matching SMUR and SMDR videos and lower than 35% acceptability on the result-matching DMSR videos, suggesting that the novel verbs in these conditions were interpreted overall as manner verbs in the traditional sense. For novel verbs in the attained-fulfilment condition, acceptability was low for SMUR (8.74%) and SMDR (12.40%) videos but high for DMSR videos (99.40%), indicating a stronger result verb interpretation.

The best-fit model with the lowest Akaike Information Criterion (AIC) returned both fulfilment type condition and video type as significant fixed effects and an interaction between the two (Table 2).

Table 1. Raw counts and percentages of positive acceptability judgments

Fulfilment type condition	SMUR	SMDR	DMSR
Intrinsic	147/152 (96.71%)	120/129 (93.02%)	12/165 (7.27%)
Moot	173/195 (88.72%)	109/149 (73.15%)	18/176 (10.23%)
Implied	126/188 (67.02%)	89/143 (62.24%)	16/181 (8.84%)
Attained	16/183 (8.74%)	21/149 (14.09%)	183/184 (99.46%)
No definition	146/190 (76.84%)	105/153 (68.63%)	63/185 (34.05%)

Abbreviations: SMUR, same-manner-unfulfilled-result; SMDR, same-manner-different-result; DMSR, different-manner-same-result.

---

[7]Manner and result judgments sometimes deviated from the expected manner/result interpretations (e.g., thinking that a DMSR video has the same manner as the target video, thinking that an SMDR video has the same result as the teaching video). This is because manner and result in our videos were small-scale (compared to manner/path differences in motion videos), so differences in them can be subtle, leading to unexpected judgments. The presence of such unexpected judgments does not indicate that participants did not follow the task; their removal is to control data quality, focussing solely on those judgements where the intended manner/result was recognised by participants. Since the removed data points came from different participants, this resulted in different numbers of responses for each participant in the dataset.

[8]As an effect size measure, an OR of 1 means that probabilities of something happening and not happening are the same, hence null effect size. An OR can take any value between 0 and infinity; the numerically further from 1 the OR, the larger the effect size.

Table 2. Generalised linear mixed model fit by maximum likelihood for acceptability

R code: glmer(Acceptability ~ FulfilmentTypeCondition * VideoType + (1 | Participant),
data = Experiment1_2522obs, family = "binomial",
control = glmerControl(optimizer = "bobyqa"))

AIC	BIC	LogLik	Deviance	Df.Resid		
1903.4	1996.7	−935.7	1871.4	2506		
Scaled residuals						
Min	1Q	Median	3Q	Max		
−21.3186	−0.3001	0.0583	0.4156	7.7373		
Random effects						
Groups		Name	Variance	SD		
Participant		(Intercept)	1.241	1.114		
No. of obs. 2522		Groups:	Participant, 103			
Fixed effects						
		Estimate	SE	*z*-value	Pr(>|z|)	Signif.
(Intercept)		5.71	1.20	4.78	<0.001	***
Intrinsic		−8.74	1.27	−6.87	<0.001	***
Moot		−8.22	1.25	−6.56	<0.001	***
Implied		−8.51	1.26	−6.78	<0.001	***
No definition		−6.54	1.23	−5.32	<0.001	***
SMUR		−8.61	1.22	−7.07	<0.001	***
SMDR		−7.98	1.21	−6.60	<0.001	***
Intrinsic: SMUR		15.63	1.38	11.29	<0.001	***
Moot: SMUR		13.48	1.28	10.54	<0.001	***
Implied: SMUR		12.28	1.27	9.70	<0.001	***
No definition: SMUR		10.79	1.25	8.67	<0.001	***
Intrinsic: SMDR		14.11	1.33	10.59	<0.001	***
Moot: SMDR		11.73	1.26	9.30	<0.001	***
Implied: SMDR		11.32	1.26	8.99	<0.001	***
No definition: SMDR		9.67	1.24	7.82	<0.001	***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '						

Abbreviations: AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion; SD, standard deviation; SE, standard error; SMUR, same-manner-unfulfilled-result; SMDR, same-manner-different-result; DMSR, different-manner-same-result.

Post hoc analysis (Table 3) was conducted comparing different fulfilment type conditions by video type to explore the number of groupings that would appear; more than two groups would call into question the accuracy of manner/result complementarity in the cognitive interpretation of verbs.

In manner-matching SMUR videos, the attained-fulfilment condition yielded the lowest acceptability. Further, acceptability for the implied-fulfilment condition was significantly lower than that for the intrinsic- and moot-fulfilment conditions.

For manner-matching SMDR videos, the attained-fulfilment condition again yielded the lowest acceptability. The intrinsic-fulfilment condition led to significantly higher acceptability than the moot-fulfilment, implied-fulfilment, and no-definition conditions.

Turning to result-matching DMSR videos, the attained-fulfilment condition this time elicited the highest acceptability. The no-definition condition took the second highest place. The intrinsic-, moot-, and implied-fulfilment conditions clustered together at the lower end.

In summary, novel verbs assigned intrinsic-/moot-/implied-fulfilment properties yielded higher acceptability in manner-matching videos (i.e., mentally represented as manner verbs), while novel verbs in the attained-fulfilment condition yielded higher acceptability in result-matching videos (i.e., mentally represented as result verbs).

Table 3. Post hoc analysis for acceptability judgments

Pairwise comparison	Z score	Bonferroni-adjusted p	OR (95% CI)
SMUR videos			
Attained − Intrinsic	−9.70	<0.001	0.00 (0.00, 0.01)
Attained − Moot	−9.94	<0.001	0.01 (0.00, 0.02)
Attained − Implied	−7.49	<0.001	0.02 (0.01, 0.09)
Attained − No definition	−8.45	<0.001	0.01 (0.00, 0.06)
Implied − Intrinsic	−4.69	<0.001	0.04 (0.01, 0.27)
Intrinsic − No definition	3.99	0.001	14.05 (2.31, 85.52)
Implied − Moot	−3.19	0.014	0.27 (0.06, 0.81)
Intrinsic − Moot	2.39	0.167 n.s.	n/a
Moot − No definition	2.19	0.289 n.s.	n/a
Implied − No definition	−1.09	1.000 n.s.	n/a
SMDR videos			
Attained − Intrinsic	−8.82	<0.001	0.00 (0.00, 0.02)
Attained − Moot	−7.16	<0.001	0.03 (0.01, 0.11)
Attained − Implied	−5.71	<0.001	0.06 (0.02, 0.23)
Attained − No definition	−6.44	<0.001	0.04 (0.01, 0.17)
Implied − Intrinsic	−4.46	<0.001	0.08 (0.02, 0.37)
Intrinsic − Moot	3.24	0.012	6.45 (1.34, 30.94)
Intrinsic − No definition	3.95	0.001	9.52 (2.01, 45.07)
Moot − No definition	0.88	1.000 n.s.	n/a
Implied − Moot	−1.57	1.000 n.s.	n/a
Implied − No definition	−0.71	1.000 n.s.	n/a
DMSR videos			
Attained − Intrinsic	6.78	<0.001	6242.66 (194.65, 200211.85)
Attained − Moot	6.56	<0.001	3715.91 (121.87, 113299.42)
Attained − Implied	6.78	<0.001	4959.28 (161.75, 152050.17)
Attained − No definition	5.32	<0.001	694.59 (24.19, 19940.59)
Intrinsic − No definition	−4.21	<0.001	0.11 (0.03, 0.46)
Moot − No definition	−3.54	0.004	0.19 (0.05, 0.68)
Implied − No definition	−4.10	<0.001	0.14 (0.04, 0.52)
Intrinsic − Moot	−0.91	1.000 n.s.	n/a
Implied − Intrinsic	0.40	1.000 n.s.	n/a
Implied − Moot	−0.54	1.000 n.s.	n/a

Abbreviations: SMUR, same-manner-unfulfilled-result; SMDR, same-manner-different-result; DMSR, different-manner-same-result; OR, odds ratio; 95% CI, 95% confidence interval.

The clustering patterns across the fulfilment type conditions show that they often fell into more than two groups, confirming the prediction that verbs (in the Realisation domain) cannot be neatly subsumed under the manner/result dichotomy in the mental lexicon. What deserves attention is that participants in the non-defined condition behaved more similarly to participants in the intrinsic-, moot-, and/or implied-fulfilment conditions than to those in the attained-fulfilment condition (see Table 1), suggesting that these verbs are naturally interpreted as having manner verb representation, but not so pure as that of the intrinsic-fulfilment condition. Finally, the fact that manner-matching SMUR and SMDR videos yielded more groupings than result-matching DMSR videos suggests that the variation among different fulfilment type conditions was driven by novel verbs with manner verb status.

3.1.7 Discussion
Experiment 1 offered two findings. First, although verbs in the four defined fulfilment type conditions could be categorised into two general groups (i.e., intrinsic-/moot-/implied-fulfilment vs. attained-fulfilment), echoing manner/result complementarity

(Rappaport Hovav & Levin, 2010), within-group variation was observed for manner verbs (i.e., novel verbs whose meaning interpretation hinged more on manner sameness than on result sameness). Second, non-defined verbs were generally interpreted as manner verbs, in particular resembling verbs in the moot-/implied-fulfilment conditions.

For the first point, the intrinsic-/moot-/implied-fulfilment conditions yielded higher acceptability in manner-matching videos (indicating manner verb representation), and the attained-fulfilment condition elicited higher acceptability in result-matching videos (indicating result verb representation). This is hardly surprising given how the novel verbs were defined. What deserves more attention is the variation among the intrinsic-/moot-/implied-fulfilment conditions, which display increasingly less dependence on manner sameness and more dependence on result sameness in acceptability judgments. This could be due to the different degrees of specificity in the intention and implications of intention realisation encoded in these verbs, showcasing how intention bonds manner and result and blurs verbs's categorical boundaries.

Having mapped out the clustering pattern of the four defined conditions, let us move on to the no-definition control. A closer look at Table 1 shows that participants in the non-defined condition tended to base acceptability primarily on manner matches, but they rated these verbs somewhere between verbs in the attained-fulfilment condition and the other three conditions. The non-defined verbs were mainly interpreted as manner verbs deviated from the well-observed tendency to interpret novel verbs as path verbs rather than manner verbs in previous Motion studies (e.g., Behrend, 1990; Ünal et al., 2021). One possible explanation is that the activities in our video stimuli were all facilitated by instruments, which has been argued to highlight manner salience (Ji & Hohenstein, 2017; Pourcel, 2004). A second account concerns the 'efficiency' of verb meaning. According to Behrend (1990, p. 688), compared to manner and result verbs, instrument verbs (e.g., *hammer*, *saw*) are more 'efficient labels' for event description, as they often combine instrumental information with manner (e.g., *hammer* conjures up a pounding manner) or result (e.g., *saw* implies that something is cut). Similarly, moot-/implied-fulfilment verbs might be more efficient labels than intrinsic-/attained-fulfilment verbs, as the former cognitively denote both manner and result, thereby maximising information. The pursuit of efficiency might have also underlain the tendency to interpret the non-defined verbs as encoding both manner and result. A third reason might be the influence of syntactic framing on meaning interpretation (e.g., Gillette et al., 1999; Naigles & Terrazas, 1998). Recall that the gerundive intransitive frame was used for the current acceptability judgment prompt (i.e., 'Is this v.-ing?'), which might have encouraged a manner verb reading. Although the novel verbs were introduced in the relatively neutral syntactic frame of 'Someone is v.-ing something' in the definitions, the prompt might have been more recent to – and thus had a greater impact on – acceptability judgments. Notwithstanding, syntax might not impact verb meaning interpretation in Realisation as strongly as in Motion research. The syntactic influence on meaning interpretation of Motion verbs partly stems from the fact that manner-of-motion verbs can seldom be used transitively with locations (e.g., *I ran the room) but path verbs can (e.g., *I entered the room*). Conversely, both manner and result verbs in Realisation can be used transitively with affected objects (e.g., *I washed/cleaned the shirt*), so compatibility with the transitive frame no longer distinguishes between these two verb categories.

Method-wise, future research can trial the frequency-based training method mentioned in Section 3.1.1 and consider the rates of result match in training videos needed to generate a specific fulfilment type reading in novel verb meaning interpretation.

Summarising, Experiment 1 attested more than two patterns of manner and result salience in the interpretation of novel verbs assigned the four fulfilment types, suggesting that these verbs do not neatly conform to manner/result complementarity in cognition. It is worth noting that the effect sizes in Experiment 1 tend to be very large. This might be because the definitions served as strict distinctions for novel verbs in different fulfilment type conditions. In verbs encountered in real language, the distinctions might not be as clear-cut, which makes it important to examine genuine verbs with different fulfilment types. This was done in Experiment 2.

3.2 Experiment 2: Mental relatedness task

Having shown that novel verbs of different fulfilment types correspond to distinct verb representations, we explored whether the same would apply to genuine English verbs. The novel verb learning paradigm was understandably no longer suitable, so we examined the clustering patterns of verbs by measuring the semantic relatedness between the verbs through a recall task in the paired-associate learning paradigm. The rationale is that semantically related words have a recall advantage compared to unrelated words, that is, the 'semantic relatedness effect' (e.g., Kowialiewski et al., 2022 and references therein).

The next step is to decide which verbs' semantic relatedness to measure. Often, manner verbs have semantically compatible result verbs that can describe the same scenes (cf. Behrend, 1990). For example, I hunted (manner verb) vs. arrested (result verb) the fugitive and I washed (manner verb) vs. cleaned (result verb) the shirt can respectively describe the same scenarios and only differ in terms of certainty of intention realisation. Given the different extents to which intrinsic-/moot-/implied-fulfilment verbs theoretically encode intention and imply its realisation, these verbs would arguably show varying cognitive distances with semantically compatible attained-fulfilment verbs. To illustrate, the distance between hunt (moot-fulfilment) and arrest (attained-fulfilment) would be greater than that between wash (implied-fulfilment) and clean (attained-fulfilment), because hunt encodes less certainty of intention realisation than wash does.

Revisiting the semantic relatedness effect, if a manner verb and its semantically compatible result verb can be recalled well based on one another, they are mentally close. Therefore, if manner verbs of different fulfilment types show a range of relatedness to their compatible result verbs, it could show that they do not constitute a homogeneous manner verb group, thereby evidencing the existence of a mental cline between manner/result verb representations.

3.2.1 Participants
A priori calculation on G*Power (version 3.1.9.4) for this experiment yielded a sample size of 48. A new group of 101 adult monolingual English speakers were recruited, among whom 3 (2.97%) were rejected for breaching instructions, 27 (26.73%) dropped out, and 3 (2.97%) were timed out by Prolific. Data from the remaining 68 (67.33%) participants (mean age = 33.81 years; age range: 20–50 years;

34 females; ethnicity: 2 Black, 2 mixed, 64 White; student status: 44 no student, 6 students, 18 unknown; employment status: 34 full-time, 7 part-time, 5 unemployed, 22 unknown) was used.

3.2.2 Materials

This experiment used 42 pairs of verbs. Each verb pair contained a target verb that was intrinsic-, moot-, implied-, or attained-fulfilment and a semantically compatible attained-fulfilment verb (e.g., *wash* [implied-fulfilment] – *clean* [attained-fulfilment]). The 42 target verbs were 12 intrinsic-fulfilment, 12 moot-fulfilment, 12 implied-fulfilment, and 6 attained-fulfilment verbs. This means that 36 pairs featured verbs with different fulfilment types, and 6 pairs contained attained-fulfilment verbs only. Attained-fulfilment verbs were compared between themselves to provide a thorough comparison for the fulfilment type hierarchy – as both verbs in such pairs are result verbs, they should reveal the shortest semantic distance compared to pairs featuring different fulfilment types. For simplicity, verb pairs containing the four types of target verbs are hereafter abbreviated as $intrinsic_T$/$moot_T$/$implied_T$/$attained_T$ pairs. The verbs in each pair were presented side by side on the screen, with target verbs appearing on the left in half of the trials and on the right in the other half for each participant.

As word frequency is a well-known factor influencing recall performance (e.g., Gregg, 1976), we collected the frequency of each verb via Google Books Ngram Viewer (https://books.google.com/ngrams/, British English corpus 2019) for the year 2018. A univariate ANOVA was performed on R (version 4.1.2, R Core Team) to explore whether word frequencies of verbs varied across the four fulfilment type groups. No significant difference was found ($F(3) = 0.34$, $p = .798$), suggesting that word frequency could not have significantly influenced recall performance. Full frequency data and analysis are available in Appendix 3.

The verb pairs were also given context sentences outlining intention in half of the trials for each participant. For example, the pair *pull* (intrinsic-fulfilment) – *pluck* (attained-fulfilment) was given the following context: *A fruit was hanging from the branch. I wanted to have it.* This factor was included to examine whether verb-external cues could strengthen the relatedness between verbs.

3.2.3 Experiment design

The present experiment had three within-subjects variables: a) context (present/absent), b) fulfilment type of the target verb (intrinsic/moot/implied/attained), and c) presentation side of the target verb (left/right). The side-of-presentation factor was considered because the recall task required responses for 'verbs on the right of the screen' based on 'verbs on the left of the screen' in the instructions. This factor is apparently inapplicable to $attained_T$ pairs, but to ensure consistency, one verb in each $attained_T$ pair was nevertheless designated as the target verb and also shifted in side of presentation. Among the 42 verb pairs, the total number of $attained_T$ pairs (i.e., 6 pairs) is half that of any other type of verb pairs (i.e., 12 pairs) because shifting in side of presentation in $attained_T$ pairs would not alter the fulfilment types of the verbs on each side.

To avoid high cognitive load and floor effects in recall performance, each participant completed only 28 trials, instead of the full list of 42 pairs. The 28 pairs comprised 2 pairs per each side of presentation and per context condition for

Table 4. Design of Experiment 2 from the perspective of a single participant

Context	Fulfilment type of target verb	Side of target verb	Number of trials
Yes	Intrinsic	Left	2
		Right	2
	Moot	Left	2
		Right	2
	Implied	Left	2
		Right	2
	Attained	Left or right	2
No	Intrinsic	Left	2
		Right	2
	Moot	Left	2
		Right	2
	Implied	Left	2
		Right	2
	Attained	Right or left	2

intrinsic$_T$, moot$_T$, and implied$_T$ pairs, as well as 2 attained$_T$ pairs per context condition. For each participant, if the target verbs in two attained$_T$ pairs appeared on the left with context, the target verbs of the other two attained$_T$ pairs would appear on the right without context, and vice versa. The experiment design from the perspective of a single participant is given in Table 4.

3.2.4 Procedure

This experiment consisted of two phases connected by a distractor task. In the encoding phase, the verbs in each pair were displayed side by side in the upper middle of the screen, one pair per trial and randomised for order. For half of the trials, context sentences appeared above the verbs. For each verb pair, participants were asked *Does the verb on the right have the same intended result as the verb on the left?* After all pairs were completed, a 2-min sound-detecting task served as a distractor ('How many times did you hear sea waves in this music audio?'). Next, participants were shown the verb previously presented on the left for each pair and recalled the verb on the right by typing it (or '?' when they could not remember a verb) into a designated textbox. No contextual sentences were shown in the recall phase, but the side of presentation within each pair was kept identical to the encoding phase. The verbs in each pair were displayed in infinitive *to do* form rather than the base form, as bare verbs can sometimes serve as nouns (e.g., the bare word *hammer* is ambiguous between a tool and an action, see Vigliocco et al., 2005). The entire task lasted about 10 min.

   A potential criticism for asking for intention judgments is that it encourages participants to think in intention/result terms. This encouragement is necessary because only when one thinks in intention/result terms can one discuss fulfilment types,[9] which by definition are based on the recognition of intended results. Moreover, intention judgments can tell us whether and how easy it is to recognise the same intention in a pair of verbs, which provides further evidence that the target verbs in

---
[9]Agent intentionality is a prerequisite for the discussion of Realisation (Talmy, 2000, pp. 262–278; p.c.).

the pairs truly differ in their denotation of intention (recall that intrinsic-fulfilment verbs are theorised to denote no intention).

Due to space limitations, we only present below the analysis for the more relevant recall task; analysis and discussion for intention judgments are presented in Appendix 5. For recall accuracy, we expected intrinsic-, moot-, implied-, and attained-fulfilment verbs to be increasingly related to their corresponding result verbs, manifesting as better recall. It was further expected that the presence of context would facilitate recall.

3.2.5 Data analysis

For recall performance, 1904 data points were collected, which were coded as '1' (correct) or '0' (incorrect/non-attempts). We performed a binomial mixed-effect logistic regression predicting recall accuracy (1/0) based on target verbs' fulfilment types (intrinsic-/moot-/implied-/attained-fulfilment), their side of presentation (left/right), their interactions, as well as a random intercept of Participant. Regression was performed on all possible combinations of variables and their interactions, and we present below the best-fit model with the lowest AIC, which removed context from the analysis. Indeed, context did not prove to significantly predict recall performance when included in any model. Post hoc analysis followed the procedure outlined in Section 3.1.5.

3.2.6 Results

Table 5 presents the raw counts and percentages of correct recalls in all types of verb pairs.

Table 6 presents the best-fit logistic regression model. Fulfilment types and side of presentation of target verbs were both returned as significant fixed effects; there was further an interaction between the two.

Pairwise comparisons (Table 7) were performed to follow up on the fixed effect of fulfilment type and the interaction. For the main effect of side of presentation, as this factor is binary (left/right), it is clear from the regression model that target verbs

Table 5. Raw counts and percentages of correct recalls

Fulfilment type of target verb	Context	Side of target verb	Correct recalls
Intrinsic	Yes	Left	42/136 (30.88%)
		Right	50/136 (36.76%)
	No	Left	40/136 (29.41%)
		Right	49/136 (36.03%)
Moot	Yes	Left	52/136 (38.24%)
		Right	50/136 (36.76%)
	No	Left	51/136 (37.50%)
		Right	41/136 (30.15%)
Implied	Yes	Left	87/136 (63.97%)
		Right	71/136 (52.21%)
	No	Left	81/136 (59.56%)
		Right	74/136 (54.41%)
Attained	Yes	Left	61/74 (82.43%)
		Right	43/62 (69.35%)
	No	Left	49/62 (79.03%)
		Right	52/74 (70.27%)

Table 6. Generalised linear mixed model fit by maximum likelihood for correct recalls

R code: glmer(RecallAccuracy ~ FulfilmentType * Side + (1 | Participant), data = Experiment2, family = "binomial", control = glmerControl(optimizer = "bobyqa"))					
AIC	BIC	LogLik	Deviance	Df.Resid	
2380.50	2430.50	−1181.30	2362.50	1895	
Scaled residual					
Min	1Q	Median	3Q	Max	
−2.74	−0.76	−0.48	0.82	2.76	
Random effects					
Groups		Name	Variance	SD	
Participant		(Intercept)	0.40	0.64	
No. of obs. 1904		Groups:	Participant, 68		
Fixed effects					
	Estimate	SE	*z*-value	Pr(>|z|)	Signif.
(Intercept)	1.56	0.24	6.55	<0.001	***
Intrinsic	−2.48	0.27	−9.32	<0.001	***
Moot	−2.10	0.26	−8.04	<0.001	***
Implied	−1.04	0.26	−3.98	<0.001	***
Target (R)	−0.64	0.30	−2.17	0.030	*
Intrinsic: Target (R)	0.95	0.35	2.70	0.007	**
Moot: Target (R)	0.43	0.35	1.24	0.216	
Implied: Target(R)	0.27	0.35	0.76	0.446	
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '					

Abbreviations: AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion; SD, standard deviation; SE, standard error.

Table 7. Post hoc analysis for recall accuracy

Pairwise comparison	*Z* score	Bonferroni-adjusted *p*	OR (95% CI)
[Fixed effect] fulfilment type			
Implied$_T$ − Intrinsic$_T$	8.28	<0.001	3.00 (2.13, 4.21)
Implied$_T$ − Moot$_T$	7.48	<0.001	2.67 (1.90, 3.74)
Attained$_T$ − Intrinsic$_T$	11.20	<0.001	7.40 (4.67, 11.70)
Attained$_T$ − Moot$_T$	10.61	<0.001	6.58 (4.17, 10.38)
Attained$_T$ − Implied$_T$	5.17	<0.001	2.47 (1.58, 3.86)
Intrinsic$_T$ − Moot$_T$	−0.88	1.000 n.s.	n/a
[Interaction] Fulfilment Type: Side			
When target verbs appeared on the left			
Implied$_T$ − Intrinsic$_T$	7.57	<0.001	4.22 (2.59, 6.89)
Implied$_T$ − Moot$_T$	5.76	<0.001	2.90 (1.80, 4.67)
Attained$_T$ − Intrinsic$_T$	9.32	<0.001	11.90 (6.01, 23.54)
Attained$_T$ − Moot$_T$	8.03	<0.001	8.17 (4.18, 16.00)
Attained$_T$ − Implied$_T$	3.98	<0.001	2.82 (1.44, 5.49)
Intrinsic$_T$ − Moot$_T$	−1.98	0.286 n.s.	n/a
When target verbs appeared on the right			
Implied$_T$ − Intrinsic$_T$	4.12	<0.001	4.22 (2.59, 6.89)
Implied$_T$ − Moot$_T$	4.84	<0.001	2.45 (1.52, 3.94)
Attained$_T$ − Intrinsic$_T$	6.48	<0.001	4.60 (2.51, 8.42)
Attained$_T$ − Moot$_T$	7.03	<0.001	5.29 (2.88, 9.73)
Attained$_T$ − Implied$_T$	3.32	0.006	2.16 (1.19, 3.92)
Intrinsic$_T$ − Moot$_T$	0.75	1.000 n.s.	n/a

Abbreviations: OR, odds ratio; 95% CI, 95% confidence interval.

appearing on the right elicited poorer recall accuracy than target verbs appearing on the left ($\beta = -0.64$, $p = .030$). Thus, the calculation in the regression model already includes the only possible pairwise comparison for this effect, so it is not included in the post hoc analysis.

Following up the fixed effect of fulfilment type, significant distinctions were found between all fulfilment types except between intrinsic$_T$ and moot$_T$ pairs, with attained$_T$ pairs eliciting the best recall performance and intrinsic$_T$/moot$_T$ the poorest. Turning to the interaction between fulfilment type and side of presentation, we found the same pattern, no matter whether target verbs appeared on the left or right.

In summary, recall performance in Experiment 2 showed increasing accuracy for intrinsic$_T$, moot$_T$, implied$_T$, and attained$_T$ pairs, although the difference between intrinsic$_T$ and moot$_T$ did not reach significance.

3.2.7 Discussion

The recall task measured the cognitive relatedness between verbs with different fulfilment types to their semantically compatible result verbs. Understandably, attained-fulfilment verbs, being result verbs, were the closest to their corresponding result verbs and yielded significantly better recall performance than the other verb pairs. What deserves more attention is that intrinsic-, moot-, and implied-fulfilment verbs did not form one homogeneous group (as would be expected of them due to their semantic status as manner verbs) but split into two groups, with intrinsic- and moot-fulfilment verbs clustering together with poorer recall performance and implied-fulfilment verbs independently forming another group with better memory accuracy. The former cluster suggests that although moot-fulfilment verbs, compared to intrinsic-fulfilment ones, denote intended results, this does not necessarily draw them mentally closer to result verbs; rather, it is the implication of intention realisation (as in implied-fulfilment verbs) that mentally demarcates manner verbs in terms of their proximity to result.

Summarising, while interpreted distinctively from attained-fulfilment (i.e., result) verbs, intrinsic-/moot-/implied-fulfilment (i.e., manner) verbs are interpreted with significant within-group variation. The better recall performance elicited by implied-fulfilment verbs than by intrinsic-/moot-fulfilment verbs supports that the former are mentally represented as closer to result verbs than the latter are. This suggests that intrinsic-, moot-, and implied-fulfilment verbs cannot be subsumed under one homogeneous group, inviting the conclusion that there are more than two types of verbs in the mental lexicon.

4. General discussion

In this study, we investigated whether verbs of the four fulfilment types (Talmy, 2000, pp. 262–268) fall into more than two verb representations in the mental lexicon. Through a novel verb learning task, Experiment 1 showed that novel verbs whose definitions entail no result (i.e., having a semantic status of manner verbs) vary in manner and result salience. Experiment 2 revealed through a recall task that manner verbs with different fulfilment types differ in terms of their mental distance with relevant results. This variation suggests that verbs of different fulfilment types have more nuanced representations in the mind than the dichotomous manner/result verb split.

Of note, although Experiments 1 and 2 offer similar results, they are not comparable due to their disparate experimental paradigms. Future studies might seek to use a single task paradigm to assess both novel and genuine words with respect to

intention/fulfilment types and manner/result. Relatedly, recall that the definitions in Experiment 1 might have heightened the distinctions between the four fulfilment types, so future studies are necessary to see whether fulfilment type readings can be generated without definitions. Another limitation is that intention was presumed to be the sole explanation for the manner/result blur. Although this derives from Talmy's fulfilment type framework, there might be other factors modulating verbs' mental representations as manner/result verbs. A further limitation is that this study examines English only, and we aim to expand this line of research to cross-linguistic settings in future studies. In addition, although both experiments reflect activities involved in general learning scenarios, external validity might be less relevant for Experiment 1 due to the involvement of novel scenes and nonce verbs, compared to Experiment 2. Moreover, in-person data collection, had it been feasible for this study, might have provided better data quality than remote participation. In-person participation allows participants to ask questions about the experimental procedures to better follow the instructions. Face-to-face participation can additionally control the practical factors involved in experimentation (e.g., screen size/lightness/resolution; internet speed), which could minimise variation in the quality of video display and thus ensure that unexpected manner/result judgments were not due to unsatisfactory display quality. Additionally, the face-to-face mode makes it easier to verify whether participants meet the sampling criteria (e.g., language status, literacy).[10] Future empirical studies would benefit from mixed modes of participation to cross-examine the validity of experimental procedures and accessibility to participants in both settings. Further, our participants were predominantly white British nationals resident in the UK; the generalisability of the current findings might increase should a broader range of demographic factors be considered. Despite these limitations, let us see what insights the present study can offer.

4.1 A mental blur between manner and result verb representations

Although intrinsic-, moot-, implied-fulfilment verbs are traditionally presumed to be manner verbs and attained-fulfilment verbs are presumed to be result verbs, manner verb interpretation shows significant within-group variation.

In light of the fulfilment type framework, intention seems to underlie this variation. The presence of intention equips manners with the potential of specific results obtaining. Just as intention can shorten the distance between causing and caused events in cognition (Section 2), it can likewise shorten the distance between manner and result in the mental lexicon. This parallelism is understandable as verbs are the most central elements in event description (Flecken et al., 2016, p. 117; Gerwien & von Stutterheim, 2018, p. 227; Hohenstein, 2005, p. 407). Thus, the dichotomous manner/result verb classification framework might not be fine-grained enough to accommodate the observed variation.

A pertinent question is why there is only variation among manner verbs but not in result verbs. The obvious answer is that there are presupposed to be three groups of manner verbs (intrinsic-/moot-/implied-fulfilment) but only one type of result verb

---

[10]That said, the participants in the current study could not have lied about these factors in an attempt to take part, because the study was only advertised to participants who had already registered their relevant information on the recruitment platform.

(attained-fulfilment), so result verbs are experimentally treated as one single group and thus not in a position to show variation. While this is perhaps a limitation, one wonders *why there are three types of manner verbs but only one type of result verb to begin with.* In other words, the blur is driven by manner verbs accommodating result, but not result verbs accommodating manner. The question thus becomes: why is it easier for manner verbs to display different levels of result, but more difficult for result verbs to display different levels of manner? The reason might be that intention is result-oriented – an intention is essentially a result not yet materialised – and thus points unidirectionally from manner to result, rather than from result to manner. Such unidirectionality makes it easier for manner verbs to denote result information but less easy for result verbs to denote manner information. It is worth examining other languages to see whether more variation occurs for result verbs.

Although we found a cognitive cline between manner and result verb representations, we are aware that Rappaport Hovav and Levin's (2010) manner/result complementarity is a *semantic* hypothesis – not a cognitive one. Therefore, we make no attempt to refute the theory semantically. Instead, we aim to show that manner and result are mentally interlinked, leading to a more nuanced perception of verb meaning beyond semantics. However, we do wish to highlight the following difference between verb meaning in semantics and cognition: while it is semantically economical to have manner/result complementarity to limit lexicon size, it is cognitively economical to relate manner to result through intention, as goal intentions provide the backbone for event apprehension (Gerwien & von Stutterheim, 2018; Zwaan & Radvansky, 1998, p. 178).

4.2 Implications for Talmyan event typology in Realisation

The variation among manner verbs with different fulfilment types has implications for event typology in the Realisation domain. According to Talmy (2000), events can be described through the satellite-framed pattern (with the framing event expressed in satellites, 'S-pattern' hereafter) or the verb-framed pattern (with framing event expressed in verbs, 'V-pattern' hereafter) (equipollent-framing will be mentioned later). In Motion, the S-pattern encodes path in the satellite and manner in the verb (e.g., 'I ran into the room'[11]), while the V-pattern encodes path in the verb and often omits manner or expresses manner through means such as the gerundive form (e.g., Spanish: *Entré a la habitación (corriendo)*, 'I entered the room (running)'). This typology has been widely attested for Motion (e.g., Gennari et al., 2002; Özçalışkan & Slobin, 1999).

Although seemingly unrelated, the underlying prerequisite for the S/V dichotomy is manner/result complementarity (Croft, 2012, p. 293). Rappaport Hovav and Levin (2010, p. 7) invoked the manner/path verbs in Motion as support for manner/result complementarity: 'motion verbs appear to fall into either [manner] or [direction], … (suggesting) a manner/direction complementarity akin to manner/result complementarity'. Namely, *manner-of-motion verbs are manner verbs, and path verbs are result verbs.* Since the S-pattern uses manner(-of-motion) verbs and the V-pattern

---

[11]The status of the preposition *into* in this example is controversial. Talmy (1985, 2000) distinguished satellites from prepositions, while other scholars such as Filipović (2007, p. 35) argued that this distinction is not necessary. In the present study, we extend the term *satellite* to any adjunct to the verb, including prepositions.

features result (path) verbs, the contrast between the S- and V-patterns to a great extent lies in the contrast between manner and result verbs, because the opposition of manner-of-motion/path verbs ensures that manner and path information do not overlap in an event expression. Manner-of-motion verbs seldom specify path, as evidenced by their compatibility with all sorts of paths (Gerwien & von Stutterheim, 2018, p. 228, e.g., *fly up/into/away/along*). Thus, if a manner-of-motion verb is used to encode a Motion event, satellites need to be added to indicate path, generating an S-pattern; otherwise, a path verb needs to be used, making a V-pattern. What would happen to the S/V dichotomy if the manner/result contrast dissolves in the mental representation of verbs?

In relation to the idea that manner verbs can point to intended results and sometimes imply their realisation, let us imagine a manner-of-motion verb simultaneously implying path, say, *bex* means 'run to enter and possibly do enter; then it would be difficult to judge whether *I bexed the room* is satellite- or verb-framed because such manner verbs would cognitively denote both manner and result/path and could somewhat serve as result/path verbs in real-life language use and no longer require the path satellite *into*.

The capacity for manner verbs to conceptually denote result information might be reminiscent of equipollent-framing (or E-pattern, Slobin, 2004; Zlatev & Yangklang, 2004), but they are essentially different. In the E-pattern, both the framing event and co-event are lexicalised *separately* by different linguistic units[12]; the only ambiguity is whether these units are verbs. In the current discussion, however, the lack of clarity lies in whether the event verb depicts the co-event only or both the co-event and the framing event; to be more specific, it is unclear whether manner verbs denote only intention (co-event) or both intention and its realisation (framing event), as modulated by their fulfilment types. Hence, it is difficult for languages to exhibit stable typological tendencies in Realisation.

Developing this argument further, let us discuss the *presence of event integration in event expressions*. Talmy (2000) formulated event integration as a *conceptual* process, which surfaces at the *linguistic* level as lexicalisation patterns. However, the thorny issue is that it is not always certain whether event integration is conceptually real by solely looking at linguistic representation. That is, how do we know whether an event expression really packs two simplex events (a framing event and a co-event) or only one?

In Motion, event integration is straightforward. In S- and E-patterns (e.g., English [S]: 'I ran into the room'; Mandarin [E]: *Wǒ pǎo-jìn-le fángjiān* 'I run-enter-PAST the room'), event integration is evidenced by the presence of the satellite or a second verb added to the main verb, so there are clearly two simplex events. In the V-pattern (e.g., *I entered the room*), although there is only one linguistic element expressing the event, namely the path verb, it invariably encodes both path and the fact of motion, thus still encoding two event aspects.

Turning to Realisation, the picture becomes obscure. In the S-pattern (e.g., *I hunted the fugitive down*), event integration is likewise signalled by the addition of satellites, so event integration has undoubtedly taken place in this kind of expression. However, when single manner verbs are used (without adverbials,

---

[12]The E-pattern can also be encoded in one single linguistic unit in the form of 'bipartite verbs' in languages such as Algonquian, Athabaskan, and Hokan (DeLancey, 1989; Slobin, 2004). The bipartite verb encodes a manner morpheme and a path morpheme. However, such verbs do not fall into the scope of manner/result complementarity, which only discusses single-root verbs.

complements, etc.), the presence of two simplex events is dubious. Given the blur in manner/result verb representations in the mind, manner verbs can sometimes be used as if they denoted result. For example, *I washed the shirt* is lexically not a macro-event as it only denotes the cleaning intention and no realisation of intention (i.e., no result), but in everyday language use it may well be used to mean *I cleaned the shirt*, describing both the presence and the realisation of the cleaning intention. Thus, when manner verbs are mentally close to result verbs, Realisation expressions with single manner verbs could also depict event integration. Namely, *event integration could occur mentally when lexically it does not.* Thus, the overt framing pattern of event expressions might no longer be a reliable indicator of event integration in Realisation.

We do not intend to negate Realisation as a valid macro-event domain, but rather to provide insights for future research in this domain. The conclusion to be drawn is that framing patterns might not be as stable and effective for Realisation as it is for Motion (and possibly other macro-event domains). A more promising direction would be to scale down from holistic event expressions and focus on 'verb-sized' events (Billman & Krych, 1998). Also, it is worth exploring how verbs come to have different fulfilment types and hence nuanced verb representations in the mind, as well as the cognitive implications of this manner/result cline.

5. Conclusion

This study is the first empirical investigation into the semantics of verbs following the fulfilment type framework proposed in Realisation macro-events in the event integration theory (Talmy, 2000). Meanwhile, it is also an attempt to extend Talmyan event research beyond Motion to less well-studied Realisation. Compared to the traditional manner/result verb classification (Rappaport Hovav & Levin, 2010), the four fulfilment types (intrinsic-/moot-/implied-/attained-fulfilment) offer a more fine-grained angle for verb meaning, moving from a semantic dichotomous manner/result split to a mental hierarchy with manner and result carrying different weight as modulated by intention linking the two. Although intrinsic-/moot-/implied-fulfilment verbs are often mentally represented as manner verbs, they are progressively closer to having a result verb representation. This blur between manner and result might cloud the typological tendencies of languages in the Realisation domain. Consequently, framing patterns might not be as stable and effective for Realisation as it is for Motion.

References

Aurnague, M. (2011). How motion verbs are spatial: The spatial foundations of intransitive motion verbs in French. *Lingvisticae Investigationes*, 34(1), 1–34.

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/10.18637/jss.v067.i01

Behrend, D. A. (1990). The development of verb concepts: Children's use of verbs to label familiar and novel events. *Child Development*, 61, 681–696.

Behrend, D. A. (1995). Processes involved in the initial mapping of verb meanings. In M. Tomasello, & W. E. Mcrriman (Eds.), *Beyond names for things: Young children's acquisition of verbs*. Erlbaum.

Behrend, D. A., & Scofield, J. (2006). Verbs, actions, and intentions. In K. Hirsh-Pasek, & R. M. Golinkoff (Eds.), *Action meets word: How children learn verbs* (pp. 286–307). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195170009.003.0012

Behrend, D. A., & Wittek, A. (2003). Toddlers use verbs as cues to an actor's intentions. *Paper presented at the meetings of the Society for Research in Child Development*, Tampa, FL.

Billman, D., & Krych, M. (1998). Path and manner verbs in action: Effects of "Skipping" or "Exiting" on event memory. In *Proceedings of the 20th annual conference of the cognitive science society*. Lawrence Erlbaum Associates.

Billman, D., Swilley, A., & Krych, M. (2000). Path and manner priming: Verb production and event recognition. In *Proceedings of the 22nd annual conference of the cognitive science society* (pp. 615–620). Lawrence Erlbaum Associates.

Buccola, B., Dautriche, I., & Chemla, E. (2018). Competition and symmetry in an artificial word learning task. *Frontiers in Psychology*, 9, 2176. https://doi.org/10.3389/fpsyg.2018.02176

Carroll, M., & von Stutterheim, C. (2010). Event representation, time event relations, and clause structure: A cross-linguistic study of English and German. In J. Bohnemeyer, & E. Pederson (Eds.), *Event representation in language and cognition* (pp. 68–82). Cambridge University Press.

Chan, C. C. Y., Tardif, T., Chen, J., Pulverman, R. B., Zhu, L., & Meng, X. (2011). English- and Chinese-learning infants map novel labels to objects and actions differently. *Developmental Psychology*, 47(5), 459–1471.

Croft, W. (2012). *Verbs: Aspect and causal structure*. Oxford University Press.

DeLancey, S. (1989). Klamath stem structure in genetic and areal perspective. In *Papers from the 1988 Hokan-Penutian languages workshop* (pp. 31–39). University of Oregon.

Engemann, H., Hendriks, H., Hickmann, M., Soroli, E., & Vincent, C. (2015). How language impacts memory of motion events in English and French. *Cognitive Processing*, 16(1), 209–213. https://doi.org/10.1007/s10339-015-0696-7

Filipović, L. (2007). *Talking about motion: A crosslinguistic investigation of lexicalization patterns*. John Benjamins Publishing Company.

Filipović, L. (2011). Speaking and remembering in one or two languages: Bilingual versus monolingual lexicalization and memory for motion events. *International Journal of Bilingualism*, 15, 466–485.

Filipović, L. (2022). First language versus second language effect on memory for motion events: The role of language type and proficiency. *International Journal of Bilingualism*, 26(1), 65–81.

Flecken, M., Carroll, M., Weimar, K., & von Stutterheim, C. (2016). Driving along the road or heading for the village? Conceptual differences underlying motion event encoding in French, German, and French-German L2 users. *The Modern Language Journal*, 99, 100–122.

Flecken, M., von Stutterheim, C., & Carroll, M. (2014). Grammatical aspect influences motion event perception: Findings from a cross-linguistic non-verbal recognition task. *Language and Cognition*, 6(1), 45–78. https://doi.org/10.1017/langcog.2013.2

Forbes, J. N., & Farrar, M. J. (1993). Children's initial assumptions about the meaning of novel motion verbs: Biased and conservative? *Cognitive Development*, 8, 273–290.

Gennari, S. P., Sloman, S. A., Malt, B. C., & Fitch, W. T. (2002). Motion events in language and cognition. *Cognition*, 83, 49–79.

Gerwien, J., & von Stutterheim, C. (2018). Event segmentation: Crosslinguistic differences in verbal and nonverbal tasks. *Cognition*, 180, 225–237.

Gillette, J., Gleitman, L. R., Gleitman, H., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, 73, 135–176.

Gregg, V. (1976). Word frequency, recognition and recall. In J. Brown (Ed.), *Recall and recognition*. John Wiley & Sons.

Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature Neuroscience*, 5(4), 382–385. https://doi.org/10.1038/nn827

Hickmann, M., Engemann, H., Soroli, E., Hendriks, H., & Vincent, C. (2017). Expressing and categorizing motion in French and English: Verbal and non-verbal cognition across languages. In I. Ibarretxe-Antuñano (Ed.), *Human cognitive processing* (vol. 59, pp. 61–94). John Benjamins Publishing Company.

Hohenstein, J. (2005). Language-related motion event similarities in English- and Spanish-speaking children. *Journal of Cognition and Development*, 6(3), 403–425.

Hohenstein, J., Eisenberg, A., & Naigles, L. (2006). Is he floating across or crossing afloat? Cross-influence of L1 and L2 in Spanish-English bilingual adults. *Bilingualism: Language and Cognition*, 9(3), 249–261.

Hohenstein, J., Naigles, L., & Eisenberg, A. (2004). *Keeping verb acquisition in motion: A comparison of English and Spanish*. In D. Halls, & S. Waxman (Eds.), *Weaving a Lexicon* (pp. 569–602). MIT Press.

Huang, C., Heyes, C., & Charman, T. (2002). Infants' behavioral reenactment of "failed attempts": Exploring the roles of emulation learning, stimulus enhancement, and understanding of intentions. *Developmental Psychology*, 38, 840–855.

Imai, M., Li, L., Haryu, E., Okada, H., Hirsh-Pasek, K., & Golinkoff, R. (2008). Novel noun and verb learning in Chinese-, English-, and Japanese-speaking children. *Child Development*, 79, 979–1000.

Ji, Y., Hendriks, H., & Hickmann, M. (2011). The expression of caused motion events in Chinese and in English: Some typological issues. *Linguistics*, 49(5), 1041–1076.

Ji, Y., & Hohenstein, J. (2017). Conceptualising voluntary motion events beyond language use: A comparison of English and Chinese speakers' similarity judgments. *Lingua*, 195, 57–71.

Ji, Y., & Papafragou, A. (2020). Is there an end in sight? Viewers' sensitivity to abstract event structure. *Cognition*, 197, 104197. https://doi.org/10.1016/j.cognition.2020.104197

Jia, H. [贾红霞], & Li, F. [李福印]. (2015). Conceptual divisions between the State Change event and the Realisation event [状态变化事件与实现事件的概念界定]. *Foreign Language and Edudation [外语教学]*, 36(1), 22–27.

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.

Kim, J. (2000). *Mind in a physical world*. MIT Press.

Kou, X., & Hohenstein, J. (2022). Fulfilment types in realisation verbs: Definition, distinctions, and diagnostics. *Lingua*, 269, 103204. https://doi.org/10.1016/j.lingua.2021.103204

Kowialiewski, B., Krasnoff, J., Mizrak, E., & Oberauer, K. (2022). The semantic relatedness effect in serial recall: Deconfounding encoding and recall order. *Journal of Memory and Language*, 127, 104377.

Lenth, R. V. (2022). *emmeans: Estimated marginal means, aka least-squares means. R package version 1.7.3.* https://CRAN.R-project.org/package=emmeans

Levin, B., & Rappaport Hovav, M. (2014). Manner and result: The view from clean. In R. Pensalini, M. Turpin, & D. Guillemin (Eds.), *Language description informed by theory* (pp. 337–357). John Benjamins Publishing Company.

Li, F. [李福印]. (2013). Two systematic errors in macro-event research [宏事件研究中的两大系统性误区]. *Foreign Languages in China [中国外语]*, 2, 25–33.

Maguire, M. J., Hirsh-Pasek, K., Golinkoff, R. M., Imai, M., Haryu, E., Vanegas, S., Okada, H., Pulverman, R., & Sanchez-Davis, B. (2010). A developmental shift from similar to language-specific strategies in verb acquisition: A comparison of English, Spanish, and Japanese. *Cognition*, 114, 299–319.

Moore, J. W., Teufel, C., Subramaniam, N., Davis, G., Fletcher, P.C. (2013). Attribution of intentional causation influences the perception of observed movements: Behavioral evidence and neural correlates. *Frontiers in Psychology*, 4, 1–11.

Naigles, L. R., & Terrazas, P. (1998). Motion-verb generalizations in English and Spanish: Influences of language and syntax. *Psychological Science*, 9, 363–369.

Nakamura, T., Marque-Pucheu, C., Leclère, C., & Laporte, E. (2021). J'ai couru au bar ('I ran to the bar'): a purpose-oriented verb of manner of motion. NAMED 2020: De/constructing motion events. 1–2 July 2020.

Özçalışkan, Ş., & Slobin, D. I. (1999). Learning how to search for the frog: Expression of manner of motion in English, Spanish, and Turkish. In A. Greenhill, H. Littlefield, & C. Tano (Eds.), *Proceedings of the 23rd annual Boston University conference on language development* (pp. 541–552). Cascadilla Press.

Papafragou, A., Massey, C., & Gleitman, L. (2002). Shake, rattle, 'n' roll: The representation of motion in language and cognition. *Cognition*, 84(2), 189–219. https://doi.org/10.1016/S0010-0277(02)00046-X

Peng, Y., Ichien, N., & Lu, H. (2020). Causal actions enhance perception of continuous body movements. *Cognition*, 194, 104060.

Pourcel, S. (2003). Rethinking 'thinking for speaking'. *Proceedings of the annual meeting of the Berkeley Linguistics Society*, 29(1), 349–358.

Pourcel, S. (2004). What makes path of motion salient? *Proceedings of the Berkeley Linguistics Society*, 30, 505–516.

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rappaport Hovav, M., & Levin, B. (2010). Reflections on manner/result complementarity. In M. R. Hovav, E. Doron, & Sichel (Eds.), *Lexical semantics, syntax, and event structure* (pp. 21–38). Oxford University Press.

Roettger, T. (2021). Context sensitivity and failed replications in linguistics – A reply to Grieve. *Linguistics*, 59 (5), 1357–1358. https://doi.org/10.1515/ling-2020-0239

Santin, M., van Hout, A., & Flecken, M. (2021). Event endings in memory and language. *Language, Cognition and Neuroscience*, 36(5), 625–648.

Sarda, L. (2019). French motion verbs: Insights into the status of locative PPs. In M. Aurnague, & D. Stosic (Eds.), *Human cognitive processing: The semantics of dynamic space in French: Descriptive, experimental and formal studies on motion expression* (vol. 66, pp. 68–107). John Benjamins Publishing Company. https://hal.science/hal-03350824

Slobin, D. I. (2004). The many ways to search for a frog: linguistic typology and the expression of motion events. In S. Strömqvist, & L. Verhoeven (Eds.), Relating events in narrative: Vol. *2*. Typological and contextual perspectives (pp. 219–257). Lawrence Erlbaum Associates.

Soroli, E., & Hickmann, M. (2011). Language and spatial representations in French and in English: Evidence from eye-movements. In G. Marotta, A. Lenci, L. Meini, & F. Rovai (Eds.), *Space in language* (pp. 581–597). Editrice Testi Scientifici.

Stosic, D. (2009). La notion de « manière » dans la sémantique de l'espace. *Langages*, 175, 103–121.

Talmy, L. (1985). Lexicalisation patterns: semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and syntactic description 3: Grammatical categories and the Lexicon* (pp. 57–149). Cambridge University Press.

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

Ünal, E., & Papafragou, A. (2019). How children identify events from visual experience. *Language Learning and Development*, 15(2), 138–156. https://doi.org/10.1080/15475441.2018.1544075

Ünal, E., Richards, C., Trueswell, J. C., & Papafragou, A. (2021). Representing agents, patients, goals and instruments in causative events: A cross-linguistic investigation of early language and cognition. *Developmental Science*, 24(6), e13116. https://doi.org/10.1111/desc.13116

Vigliocco, G., Vinson, D. P., & Siri, S. (2005). Semantic similarity and grammatical class in naming actions. *Cognition*, 94(3), B91–B100. https://doi.org/10.1016/j.cognition.2004.06.004

Winter, B., & Grice, M. (2021). Independence and generalizability in linguistics. *Linguistics*, 59(5), 1251–1277. https://doi.org/10.1515/ling-2019-0049

Yarkoni, T. (2020). The generalizability crisis. *Behavioral and Brain Sciences*, 45, 1–37. https://doi.org/10.1017/S0140525X20001685

Zacks, J., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, 127, 3–21. https://doi.org/10.1037/0033-2909.127.1.3

Zlatev, J., & Yangklang, P. (2004). A third way to travel: The place of Thai in motion event typology. In S. Strömqvist, & L. Verhoeven (Eds.), *Relating events in narrative: Typological and contextual perspectives* (pp. 159–190). Lawrence Erlbaum.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123(2), 162–185. https://doi.org/10.1037/0033-2909.123.2.162