COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 46 (2022) e13140

ISSN: 1551-6709 online
DOI: 10.1111/cogs.13140

Evidence for a Shared Instrument Prototype from English, Dutch, and German

Lilia Rissman,[a] Saskia van Putten,[b] Asifa Majid[c]

[a]*Department of Psychology, University of Wisconsin-Madison*
[b]*Center for Language Studies, Radboud University*
[c]*Department of Experimental Psychology, University of Oxford*

Received 19 July 2021; received in revised form 24 March 2022; accepted 31 March 2022

Abstract

At conceptual and linguistic levels of cognition, events are said to be represented in terms of abstract categories, for example, the sentence *Jackie cut the bagel with a knife* encodes the categories Agent (i.e., *Jackie*) and Patient (i.e., *the bagel*). In this paper, we ask whether entities such as *the knife* are also represented in terms of such a category (often labeled "Instrument") and, if so, whether this category has a prototype structure. We hypothesized the Proto-instrument is a *tool*: a physical object manipulated by an intentional agent to affect a change in another individual or object. To test this, we asked speakers of English, Dutch, and German to complete an event description task and a sentence acceptability judgment task in which events were viewed with more or less prototypical instruments. We found broad similarities in how English, Dutch, and German partition the semantic space of instrumental events, suggesting there is a shared concept of the Instrument category. However, there was no evidence to support the specific hypothesis that tools are the core of the Instrument category—instead, our results suggest the most prototypical Instrument is the direct extension of an intentional agent. This paper supports theoretical frameworks where thematic roles are analyzed in terms of prototypes and suggests new avenues of research on how instrumental category structure differs across linguistic and conceptual domains.

Keywords: Semantics; Language; Thematic roles; Event cognition; Concepts; Tools

Correspondence should be directed to Lilia Rissman, Department of Psychology, University of Wisconsin-Madison, Madison, WI, USA. E-mail: lrissman@wisc.edu

1. Introduction

When people witness events in the world, such as a baseball player hitting a ball, they represent entities in the event in terms of abstract categories: the player is the one performing the action (often labeled the Agent) while the ball is the one being acted upon (often labeled the Patient) (Hafri, Trueswell, & Strickland, 2018; Rissman & Lupyan, 2021; Rissman & Majid, 2019). These categories (sometimes called *thematic roles*) are encoded in language through morphosyntactic structures, for example, in the English sentence *the baseball player hit the ball*, the fact that *the baseball player* appears in Subject position indicates this individual is the Agent (Dowty, 1989; Fillmore, 1968; Gruber, 1965; Jackendoff, 1987; Levin & Rappaport-Hovav, 2005). Thematic roles are implicated in a variety of cognitive processes, including learning, linguistic processing, and event interpretation (see Rissman & Majid, 2019; Strickland, 2017 for review). Understanding the structure of thematic roles is therefore essential for constructing theories of these cognitive processes. While the category structure of Agent and Patient is fairly well understood, the structure of other thematic roles is poorly understood—here we investigate one such role, a category often labeled Instrument (e.g., the bat in *the baseball player hit the ball with a bat*).

Instrumental markers (e.g., the English preposition *with*) are well-documented descriptively across languages and Instrument appears frequently on lists of proposed thematic roles. Knowledge of instruments may also be part of core cognition, given infants' early understanding of tool use and the importance of tool cognition in human evolution (Haidle, 2014; Neldner et al., 2020; Stavans & Baillargeon, 2018). Instruments are often characterized as the means of performing an action or an object that is used to perform an action, and in English, Instrument is most commonly identified with the lexical items *with* and *use*. A range of theorists have proposed semantic features that delineate the structure of this category, such as the importance of an Instrument being a causal intermediary (Croft, 1991; Koenig, Mauner, Bienvenue, & Conklin, 2008; Rissman & Rawlins, 2017; Schlesinger, 1995; Talmy, 1976). Nonetheless, these proposals leave a range of questions unanswered: if Instrument is a category in conceptual representation, what semantic features determine whether an entity is a better or worse example of an Instrument? Does this category have prototype structure and if so, what is the prototypical Instrument? We address these questions using data from an elicited language production task and a linguistic acceptability judgment task. We take a cross-linguistic approach, comparing category structure across English, Dutch, and German. Investigating multiple languages allows us to investigate the stability of the Instrument category. For the remainder of this paper, we use capitalized "Instrument" to refer to a proposed semantic/conceptual category, and lowercase "instrument" to refer to an individual in an event or the complement of a lexical item such as *with* or *use*.

1.1. What is an Instrument?

For thematic roles such as Agent, Patient, and Goal (e.g., *Jackie threw the ball into the net*), a wealth of empirical evidence points to their cognitive prominence. Infants represent events in terms of these categories and use them to learn language (Carey, 2009; Lakusta, Spinelli,

& Garcia, 2017; Tatone, Geraci, & Csibra, 2015; Yin & Csibra, 2015), adults activate them when interpreting events (Hafri, Papafragou, & Trueswell, 2013; Hafri et al., 2018; Rissman & Lupyan, 2021), deaf individuals creating new sign languages readily innovate morphosyntactic devices for these categories (Ergin, Meir, Aran, Padden, & Jackendoff, 2018; Flaherty, 2014; Goldin-Meadow, Brentari, Coppola, Horton, & Senghas, 2015; Rissman & Goldin-Meadow, 2017), and across the world's languages, these categories are robustly encoded through morphosyntactic structures (Comrie, 2013; Dryer, 2013; Hartmann, Haspelmath, & Cysouw, 2014; Siewierska, 2013).

For the proposed thematic role of Instrument, by contrast, the cognitive prominence and stability of the category is less clear. English-speaking adults often omit instruments when describing instrumental events (e.g., saying *the man cut the bread* when the action was performed with a knife) (Brown & Dell, 1987; Grigoroglou & Papafragou, 2019; Lockridge & Brennan, 2002). When viewing events, English-speaking adults are also slower to recognize instruments than agents, patients, or goals, as revealed through patterns of eye-gaze (Wilson, Papafragou, Bunger, & Trueswell, 2011). In addition, both English- and Turkish-learning children are less likely to notice changes to an instrument than to a patient or goal in a change-detection task (Ünal, Richards, Trueswell, & Papafragou, 2021).

With and *use* are thought to introduce the category Instrument—these lexical items, however, can introduce a range of event participant types which are intuitively more or less instrumental, such as in *Jackie filled the glass with orange juice* and *Ross used his time wisely*. Instrumental participants can also be introduced through other morphosyntactic means, as in *the wrecking ball destroyed the building*. It is unclear whether these event participants (e.g., *orange juice, time*, and *wrecking ball*) should be analyzed in terms of a category called Instrument, or what the boundaries of this category might be (see Koenig et al., 2008; Nilsen, 1973; Rissman & Rawlins, 2017; Schlesinger, 1995).

Attempts to define a category Instrument are further complicated by the fact that *with* and *use* do not have identical meanings (Chomsky, 1972; Lakoff, 1968; Rissman, 2013; Rissman & Rawlins, 2017). For example, Rissman and Rawlins (2017) propose that *use* but not *with* requires an intentional agent, allowing *Michelle bumped the statue with her purse* but prohibiting *#Michelle used her purse to bump the statue* in the same accidental context.[1] Conversely, *with* but not *use* requires the instrument be part of the causal force directed toward the patient, allowing *Jan used a cutting board to chop the carrots* but prohibiting *#Jan chopped the carrots with a cutting board* in the same context. These data demonstrate that neither *with* nor *use* denote a unitary concept of Instrument. In fact, Rissman and Rawlins (2017) employ Agent and Patient, but not Instrument, as representational primitives in their analyses of *with* and *use*. These authors characterize the role Instrument only in broad terms as "an entity, either concrete or abstract, acted on by an agent as part of a larger event" (p. 509).

Across languages, the clusters of meaning expressed by individual instrumental markers (e.g., *with*) are also diverse, spanning a wide range of event components such as location, theme, and manner (Bickel, Zakharko, Bierkandt, & Witzlack-Makarevich, 2014; Luján & Ruiz Abad, 2014; Narrog & Ito, 2007; Stolz, 2001). In Russian, for example, instrumental case can be used for events such as slicing bread with a knife but can also mark a passive agent (Janda, 2013). This extension is not possible for English *with* (e.g., *the book was written*

by Tolstoy, not *#the book was written with Tolstoy*). The wide range of meanings encoded by instrumental markers across languages raises further questions about the prominence and coherence of the proposed category Instrument.

1.2. Is there a prototypical Instrument?

Variability within and across languages does not on its own rule out the possibility that there is a conceptual category Instrument that is important in cognition and development, a category that may even be universal and innate. It may be that people represent events and sentences in terms of an Instrument category that has prototype structure: individual instrumental markers may encode this shared prototype, even if they encode the periphery of the category in different ways (see Rice, 1996; Rosch, 1978; Taylor, 2003).

According to classical theory, categories are defined in terms of necessary and sufficient conditions, and no member of a category has a more prominent status than any other member (see Taylor, 2003 for review). The categories delineated by natural language lexical items do not appear to be structured in this way. In particular, natural language categories demonstrate prototypicality effects: for example, people have gradient judgments about whether a particular animal is a good example of a bird, and when people are asked to list types of birds, some members of this category are listed more quickly and consistently than others (see Geeraerts, 2010; Hampton, 2006; Murphy, 2002; Rosch, 1978 for review). Under a prototype model, lexical categories such as *bird* are represented in terms of a set of attributes (e.g., having feathers) that demarcate an abstract conceptual core.

Agents and Patients are often characterized as individuals who perform an action and individuals who are acted upon, respectively. Nonetheless, thematic roles are better characterized as having prototype structure (Ackerman & Moore, 2001; Dowty, 1991; Gärdenfors, 2014; Grimm, 2011; Hopper & Thompson, 1980; Lakoff & Johnson, 1980; Luraghi, 2003; Primus, 1999). Notably, Dowty (1991) invokes "Proto-Roles" to explain why the arguments of English verbs surface in Subject versus Direct Object position. In Dowty's framework, the properties characteristic of a Proto-Agent are: (1) being volitional, (2) being sentient, (3) playing a causal role, (4) moving, and (5) existing independently of the event. This analysis has been validated and extended in a variety of subsequent linguistic research (Ackerman & Moore, 2001; Grimm, 2011; Kako, 2006; Luraghi, 2003; Primus, 1999; Reisinger et al., 2015; White, Rawlins, & Van Durme, 2017). Many of the Proto-Properties proposed by Dowty are also constitutive of the Agent and Patient categories that English speakers extract from visually presented events (Rissman & Lupyan, 2021).

Is there a Proto-instrument role? If yes, what are the properties that determine whether an individual is more or less prototypically instrumental? The literature on instrumental semantics implicitly assumes some notion of a prototype. A long-utilized diagnostic for identifying Instruments in English is the following: a noun phrase (NP) in a sentence is an Instrument if the *with*- and *use*-variants of a sentence are both grammatical and have the same meaning (Koenig et al., 2008; Lakoff, 1968; Nilsen, 1973). For example, *the knife* in *Jayda sliced the brisket with a knife* would be an Instrument because *Jayda used a knife to slice the brisket* is grammatical and has a comparable meaning to the *with*-version of the sentence. This

diagnostic assumes a conceptual category Instrument with prototype structure: although *with* and *use* do not directly denote the prototype, they jointly point toward what the prototype is.

Here we test the hypothesis that there is a Proto-instrument role organized around the prototype of a *tool* (the *tool-prototype* hypothesis). In studies of tool use among nonhuman animals, a tool is typically defined as a physical object which is manipulated intentionally in order to cause a change in another object or individual (Plotnik & Clayton, 2015; Seed & Byrne, 2010; Vaesen, 2012). This definition aligns closely with how Instruments are characterized in the linguistics literature. For example, Luraghi (2001, p. 388) characterizes an Instrument as "an inanimate manipulable entity which occurs in a controlled state of affairs, where an agent acts intentionally." Tools are essential to building human civilization and culture, as they allow us to manipulate and alter our environment in ways not possible through the affordances of our physical bodies alone. Tool use is also culturally universal (Brown, 2004). As such, the concept of a tool plausibly centers the proposed category Instrument.

Before we outline our research approach, we note that the relationship between prototype analyses of thematic roles and prototype analyses of nouns such as *bird* is not entirely parallel. For the latter type of category, it is the words themselves that are thought to be represented in terms of prototypes. Thematic roles, by contrast, are typically linguistically encoded through morphosyntactic structures (e.g., Subject, ergative case) rather than through lexical labels (e.g., *bird*). *Agent* and *patient* are of course English words, but the common meanings of these words do not closely correspond to the notions of Agent and Patient that we are investigating here. And although there is a mapping between Proto-Agent and Subject in English, Subject is not itself a semantic category. For this reason, thematic roles are often assumed to be categories in conceptual structure or categories at the semantic/conceptual interface, rather than strictly linguistic categories (Dowty, 1991; Jackendoff, 1983, 1987). In other words, the meanings of individual morphemes such as *with* are distinct from the representations of the roles.

These assumptions about thematic role representation are informed by the semantic maps approach to characterizing similarity in thematic role space (Croft, 2003; Haspelmath, 2003; Luján, 2010; Malchukov, Haspelmath, & Comrie, 2010; Narrog & Ito, 2007; Rice & Kabata, 2007). For example, Haspelmath (2003) proposes a semantic/conceptual space in which the instrumental role is most closely related to the comitative role (e.g., *Crystal went to the store with Michelle*) and the passive role (e.g., *the wig was stolen by Gigi*). This similarity space across roles is proposed to be universal, even if particular lexical items carve it up in different ways.

1.3. Approach

We extended prototype analyses of thematic roles to the instrumental domain, hypothesizing that the prototypical Instrument is a tool. We tested this hypothesis through two studies: an event description study and an acceptability judgment study. In the first study, participants viewed videos and described them. In the second study, participants viewed a video and a sentence and judged whether the sentence was a natural description of the video. The most critical feature of our design was the range of videos that participants described and judged.

Table 1 provides the 10 event conditions tested: one Tool condition, seven non-Tool conditions, and two control conditions. The seven nontool event types differed from tool events by one or more dimensions of event representation. Each of these dimensions has been argued to be relevant to understanding Instrument category structure or has been shown to distinguish different instrumental markers within or across languages (see references in Table 1). Rather than presuppose that each of these events has an Instrument, we refer to each event as having a Target participant and we investigate how these Target participants are encoded morphosyntactically.

Most prior studies on thematic role structure have used linguistic data as a window into underlying conceptual structure. We follow the semantic maps literature (reviewed above) in assuming that linguistic similarity reflects conceptual similarity in event representation (see also Majid, Boster, & Bowerman, 2008). Although individual lexical items can have broad extension (and some form $\sim$ meaning mappings are homonymous), we assume that word meanings generally reflect coherent regions of conceptual space. If a Proto-Instrument category constrains the meanings of instrumental markers in individual languages, then analyzing the distribution of these markers (i.e., whether two different events are described using the same marker) sheds light on the conceptual similarity between events. We also adopt a linguistic approach for the sake of comparison, as Dowty (1991) proposed Proto-Agent and Proto-Patient categories based on linguistic data. We do not assume that linguistic and nonlinguistic tasks (or different types of linguistic tasks for that matter) reflect conceptual knowledge in identical ways, a point we return to in Section 4.

We made two additional assumptions in linking the data we collected in Study 1 on event description and Study 2 on acceptability judgments to a hypothesized Proto-Instrument category. First, following classic work by Brown and Lenneberg (1954), we assume that more prototypical members of a Proto-Role will be described using a smaller range of lexical items than less prototypical members of the Proto-Role. If an event participant is described using a large number of lexical items, this may reflect that the participant is encoded in terms of multiple event schemas. For example, Targets in the Locatum category can be described using either instrumental language (e.g., *Gigi used icing to decorate the cake*) or spatial, caused-motion language (e.g., *Jayda put icing onto the cake*). Our second assumption follows the work of Rosch (1975), where participants were asked to judge whether various entities were good examples of categories such as *birds*, *furniture*, and *toys*—consistently affirmative judgments were taken as evidence that an entity was a more prototypical example of a category. Extending this logic to the domain of sentences, we assume if a particular lexical item ℓ can be used to describe members of a Proto-Role, then an utterance containing ℓ will be most acceptable for the most prototypical members of the Proto-Role. We collected both description (Study 1) and judgment (Study 2) data because productive and receptive tasks shed light on speakers' knowledge in different ways. We assume if a particular feature of event representation is characteristic of Proto-Instruments, then events with these features will elicit prototypical language behavior in both studies.

We conducted the two studies with adult speakers of British English, Dutch, and German. These three Germanic languages of the Indo-European family are closely related to each other. In addition, speakers from these populations have similar cultural and educational

Table 1
Experimental conditions, example events, description, and motivation. The Target object in each video is underlined

Condition	Example Event	Description/Motivation
Tool	A woman breaks a plate with a hammer	The Target is the tool.
Contact	A woman taps a cat on the head with a cat toy	Identical to tool events except the patient does not undergo a change of state. Establishes whether causally changing a participant is characteristic of Proto-Instruments.
Body Part	A man knocks over a music stand with his hand	Identical to tool events except the agent performs the action with a part of their own body. Establishes whether being a discrete physical object is characteristic of Proto-Instruments.
Accidental Agent	A woman sweeps with a broom, accidentally knocking over a bottle	An agent intentionally performs one action with the Target thereby accidentally performing a second action with the Target. Establishes whether agent intentionality is characteristic of Proto-Instruments (see discussion of Rissman & Rawlins, 2017, above).
Means of Transit	Planning a trip on Google Maps from Rome to Moscow by plane	The Target is a mode of transportation such as a train or plane. Establishes whether being manipulable by hand is characteristic of Proto-Instruments.
Locatum	A woman fills a glass with orange juice	The Target is a theme in a caused-motion event. Iwata (2008), Jackendoff (1990), and Pinker (1989) analyze these events in terms of spatial transfer; by contrast, Koenig et al. (2008) analyze the Target in these events as an Instrument, given the felicity of both *a woman filled the glass with orange juice* and a *woman used orange juice to fill the glass*. Establishes whether a spatial-theme analysis conflicts with an instrumental analysis.
Indirect Action	A woman climbs a ladder to open a window	The Target is not a direct part of the force exerted by the agent. Establishes whether causal directness is characteristic of Proto-Instruments (see discussion of Rissman & Rawlins, 2017, above).

(*Continued*)

Table 1
(Continued)

Condition	Example Event	Description/Motivation
Inanimate Agent	A train rolls down a track, which bumps a red car, which bumps a truck	The Target is an intermediary in a causal chain from an agent to a patient but the agent is not animate. Previous researchers have defined an Instrument as a causal intermediary, regardless of the animacy of the agent (Croft, 1991; Fillmore, 1968; Koenig & Davis, 2006; Talmy, 1976). For example, Croft (1991, p. 178) characterizes an Instrument as "intermediate in a causal chain between the subject (initiator) and the direct object (final affected entity)." This condition establishes whether agent animacy is characteristic of Proto-Instruments.
Put-Theme (control condition)	A man puts a box on a shelf	The Target is moved from one location to another. These events involve three central participants but have not been characterized as instrumental, delineating an outer boundary on a possible Proto-Instrument category.
Give-Theme (control condition)	A woman gives a mug to another person	The Target is transferred from one person to another. As with Put-Theme events, these events have not been characterized as instrumental and delineate an outer boundary on a possible Proto-Instrument category.

backgrounds (relative to the full range of human culture). We tested English because most previous studies of instrumental semantics are based on English—how English speakers describe and evaluate events therefore provides the most direct test of previous hypotheses about instrumental category structure that were derived primarily from this language. Studies based on English are by and large intended to generalize beyond English—as a test of how well previous hypotheses generalize, we included speakers of Dutch and German. It should not be taken for granted that these three languages carve up instrumental event space in the same way. These languages, in fact, diverge in how verbs lexicalize cutting and breaking events (Majid, Gullberg, van Staden, & Bowerman, 2007), putting and taking events (Kopecka & Narasimhan, 2012), giving events (Newman, 1998), and human locomotion events (Goddard, Wierzbicka, & Wong, 2016; Malt et al., 2014). For example, using cluster analysis, Majid et al. (2007) found different clusters of events of cutting: while German and English had a single cluster of events involving cutting with a single blade (knife) and double blade (scissors) (both be described with the same verb, *schneiden* and *cut* respectively), Dutch had

two distinct clusters because it makes an obligatory distinction between single-blade cutting (*snijden*) and double-blade cutting (*knippen*). Events of placement also show variation across these closely related languages, but here German and Dutch pattern together using caused posture verbs (e.g., *to lay, to stand*), rather than a general verb like English *put* (Kopecka & Narasimhan, 2012). Similarities uncovered between these three languages—which diverged around approximately 2000 years ago (Atkinson & Gray, 2006; Bouckaert et al., 2012)—would shed light on the stability of the Instrument category.

2. Study 1

2.1. Participants

One hundred and eighteen adults participated (43 British English speakers: $N_{female} = 30$, $N_{male} = 13$; age $M = 22$, range = 18–56; 39 Dutch speakers: $N_{female} = 24$, $N_{male} = 15$, age $M = 27$, range = 18–61; 36 German speakers: $N_{female} = 19$, $N_{male} = 17$, age $M = 22$, range = 19–26). An additional four participants were tested but excluded for being native speakers of American English. Dutch and German speakers were tested at Radboud University in the Netherlands. English speakers were tested at Radboud University and at University of York in England. Participants received either course credit or £5/€5 for participating. Informed consent was obtained for participants in Study 1 and Study 2.

2.2. Design and materials

Participants described five videos from each of 10 conditions. The Tool condition featured events of canonical tool use. In an additional seven conditions, one of the event participants had instrumental properties but was not a canonical tool. In two final conditions, the events were noninstrumental; they instead featured transfer and caused motion. These 10 conditions are shown in Table 1, along with example events. Each video included a Target object which is underlined in the example videos in Table 1. See Appendix A for brief descriptions of all videos, which can also be accessed at https://osf.io/3r28k/.

Events were live-action videos filmed in the lab and around campus, with the exception of events in the Means of Transit condition. For this condition, the aim was to elicit descriptions where mode of transit (e.g., train, bicycle) was construed as a means of getting from one place to another. This construal is difficult to access if participants only see a live-action event of someone riding on a train, for example. We therefore showed a video of someone planning a trip on Google Maps, with screen capture showing someone typing in a starting point, then a destination, then a means of travel (e.g., walking, driving).

As a validation of the videos that we selected for each condition, we separately collected intention norms and change norms for all videos except Means of Transit videos. For the intention norming task, 17 adult English speakers on Amazon Mechanical Turk rated the degree of intentionality of the agent on a 1–5 scale (e.g., for the Tool video in Table 1, "To what extent did the woman choose to smash the plate?"). For the change norming task, 18 adult English speakers on Amazon Mechanical Turk rated on a 1–5 scale whether the patient

Fig. 1. Initial freeze-frame image for the video of a woman wiping a table with a cloth (Contact condition).

underwent a change (e.g., for the Tool video in Table 1, "To what extent did the plate change as a result of the action?"). Mean ratings are available in Appendix A.

We conducted a separate online task with 25 British English speakers where people typed a description of what they saw happening in the videos without being prompted to mention any of the entities in the video. When speakers described instrumental events, they often omitted the Target (e.g., describing an event of a man cutting bread with a knife as *a man was cutting some bread*) (see Grigoroglou & Papafragou, 2019).[2] Given this, in Study 1 we highlighted the event participants of interest by placing red circles around them, except for Means of Transit events. Prior to each video, there was a two second freeze-frame in which three event participants were circled in red. The circles then disappeared and the video began. Fig. 1 shows the freeze-frame image for the video of a woman wiping a table with a cloth. In the Put-Theme and Give-Theme events, the circles were around the source, goal, and theme. In all other events, the red circles were around the agent, patient, and instrument.[3]

2.3. Procedure

Participants viewed each of the 50 events on a computer screen and described the events out loud, and descriptions were video- and audio-recorded for later transcription and coding. Participants were told that they could describe the videos however they liked so long as they mentioned the objects and individuals in red circles. Each participant saw the events in a unique random order. We gave participants four practice videos to familiarize them with task requirements, particularly the use of red circles, and the different format of the Means of Transit events. If a participant failed to mention one of the objects in red circles during a practice video, they were corrected and given another opportunity to describe the video. For

the Means of Transit events, participants were told that they would see someone planning a trip on Google Maps, and they should describe the trip as if they took it themselves, as if it was something that actually happened.

2.4. Coding

The English, Dutch, and German descriptions were transcribed by native speakers. For each description, we coded which lexical item introduced the Target DP as its syntactic complement. We refer to this lexical item as the "term." For example, one participant described the Tool video in Table 1 saying: *the woman smashed the plate using a hammer*. Here the term is *using*.[4] When the Target was expressed through a denominal verb such as *to bike* or *to paint*, we also coded the verb as a "term." In all three languages, verb particle constructions (e.g., *hold out*, *knock over*) were coded as being the same term as the bare verb (e.g., *hold*, *knock*). Descriptively, the verb particle and bare verb variants do not appear to be distributed differently across the videos (i.e., this distinction does not appear to capture semantic variance in the stimuli).

By their nature, events can be construed and segmented in different ways. Therefore, if a single description from a participant included multiple terms, each term was counted. For example, one participant described the Tool video in Table 1 saying: *she picks up the hammer and smashes the plate with the hammer*. Here both *pick up* and *with* were used to introduce the Target (the hammer) and we included both terms in our analysis. Including multiple mentions of the Target allowed us to capture whether some events are more variable than others in how they are segmented, and whether this variability itself varies across languages. At the same time, we did not code terms that only conveyed the presence of the Target but were not used to describe a component of the main action itself (for example, in English: *be, have, see*; in Dutch: *hebben, kijken*; in German: *sehen*). For example, one participant described the Tool video in Table 1 saying: *I see a woman, a hammer and a plate. The woman smashes the plate with the hammer*. For this trial, the only coded term is *with*. In addition, we did not code terms where the Target was mentioned in an action that was distinct from the main action. For example, for the Accidental Agent video in Table 1, our purpose was to test how speakers would describe the Target in the accidental segment of the event, where the woman knocks over the bottle. So the description *a woman used the brush to sweep and accidentally knocked over a bottle* was coded as lacking mention of the Target, because the broom was not mentioned as part of the accidental event. Put another way, it would be misleading to code *use* as the term for this description, as *use* characterizes the role of the Target in the intentional segment of the event, not the accidental segment.

We excluded from analysis trials where: (1) the Target was not mentioned, or (2) the Target only appeared as the Subject of a sentence, as in *the wooden train tracks fell over* (here the Target is one of the train tracks). We excluded 7% ($n = 167$) of English, 8% ($n = 168$) of Dutch, and 10% ($n = 197$) of German trials for these reasons. Appendix B lists the percentage of included trials for each video for each language.

Table 2
Distributions of most frequent terms in English, Dutch and German

English terms	proportion of total	Dutch terms	English gloss	proportion of total	German terms	English gloss	proportion of total
with	0.175	met	*with*	0.370	mit	*with*	0.409
use	0.110	op	*on*	0.083	auf	*on*	0.103
put	0.089	leggen	*put*	0.073	nehmen	*take*	0.074
pick up	0.070	pakken	*take*	0.073	legen	*put*	0.049
on	0.063	geven	*give*	0.058	werfen	*throw*	0.034
take	0.049	tegen	*against*	0.057	in	*in*	0.033
place	0.046	gooien	*throw*	0.032	geben	*give*	0.031
hit	0.040	stoppen	*put*	0.025	packen	*grab*	0.024
knock	0.038	in	*in*	0.024	stellen	*put*	0.024
using	0.036	zetten	*put*	0.023	gegen	*against*	0.023
drop	0.034	doen	*do*	0.021	benutzen	*use*	0.020
in	0.029	laten vallen	*drop*	0.016	stoßen	*push*	0.020
pass	0.029	gebruiken	*use*	0.016	reichen	*hand*	0.019
by	0.024	rollen	*roll*	0.015	fallenlassen	*drop*	0.016
into	0.024	opklimmen	*climb up*	0.013	tun	*do*	0.013
throw	0.021	in de hand hebben	*hold in the hand*	0.012	mit Hilfe	*with help*	0.012
hand	0.016	omstoten	*knock over*	0.010			
onto	0.016						
hold	0.015						
climb	0.013						
give	0.011						

2.5. Results

2.5.1. Clustering analysis

For each language, we first analyzed the similarity structure of the events by clustering the 50 videos according to how often the same terms were used to describe each video. As is typical of naming studies, responses showed a Zipfian distribution with a few high frequency terms and a long tail of low frequency terms (overall across videos, English speakers produced 86 terms, Dutch 95 terms, and German 103 terms). Since the purpose of this initial analysis was to quantify the similarity of videos relative to each other, we excluded terms used for only a single video. Some singleton terms were produced only once across the entire dataset, whereas others were common but appeared only for a single video (e.g., English *fly*, Dutch *vliegen*, and German *fliegen* for the event of someone flying from Rome to Moscow). After singleton terms were excluded, the English data contained 39 terms (85% of dataset where a term was used to introduce the Target), Dutch 43 terms (90% of dataset) and German 44 terms (86% of dataset). We return to singleton terms in Section 2.5.2. Table 2 shows the most common terms for each language (each term accounting for at least 1% of the dataset). Note that these terms are a subset of the terms analyzed in the clustering analysis.

We constructed a video-by-term matrix for each language such that each cell indicated the frequency with which a term was used to describe a video. Data were analyzed using Hierarchical Clustering on Principal Components (HCPC) through the *FactoMineR* package (Husson, Lê, & Pagès, 2017; Lê, Josse, & Husson, 2008) in R (R Core Team, 2022). We first conducted a correspondence analysis on each language matrix (Greenacre, 2007) using the function *CA()*. Correspondence analysis is a dual factoring technique that takes the original matrix and produces a reduced number of dimensions that capture a substantial proportion of variance in the data. We then submitted the correspondence analysis for each language to agglomerative hierarchical clustering on principal components using the function *HCPC()* using Ward's method for identifying optimal clusters. In this procedure, videos that load most highly on dimensions capturing the most variance in the correspondence analysis are clustered most distinctly in the clustering analysis. HCPC is particularly useful for visualizing data where a large number of dimensions in the correspondence analysis are needed to adequately account for variability in the data. This was the case with our dataset: the 10 dimensions with the highest eigenvalues in the correspondence analysis accounted for 81% of total variance in the English data, 82% for Dutch, and 83% for German. Whereas plotting the correspondence analysis dimensions only depicts similarity for two dimensions at a time, the clustering visualization depicts overall similarity (see Figs. 2, 3, and 4). Three hierarchical trees were produced, one for each language. In each tree, leaves correspond to videos, referred to in shorthand and indicated with brackets, for example, [tool_break].

Finally, we compared the three language trees (i.e., dendrograms) to one another using the *dendextend* package for R and the *tanglegram()* function (Galili, 2015). Three comparisons are shown: English versus Dutch (Fig. 2), English versus German (Fig. 3), and Dutch versus German (Fig. 4). Colors indicate which videos cluster together in the HCPC analysis for each language. Critically, colors should be interpreted language internally. In Fig. 2, for example, the purple clusters in English versus Dutch are assigned arbitrarily—what matters is that events are coded the same color within a language dendrogram. In each figure, videos are linked across dendrograms by a solid line: lines that have a steeper angle indicate videos that cluster differently across languages.

In hierarchical dendrograms, the order of branches can be rearranged at junctions, just like a baby cot mobile, and still preserve the critical relational information between nodes. We used the functions *entanglement()* and *untangle()* in the *dendextend* package to identify trees where the arrangement of branches minimized the discrepancy between each language pair, resulting in straighter lines in Figs. 2–4 where such correspondences could be found across the trees. *Entanglement()* calculates the degree of crossover for a pair of trees; *untangle()* uses multiple methods to find alternate trees that reduce the entanglement value. For each pair of trees, we used the method that resulted in the lowest entanglement value (this was the "step2side" method).

Figs. 2–4 revealed broad similarities across languages, as well as places where the languages were semantically distinct. First, each language is characterized by a large cluster composed primarily of Tool, Contact, and Body Part videos (we will refer to this cluster as the "central cluster"). In each language, speakers largely used the same terms to describe these three types of events. In English, the most frequently used terms for central cluster

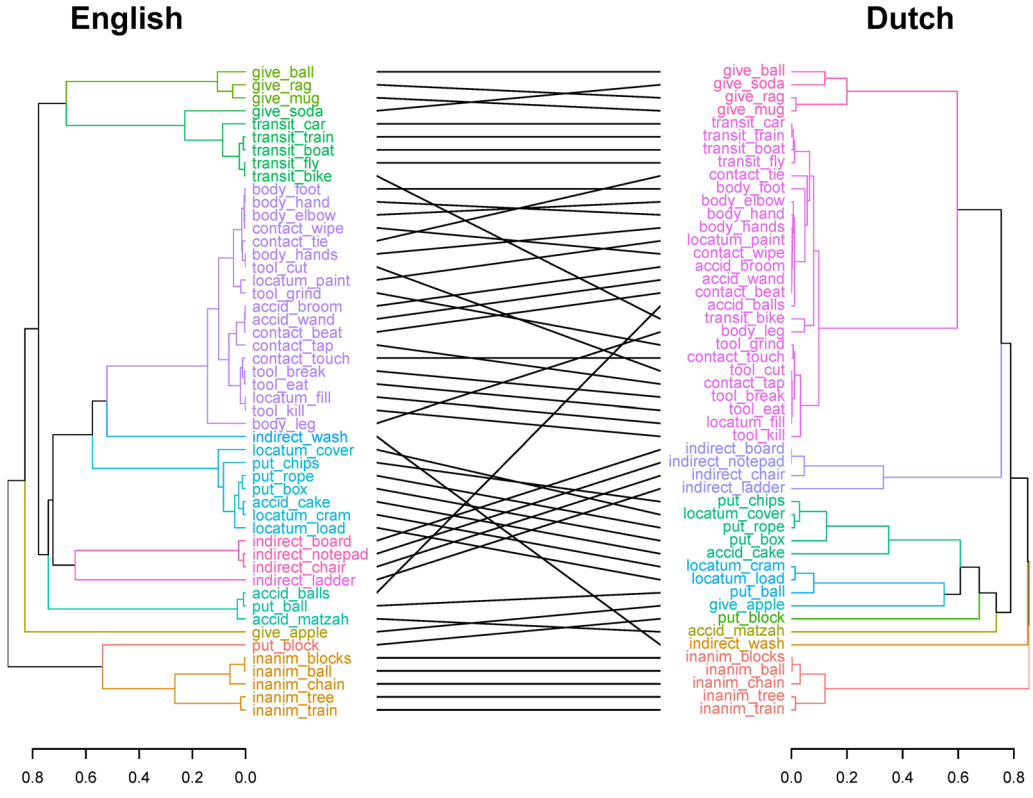

Fig. 2. Dendrogram correspondence between English and Dutch. Straighter lines indicate closer correspondences between the position of each video acrossdendrograms.

videos were (in descending order): *with, use, pick up*, and *using*; in Dutch these were *met* ('with') and *pakken* ('grab'), and in German, *mit* ('with') and *nehmen* ('take').

A second similarity across the three languages was that Inanimate Agent events were the most distant from the central cluster in all three languages. We included these events in our study to test whether being a causal intermediary is characteristic of Proto-Instruments. Our results suggest that this event property on its own is not sufficient for Instrumenthood, as the Inanimate Agent events were more distant from the central cluster than the Put-Theme and Give-Theme events. This suggests that Targets in the Inanimate Agent events are not even atypical Instruments. Apparently, being acted on by an agent is more a defining property of Instrumenthood then causal directness, at least in English, Dutch, and German.

Cross-linguistic differences emerged in the six other event conditions and were most pronounced for the Indirect Action and Means of Transit conditions. In English and Dutch, Indirect Action events occupied an intermediate position outside the central cluster, but not as distant from the central cluster as Inanimate Agent events. For example, English participants sometimes used *with, use*, and *using* to describe Indirect Action videos (e.g., *a woman uses a chair to grab a plant pot out of her reach;* [indirect_chair]). By contrast, in German, Indirect

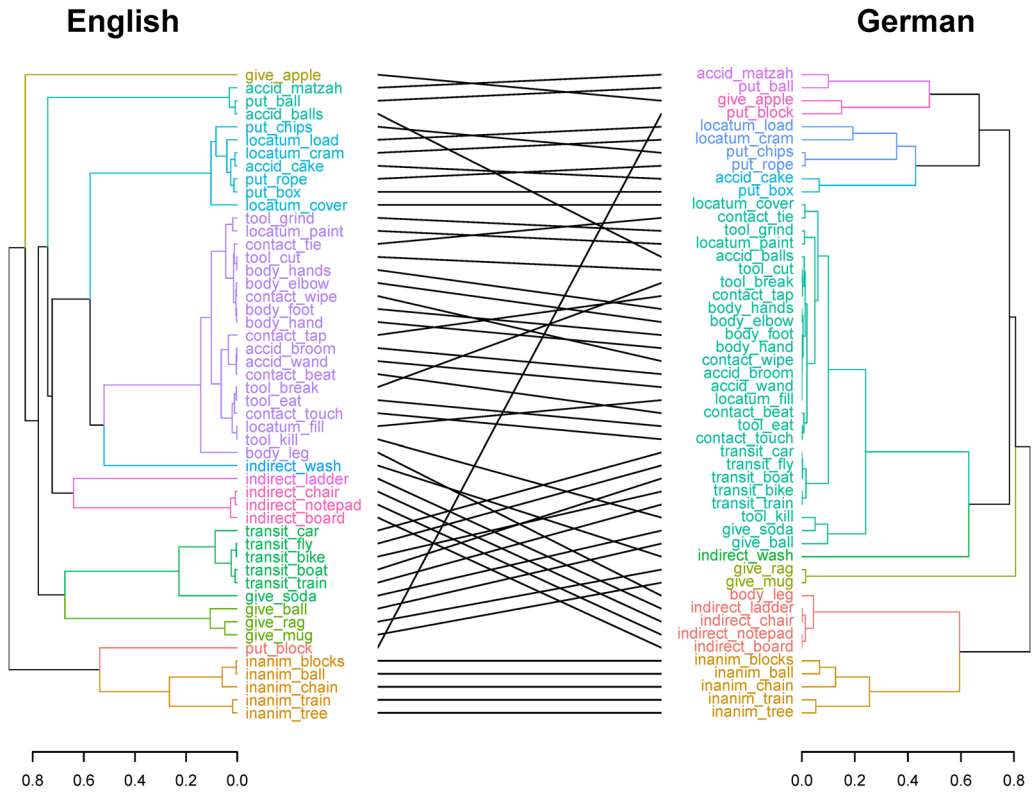

Fig. 3. Dendrogram correspondence between English and German. Straighter lines indicate closer correspondences between the position of each video across dendrograms.

Action events were relatively distant from the central cluster. Another cross-linguistic difference emerged for Means of Transit videos: these were part of the central cluster in Dutch and German, but not in English. Dutch *met* and German *mit* were frequently used to describe the means of transit, as in the Dutch response *Ik ga met een ferry van Port Angeles naar Victoria* ('I go by ferry from Port Angeles to Victoria'; [transit_boat]). English responses, by contrast, never featured *with* and rarely featured *using* and *use* for Means of Transit videos.

As a group, the five Accidental Agent videos did not cluster together. In all three languages, the videos of a man accidentally knocking over a music stand with a wand [accid_wand] and a woman accidentally knocking over a plastic bottle with a broom [accid_broom] were in the central cluster. As described in Section 1.2, each of these videos had an intentional phase (e.g., sweeping the floor) and an accidental phase (e.g., knocking over the bottle). We only coded terms that described the role of the instrument in the accidental phase, so it is notable that these two Accidental Agent videos clustered with Tool videos.

The clustering analysis suggests that in all three languages, there is an Instrument prototype with a shared central focus, although there were cross-linguistic differences at the periphery. Most of the Tool, Contact, and Body Part videos were in a central cluster in all

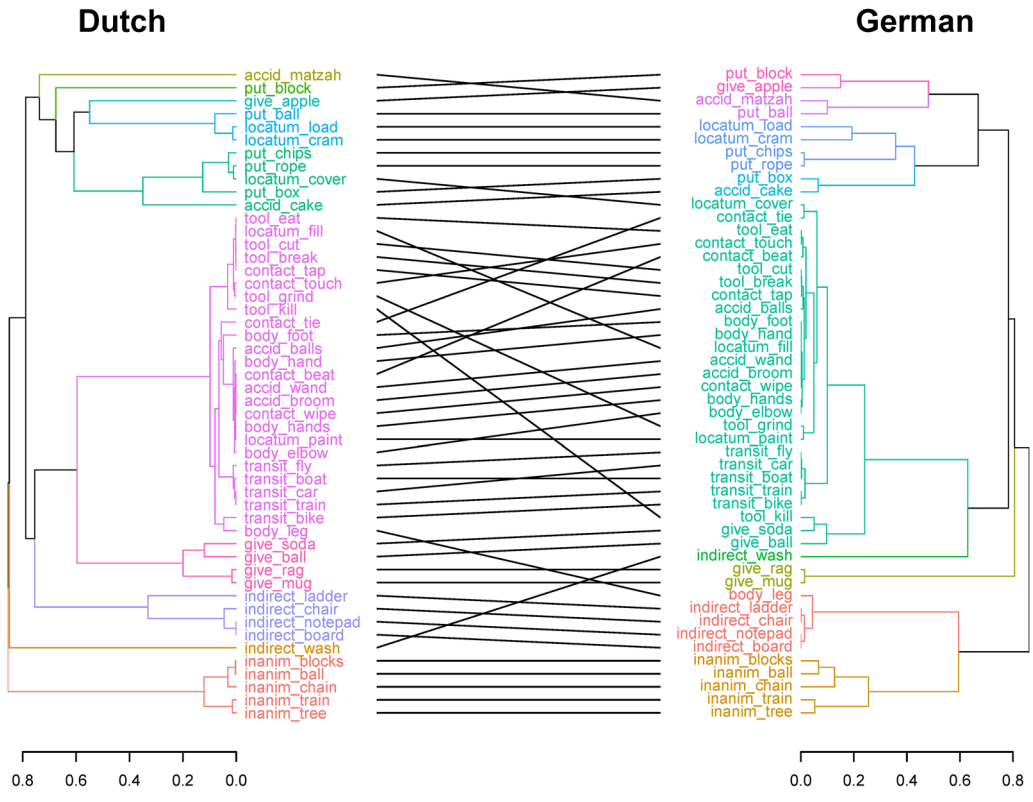

Fig. 4. Dendrogram correspondence between Dutch and German. Straighter lines indicate closer correspondences between the position of each video across dendrograms.

three languages—this cluster likely reflects prototypical qualities of Instrumenthood. While the tool-prototype hypothesis predicts the Tool condition is most prototypical, the results from how people describe events suggests that the Instrument prototype is in fact broader than Tool alone.

2.5.2. Central cluster naming agreement

If this central cluster of videos reflects a prototype across English, Dutch, and German, then this could be because (1) some central cluster videos are more prototypical than others but the three languages agree regarding the typicality structure across videos, or (2) the central cluster videos are all equally prototypical. The tool-prototype hypothesis predicts the first situation, specifically that Tool videos will be most prototypical. In the analysis that follows, we used term diversity as a diagnostic of prototypicality. As described in Section 1.2, we assume that events described using more consistent language are more prototypical. We used Simpson's diversity index (D) to calculate term diversity for each video in each language (Simpson, 1949). This metric takes into account both number of term types and overall number of tokens per video and has been used in previous comparative linguistic studies (Kim, Elli, & Bedny,

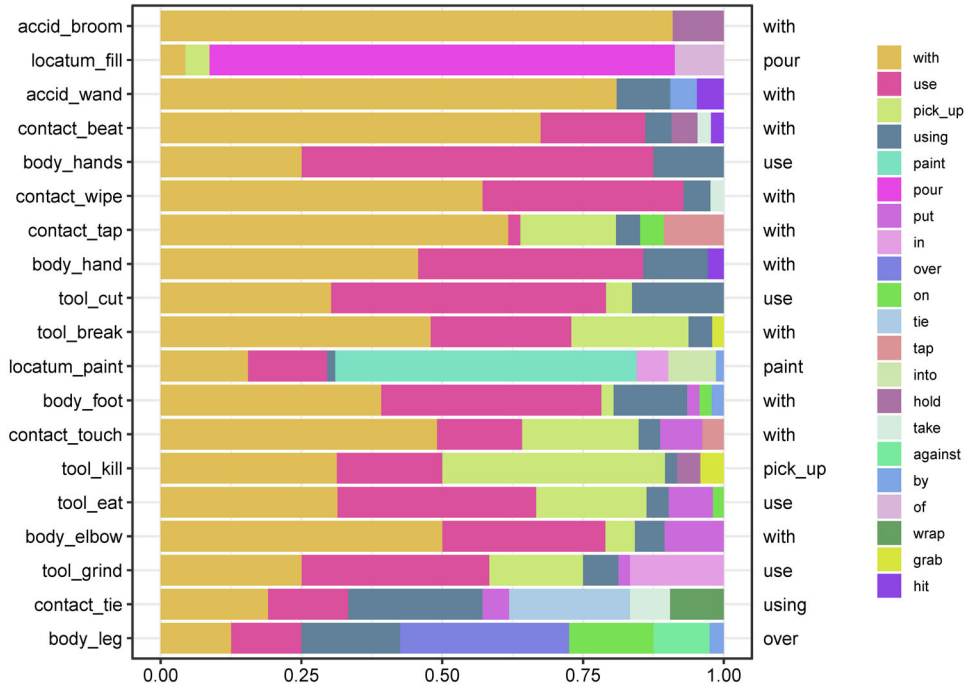

Fig. 5. Distribution of terms across central cluster videos in English. The modal term for each video is listed to the right of the plot. Videos are ordered by *D*-value, from highest (most naming agreement) to lowest (least naming agreement). Terms are ordered in the legend from most to least frequent across all central cluster videos. Terms that were produced only once or twice in total are not shown.

2019; Majid & Burenhult, 2014; Majid et al., 2007; Majid et al., 2018). *D*-values range from 0 to 1, where 0 corresponds to low naming agreement (i.e., no consistency in video descriptions) and 1 corresponds to high naming agreement. For this analysis, singleton terms were included to capture the full range of responses.

Figs. 5–7 show the distribution of terms for the central cluster in each language. It is apparent that once singleton terms are included, some videos have markedly different term distributions from other central cluster videos. In all three languages, speakers predominantly used singleton terms to describe [locatum_fill] and [locatum_paint]. German speakers overwhelmingly used singleton terms for [give_ball] and never used *mit* for [give_soda]. In addition, two of the Means of Transit videos elicited modality-specific terms in Dutch (for [transit_bike]: *fietsen* "bike") and in German (for [transit_fly]: *fliegen* "fly"). These videos suggest within-cluster variability not captured in the earlier analyses.

Table 3 shows the average *D*-values for central cluster videos in each condition in each language. We predicted that if Tool events are most prototypical, *D*-values would be highest for Tool videos. We found instead that in all three languages, naming agreement was comparable between Tool and Contact videos and was higher for Accidental Agent videos. In Dutch and German, Body Part videos had higher naming agreement than Tool videos. As there were

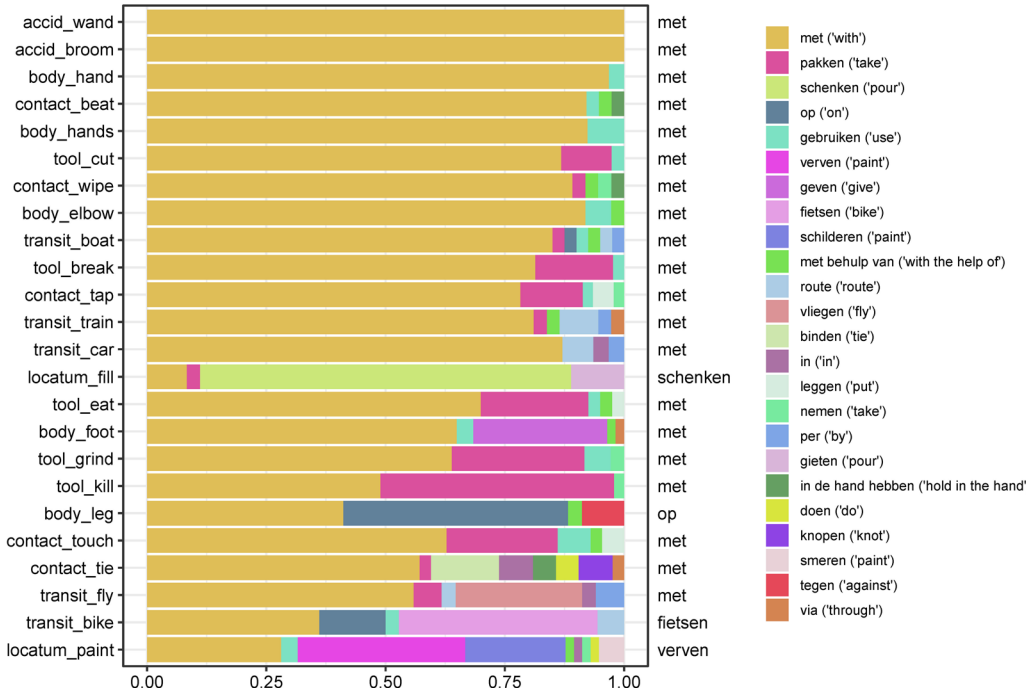

Fig. 6. Distribution of terms across central cluster videos in Dutch. The modal term for each video is listed to the right of the plot. Videos are ordered by D-value, from highest (most naming agreement) to lowest (least naming agreement). Terms are ordered in the legend from most to least frequent across all central cluster videos. Terms that were produced only once or twice in total are not shown.

five or fewer videos in each condition in Table 3, we were not able to use inferential statistics to analyze whether the conditions vary with respect to naming agreement. Nonetheless, descriptively we did not find that naming agreement was highest among Tool videos.

To formally test the similarity across languages, we calculated the correlations between the *D*-values for the 18 videos[5] that were part of the central cluster in all three languages and found the languages were positively correlated: English and Dutch, $r(16) = .66$, $p < .01$, Dutch and German, $r(16) = .80$, $p < .001$). The correlation between English and German was marginally significant: $r(16) = .41$, $p = .09$. This suggests that similar features of event representation influenced diversity of linguistic encoding in the three languages, although English and German are most dissimilar.

Counter to our predictions, naming agreement among central cluster videos was not highest for Tool videos. As a further test of the tool-prototype hypothesis, we analyzed the use of "core" instrumental terms across central cluster videos. As described earlier (Section 2.4), we coded multiple terms when participants mentioned the Target in multiple ways (e.g., *the woman picked up the hammer and broke the plate with it*). Figs. 5–7 show terms such as *pick up* and German *nehmen* ("take"), which encode instrumentality only indirectly, were used for some videos more often than others, leading to different *D*-values. Figs. 5–7 do not show,

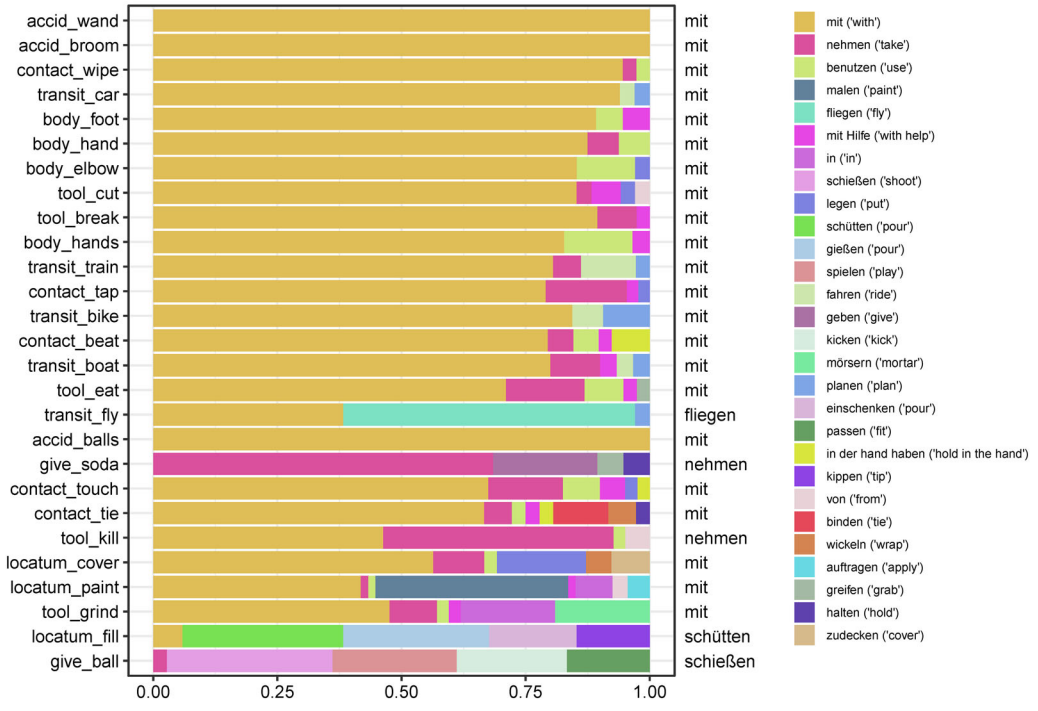

Fig. 7. Distribution of terms across central cluster videos in German. The modal term for each video is listed to the right of the plot. Videos are ordered by D-value, from highest (most naming agreement) to lowest (least naming agreement). Terms are ordered in the legend from most to least frequent across all central cluster videos. Terms that were produced only once or twice in total are not shown.

however, the proportion of descriptions for which indirect-instrumental terms such as *pick up* were the only way the Target was mentioned. It is possible that when only core instrumental terms are considered, Tool videos are more likely to elicit these core terms than other videos. We operationalized core instrumental terms as the dominant terms used for the central cluster: English *with, use* and *using*, Dutch *met*, and German *mit*. To maintain comparability across the three languages, we also treated as core instrumental terms the translational equivalents of *use* in Dutch and German, *gebruiken* and *benutzen*. Dutch and German lack an equivalent of *using* in this data (see Table 2).

For the central cluster videos in each language, we analyzed whether or not speakers produced a core instrumental term on each trial using mixed-effects logistic regression and the *lme4* package for R (Bates, Maechler, Bolker, & Walker, 2014). Models included Subject and Video random intercepts and the fixed effect Condition, with Tool as the reference level of Condition (models with by-subject random slopes failed to converge). In English and Dutch, core instrumental terms were used less often in the Locatum condition than the Tool condition (English Locatum: $\beta = -3.43$, $SE = 1.21$, $p < .01$; Dutch Locatum: $\beta = -2.88$, $SE = .98$, $p < .01$). The Tool condition did not differ significantly from any other conditions in Table 3. In German, core instrumental terms were used less often in the Locatum than Tool condition

Table 3
Mean *D*-values and mean proportion of trials where a core instrumental term was used by condition for central cluster videos in each language

English			Dutch			German		
Condition	*D*	Proportion of Trials with Core Terms	Condition	*D*	Proportion of Trials with Core Terms	Condition	*D*	Proportion of Trials with Core Terms
Tool	0.27	0.81	Tool	0.55	0.78	Tool	0.50	0.77
Contact	0.32	0.82	Contact	0.50	0.84	Contact	0.54	0.89
Body	0.25	0.81	Body	0.60	0.83	Body	0.75	0.95
Accidental	0.71	0.94	Accidental	1.00	1.00	Accidental	0.69	0.93
Transit	NA	NA	Transit	0.45	0.65	Transit	0.62	0.72
Locatum	0.31	0.29	Locatum	0.24	0.27	Locatum	0.31	0.50

Note. Means of Transit videos were not in the central cluster in English.

($\beta = -1.76$, $SE = 0.85$, $p < .05$), and were used more often in the Body Part than Tool condition ($\beta = 2.04$, $SE = 0.88$, $p < .05$). No other contrasts were significantly different. These results show that, once singleton terms have been included, Locatum videos in the central cluster were less likely to be described with core instrumental terms than Tool videos. Nonetheless, we did not find support for the tool-prototype hypothesis. Tool videos were not more likely than central cluster videos in the Contact, Body Part, Accidental Agent, and Means of Transit conditions to elicit core instrumental terms.

2.6. Discussion

In Study 1, we found strong evidence for a shared semantic space influencing linguistic encoding across English, Dutch, and German. Each language clustered the videos in similar ways. Specifically, Contact and Body Part events clustered with Tool events in all three languages. Inanimate Agent events were more dissimilar from Tool events than Put-Theme and Give-Theme events were from Tool events in all three languages. At the same time, there were some differences between languages too: Indirect Action events were more dissimilar from Tool events in German than in English and Dutch. In addition, Means of Transit events were in the central cluster in Dutch and German but not in English.

As far as diagnostics of prototypicality, however, our analysis of the central cluster videos did not provide evidence for the tool-prototype hypothesis. Naming agreement among central cluster videos was not numerically highest for videos in the Tool condition. In addition, Tool events were not most likely to elicit core instrumental terms. Across the central cluster videos, no one event condition emerges as a clear prototype. As far as the high naming agreement for Accidental Agent events, two factors are worth noting: first, speakers often failed to mention the role of the Target in the accidental component of the event (see Appendix B), suggesting the Target had low prominence in speakers' representations of accidental events. Second, Accidental Agent events had high naming agreement in part

because *use/using/gebruiken/benutzen* are infelicitous for these videos and were never produced. Naming agreement should, therefore, not be taken on its own to indicate that Accidental Agent events are prototypical. This has to be contextualized with respect to the actual descriptions given.

Language production results provide one window on category structure within and across languages. In Study 2, we report a complementary source of evidence: participants' judgments about whether a particular instrumental sentence is a felicitous description of an event. We tested sentences with the preposition *with* and the periphrastic verb *use* and their translational equivalents in Dutch and German because these terms were frequently produced by participants in Study 1 and these terms have been argued to specifically lexicalize instrumental meaning.

3. Study 2

3.1. Participants

One hundred and eighty-seven adults participated (61 British English speakers: $N_{female} = 33$, $N_{male} = 28$, age $M = 33$, range = 18–82; 64 Dutch speakers: $N_{female} = 51$, $N_{male} = 13$, age $M = 24$, range = 18–66; and 62 German speakers: $N_{female} = 57$, $N_{male} = 5$, age $M = 21$, range = 18 – 31). An additional 17 adults were tested but were excluded either due to failure on control trials (six English, three Dutch and six German speakers) or problems with video playback (one English and one Dutch speaker). We tested roughly double the number of participants as in Study 1 because in Study 2, each participant saw only half of the full set of video-sentence pairs (see Section 3.2). British English speakers were tested on the crowdsourcing platform Prolific. Dutch and German speakers were recruited at Radboud University in the Netherlands. Participants received either course credit or £5/€5 for participating.

3.2. Design and materials

Participants viewed each of the 50 videos from Study 1. Each video was paired with a single written sentence in English, Dutch or German and participants were asked to judge on a 6-point scale how natural the sentence was as a description of the video (see Simms, Zelazny, Williams, & Bernstein, 2019 for an empirical assessment of the benefits of 5- vs. 6- vs. 7-point scales). We tested two types of syntactic frames: a prepositional frame and a periphrastic verb frame. Across languages, the prepositional frame sentences featured *with* (English), *met* (Dutch), or *mit* (German). The periphrastic verb sentences featured *use* (English), *gebruiken* (Dutch) or *benutzen* (German). With respect to frame type, sentences were presented in a within-subject Latin-square design, with participants judging 25 prepositional sentences and 25 periphrastic verb sentences. Frame type was balanced across conditions.

In addition to these 50 video-sentence pairs, participants also completed eight positive control and eight negative control trials. To be included in the study, participants needed to distinguish the positive and negative control trials by at least three points on the 6-point

Table 4
Syntactic frames and example sentences used in Study 2

Prepositional Frame		**Periphrastic Frame**
example video: [tool_cut]		example video: [indirect_board]
A N1 V+INFL *the* N2 *with a* N3 A woman sliced the baguette with a knife.	**English**	*A* N1 *used a* N3 *to* V *the* N2 A woman used a chopping board to cut up the carrot.
Een N1 V+INFL *de* N2 *met een* N3 Een vrouw snijdt het stokbrood met een mes.	**Dutch**	*Een* N1 *gebruikt een* N3 *om de/het* N2 *te* V+INFL Een vrouw gebruikt een snijplank om de wortel te snijden.
Ein(e) N1 V+INFL *der/die/das* N2 *mit einem/einer* N3 Eine Frau scheidet das Baguette mit dem Messer.	**German**	*Ein(e)* N1 *benutzt ein/eine/einen* N3 *um ein/eine/einen* N2 *zu* V+INFL Eine Frau benutzt ein Schneidebrett, um eine Karotte zu schneiden.

scale. The function of these control trials was to assess participants were using the scale appropriately and not responding at random.

Each frame in each language had a default syntactic form. Table 4 shows the prepositional frames with example sentences from the [tool_cut] video and the periphrastic frames with example sentences from the [indirect_board] video. The Target was mentioned in the N3 slot. N1 corresponded to the agent and N2 to the patient/goal, depending on the video.

We selected open-class words in the NPs by choosing frequent nouns from descriptions speakers produced in Study 1. For example, for the video of a woman wiping a table, English-speaking participants used the lexical item *cloth* more often than *rag*. When only the hand of the agent was visible in the video, N1 was *someone/iemand/eine person*. For some videos, the sentences had different determiners than in the standard syntactic form in Table 4. When the Target was a body part, we used the possessive determiner, for example *a man smashed the cupcake with his elbow*. For mass nouns, we used the bare plural form, for example, *socks, orange juice, blue paint, rope*. Parallel adjustments were made in Dutch and German. For a few Dutch sentences, a preposition was inserted before NP2 to improve the felicity of the sentence: *Een man schopt naar een andere man met een bal*, "a man kicks to another man with a ball."

We selected the verbs for each sentence in a two-step process. First, if speakers in Study 1 ever used a *with/met/mit* or *use/gebruiken/benutzen* description for a particular video, we selected a verb that was frequently used in those descriptions. For example, for the video of a woman putting a towel on a doll, English speakers sometimes produced *the woman covered the baby doll with the towel*; so *cover* was the verb for this video, even though *put* and *place* were more frequently produced for this video. Second, if *with/use* descriptions were never produced for a video, we selected a verb that made the sentence as felicitous as possible as a description of the video. For example, for the video of a man putting a box on a shelf, the sentence *a man put the shelf with a box* is fairly unacceptable in English. We therefore chose

a more felicitous English verb for this video, *load*, that is, *a man loaded the shelf with a box*. We used the list of English verbs in Levin (1993) to help select English verbs. We also favored basic level verbs over more specific verbs (e.g., *fill* over *infuse*).

We used sentences with *on/op/auf* for the control trials. For example, for a video of a woman putting a mug on a table, the sentence *a woman put a mug on the table* was a positive English control. By contrast, for a video of someone taking chips out of a bag, the sentence *someone put the crisps on the bag* was a negative English control. The same videos were used for positive and negative control trials across the three languages. Including the control trials, participants were tested on 66 video-sentence pairs. These pairs were presented in a random order. Sentence stimuli may be accessed at https://osf.io/3r28k/.

3.3. Procedure

Participants completed the study online. They were instructed they would watch short video clips and read sentences describing videos. Their task was to judge how natural the sentence was as a description of what was happening in the video. Participants were given examples of natural and unnatural video descriptions prior to the experimental trials. Participants also received specialized instruction and an example for the Google Maps videos (Means of Transit condition).

In each trial, a video and a sentence description were displayed at the same time. There was a one-second lag before the video began, allowing participants to read the sentences. The sentence remained visible after the video completed playing. Participants then used the numbers 1 through 6 on the keyboard to rate the naturalness of the match between the video and the sentence. We showed them a visual to help them remember the directionality of the scale. The endpoints of the scale were labeled *completely unnatural/completely natural* in English, *helemaal niet natuurlijk/helemaal natuurlijk* in Dutch, and *vollkommen unnatürlich/vollkommen natürlich* in German. Participants had the option of pressing the 0 key on the keyboard if they had problems with video playback.

3.4. Results

3.4.1. Acceptability across conditions

Naturalness ratings spanned the full range of the scale in each language, providing evidence on which events correspond to the core uses of each term. The three languages were in strong agreement as to which videos were more or less compatible with prepositional and periphrastic terms (see Fig. 8). For prepositional terms, the average ratings across all 50 videos were significantly positively correlated (English *with* and Dutch *met,* $r(48) = .78$, $p < .001$, English *with* and German *mit*, $r(48) = .73$, $p < .001$, Dutch *met* and German *mit*, $r(48) = .95$, $p < .001$). Cross-language correlations were also high for the periphrastic terms: English *use* and Dutch *gebruiken,* $r(48) = .89$, $p < .001$, English *use* and German *benutzen,* $r(48) = .84$, $p < .001$, Dutch *gebruiken* and German *benutzen,* $r(48) = .95$, $p < .001$). This is consistent with Study 1 suggesting a shared semantic space across languages.

The tool-prototype hypothesis predicts that naturalness ratings will be highest for the Tool videos. To test this, we conducted a mixed-effects ordinal regression analysis using the *brms* package for R (Bürkner, 2017, 2018), which uses Bayesian statistics to estimate the thresholds

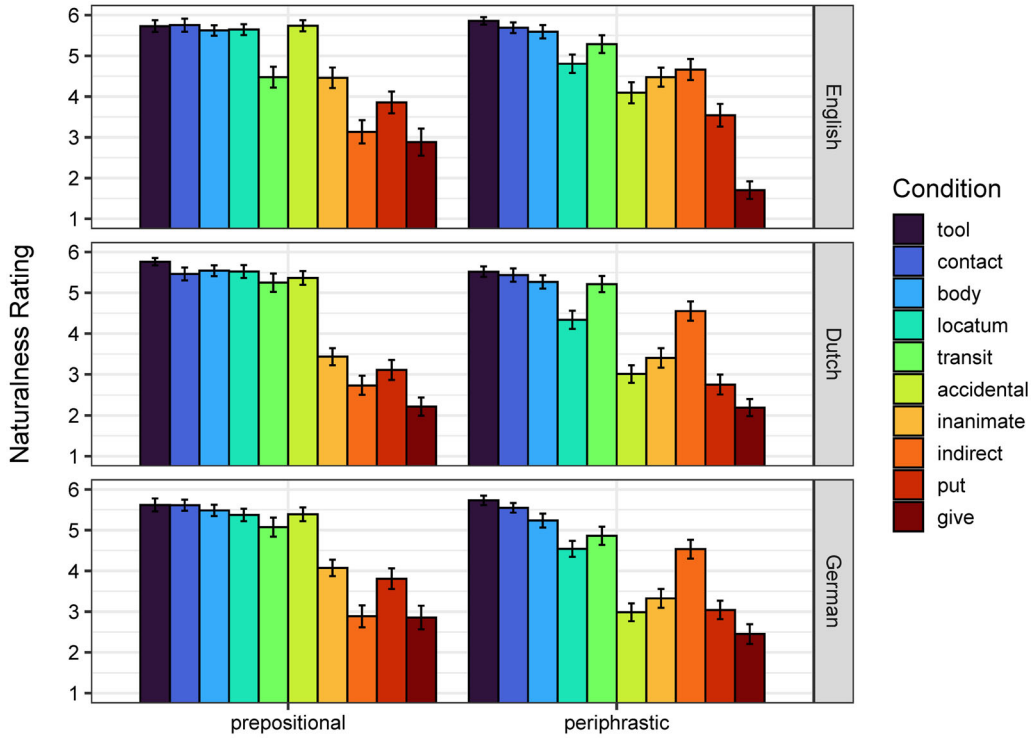

Fig. 8. Mean naturalness ratings for each condition,sentence frame, and language (prepositional terms: English *with*, Dutch *met*, German *mit*; periphrastic terms: English *use*, Dutch *gebruiken*, German *benutzen*). Error bars show 95% confidence intervals of the mean.

for each of the levels of an ordinal variable. Models included random intercepts for Subject and Video and the fixed effect Condition, with Tool as the reference level. We fit models for each of the six terms tested (three prepositional and three periphrastic). Table 5(a and b) shows the coefficient estimates and 95% confidence intervals for each of the individual conditions relative to Tool for each term. Conditions where the confidence interval does not include zero are in bold, demonstrating ratings in this condition were significantly different from Tool.

Table 5(a and b) shows that across languages and terms, ratings for the Tool videos were not significantly higher than ratings for all the other conditions, contrary to the tool-prototype hypothesis. For all six terms, ratings for Contact videos were similar to ratings for Tool videos, and for all terms except German *benutzen*, Body Part ratings were as high as Tool ratings. For the prepositional variants in all three languages, the Locatum and Accidental Agent conditions did not differ significantly from the Tool condition.

This analysis does not take into account the similarity structure revealed by participants' descriptions in Study 1, specifically whether Tool videos have higher ratings than other videos in the central cluster. As a more direct test of the tool-prototype hypothesis, we conducted an additional set of ordinal regressions, including only those videos that were part of the central cluster in each language.[6] For each term, we tested whether ratings were higher for Tool

Table 5
(a and b) coefficient estimates and 95% confidence intervals for each level of the condition fixed effect for each term, with Tool videos as the reference level

	Prepositional Frame					
	English *with*		Dutch *met*		German *mit*	
Reference Level = tool	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
Contact	0.33	−0.70 to 1.39	−0.54	−1.39 to 0.30	−0.13	−0.99 to 0.69
Body	−0.17	−1.13 to 0.83	−0.51	−1.40 to 0.36	−0.41	−1.24 to 0.40
Locatum	−0.12	−1.09 to 0.86	−0.5	−1.37 to 0.37	−0.54	−1.40 to 0.29
Transit	**−1.41**	**−2.38** to **−0.42**	−0.73	−1.56 to 0.17	−0.8	−1.64 to 0.03
Accidental	0.06	−0.91 to 1.07	−0.77	−1.62 to 0.08	−0.49	−1.36 to 0.32
Inanimate	**−1.47**	**−2.49** to **−0.48**	**−2.5**	**−3.38** to **−1.63**	**−1.8**	**−2.64** to **−1.00**
Indirect	**−2.35**	**−3.37** to **−1.38**	**−3.03**	**−3.91** to **−2.22**	**−2.68**	**−3.53** to **−1.84**
Put	**−1.88**	**−2.85** to **−0.87**	**−2.78**	**−3.69** to **−1.95**	**−2**	**−2.86** to **−1.18**
Give	**−2.7**	**−3.69** to **−1.72**	**−3.59**	**−4.47** to **−2.75**	**−2.73**	**−3.57** to **−1.90**
	Periphrastic Frame					
	English *use*		Dutch *gebruiken*		German *benutzen*	
Reference Level = tool	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
Contact	−0.4	−1.28 to 0.49	−0.06	−0.83 to 0.79	−0.38	−1.11 to 0.35
Body	−0.58	−1.48 to 0.31	−0.35	−1.18 to 0.51	**−0.76**	**−1.51** to **−0.03**
Locatum	**−1.57**	**−2.49** to **−0.71**	**−1.26**	**−2.07** to **−0.45**	**−1.55**	**−2.30** to **−0.86**
Transit	**−1.01**	**−1.92** to **−0.09**	−0.38	−1.19 to 0.45	**−1.22**	**−1.97** to **−0.48**
Accidental	**−2.15**	**−3.02** to **−1.28**	**−2.33**	**−3.14** to **−1.45**	**−2.87**	**−3.60** to **−2.14**
Inanimate	**−1.89**	**−2.83** to **−1.03**	**−2.06**	**−2.90** to **−1.26**	**−2.58**	**−3.30** to **−1.86**
Indirect	**−1.64**	**−2.58** to **−0.76**	**−1.1**	**−1.91** to **−0.28**	**−1.61**	**−2.35** to **−0.88**
Put	**−2.43**	**−3.32** to **−1.57**	**−2.56**	**−3.37** to **−1.72**	**−2.84**	**−3.57** to **−2.13**
Give	**−4.13**	**−5.04** to **−3.25**	**−3.17**	**−4.00** to **−2.33**	**−3.45**	**−4.16** to **−2.70**

Note. Bolded entries indicate conditions where the confidence interval does not include zero—these conditions are significantly different from the Tool condition.

videos than non-Tool videos. For German *benutzen*, Tool videos were rated higher than the other central cluster videos ($\beta = 1.17$, 95% CI = .21–2.13). Nonetheless, for the five other terms, ratings for these two groups of videos did not differ significantly (English *with*: $\beta = -.12$, 95% CI = −.83 – .55; Dutch *met*: $\beta = .54$, 95% CI = −.09 – 1.16; German *mit*: $\beta = .36$, 95% CI = −.11 – .85; English *use*: $\beta = .87$, 95% CI = -.19–2.04; Dutch *gebruiken*: $\beta = .44$, 95% CI = −.42–1.26). These results show that among videos that elicited similar descriptions in Study 1, instances of these descriptions in Study 2 were not rated as more natural for Tool videos than the other conditions (with the exception of German *benutzen*-descriptions).

3.5. Discussion

Study 2 demonstrated broad semantic similarity across the three languages: ratings were highly correlated for the three prepositional variants and for the three periphrastic variants. At

the same time, we did not find support for the tool-prototype hypothesis: for each of the six terms, ratings were comparable between Tool videos and videos from at least one other condition. In Section 1.1 we described the English diagnostic than an Instrument can be identified by considering the intersection of acceptable *with*-sentences and acceptable *use*-sentences. This criterion can be extended to the current experiment: prototypical members of the Instrument category will elicit high ratings for both prepositional and periphrastic variants of the test sentences. Assuming this criterion, we found the three languages differed as to which conditions were most prototypical. In English, Tool, Contact and Body Part videos received the highest ratings; by contrast, in Dutch, ratings for the Tool, Contact, Body Part, and Means of Transit conditions were equally high; and in German, ratings for the Tool and Contact conditions were equally high. This suggests that despite broad overall similarity, the characteristics of Proto-Instruments differ in subtle ways across these languages.

Speakers in Study 2 evaluated the prompt descriptions globally, and we assume that all aspects of these descriptions influenced speakers' judgments (e.g., sentence frame, choice of verb, choice of determiner). The descriptions necessarily contained different nouns and verbs across videos, and it is possible that some of the variation in judgments was influenced by specific nouns and verbs (e.g., if the nouns used to label Body Part Targets in German were slightly less felicitous than the nouns used for Body Part Targets in English and Dutch). We think it is unlikely, however, that this can account for the contrasts we see in Table 5(a and b), as we selected nouns and verbs based on what was most common in the descriptions from Study 1.

4. General discussion

Event categories such as Agent, Patient, and Goal have been shown to play an important role in shaping event perception, language learning, and language processing; they have also been argued to be part of core cognition. Many thematic roles beyond these three have been proposed to be part of linguistic and conceptual representation, however, and testing the theoretical relevance of thematic roles requires investigating the structure of a wider range of roles. In this paper, we examined the proposed Instrument role, asking whether this category has prototype structure and if so, what the prototype is. Specifically, we hypothesized the prototypical Instrument is a tool: a physical object which is manipulated intentionally in order to cause a change in some other object or individual. Our results suggest that instrumental events are represented in terms of a conceptual core that is shared across English, Dutch, and German. While this conceptual core includes tool events, it is broader than predicted by the tool-prototype hypothesis. The consistency of the category structure revealed across English, Dutch, and German may indicate that Instrument is a prominent and stable category, despite previous suggestions to the contrary (see Section 1.1). The consistency of the category structure also supports the theoretical approach of analyzing thematic roles in terms of prototypes (see Dowty, 1991)

For semantic domains such as color, it is well understood how languages impose different categories on conceptual-perceptual space, and the causes and consequences of linguistic variability have been extensively investigated (Abbott, Griffiths, & Regier, 2016; Berlin & Kay,

1969; Davidoff, Davies, & Roberson, 1999; Gibson et al., 2017; Heider, 1972; Josserand, Meeussen, Majid, & Dediu, 2021; Regier, Kay, & Cook, 2005; Skelton, Catchpole, Abbott, Bosten, & Franklin, 2017; Winawer et al., 2007). By comparison, little psycholinguistic research has addressed whether thematic role markers in different languages impose different category structures on the space of event participants, or how linguistic categories shed light on underlying conceptual structure. One possible reason for this gap is that event participants are not typically "labeled" through open class words but rather are marked through morphosyntactic devices that are distributed across the utterance. Our study demonstrates an empirical framework for investigating cross-linguistic encoding of thematic role space, opening up new pathways of research on what is variable, what is universal (if anything), and what the cognitive consequences are of cross-linguistic variability in encoding of event participant categories.

4.1. What is an Instrument? Revisited

In previous studies of instrumental semantics, a variety of event features have been proposed to be definitive of Instrumentality—that an Instrument is a causal intermediary in an event, for example (see Table 1). In our studies, Tool, Contact, and Body Part videos emerged as prototypical in all three languages: they clustered together and were likely to be described using core instrumental terms. Instrumental descriptions of these videos also elicited high acceptability ratings in all three languages (both prepositional and periphrastic variants of the descriptions). These results suggest the following characterization of the prototypical Instrument in English, Dutch, and German: as the direct extension of an intentional agent. This finding supports previous accounts in which the Instrument role is inextricably linked to the Agent role (Dowty, 1991; Rissman, Rawlins, & Landau, 2015; Schlesinger, 1989).

Whether or not an instrument causes a change of state in a patient does not appear to be a definitive feature of Instrumenthood, as we found no evidence across tasks, measures, or languages that Tool videos are more prototypical than Contact videos. This is consistent with linguistic proposals in which affectedness is a matter of degree (Beavers, 2011)—in the examples of a plate being smashed (Tool condition) versus a cat being tapped on the head (Contact condition), the plate and cat are both affected and the difference between them is only in the end state. In addition, being a causal intermediary in the absence of an intentional agent does not appear to be sufficient for Instrumental prototypicality, for two reasons: in all three languages, Inanimate Agent events were in the furthest cluster from Tool events, and for Accidental Agent events, periphrastic instrumental sentences were rated as fairly infelicitous.

The results for Body Part videos support the idea that prototypical Instruments are direct extensions of intentional agents. In Study 1, Body Part videos elicited core instrumental terms to a comparable degree as Tool videos did, and in Study 2, it was only for German *benutzen* that Tool ratings were higher than Body Part ratings. This finding is noteworthy, as in the animal literature, an animal being able to manipulate an object outside the scope of their own body is thought to be definitive of tool use. In addition, human material culture is due in large part to our ability to manipulate external objects to overcome the limitations of our bodies. Nonetheless, the speakers we tested appeared to readily categorize a subpart of the

body as an instrument, at least for the purposes of language. Construing body parts as Instruments is natural perhaps because body parts are quintessentially embedded within an animate agent.

Beyond pointing toward the prototypical Instrument, our results suggest an Instrumental category with gradient structure—for some events, the Target shares some but not all instrumental properties. Means of Transit, Locatum, and Accidental Agent events demonstrated some similarity with Tool videos but were also encoded relatively variably across languages and across terms. In Dutch and German, Means of Transit videos clustered with Tool videos and also elicited high ratings for sentences with *met* and *mit*. These patterns were not observed in English, suggesting that a means of transit is a more atypical member of an instrumental category in English than it is in Dutch or German. For Locatum events, we observed instrumental construals of events in conflict with spatially-oriented construals: in all three languages, many of the Locatum videos clustered with Put-Theme videos. For example, Fig. 5 shows that English speakers often used *with* and *use* for [locatum_paint] (e.g., *the person is painting on a piece of paper, covering it entirely with blue paint*). At the same time, the spatially-oriented terms *in* and *into* were also well-attested for this video (e.g., *a woman picked up a paintbrush and dipped it into some blue paint and then painted a plain piece of paper*). Speaking also to the ambiguous status of Locatum videos, in all three languages, these videos elicited high ratings for prepositional sentences but not for periphrastic ones. Turning to Accidental Agent videos, a perhaps surprising result is that in all three languages, these videos had high naming agreement (see Table 3) and ratings for prepositional sentences were as high as ratings for Tool videos. These results suggest that for events where the agent is acting unintentionally, *with/met/mit* are not marginal but entirely felicitous. As described in Section 2.6, we interpret the high naming agreement for Accidental Agent events with caution, as speakers frequently omitted the Target in their descriptions and periphrastic instrumental sentences were not highly rated for these events.

Indirect Action events were a step further away from Tool events: they clustered relatively distantly from Tool events in all three languages, and ratings for Indirect Action videos were lower than ratings for Tool videos for all six terms. This finding highlights the importance of being a direct part of the causal force exerted by the agent for Instrumental categorization, as far as English, Dutch, and German are concerned. The videos that patterned most distantly from Tool events in the semantic space were Inanimate Agent, Put-Theme, and Give-Theme events. In Study 2, ratings for Put-Theme and Give-Theme videos were among the lowest, for all six terms. These results confirm our expectation that Put-Theme and Give-Theme events would serve as an outer boundary on the Instrument category. People appear to represent such events in terms of abstract schemas that are fundamentally different from our representation of instrumental events: the schema of caused change of location for Put-Theme events and the schema of transfer between two individuals for Give-Theme events (Jackendoff, 1983, 1990).

A major outstanding question is whether the Proto-Instrument core that emerged for English, Dutch, and German (Tool, Contact, and Body Part events) is shared across a wide range of genetically diverse languages. While Instrument may not be needed as a semantic primitive to account for the meanings of instrumental markers such as *with* and *use* (see Section 1.1), Instrument may still have broad relevance as a linguistic category if instrumental

markers universally encode the same prototype. By contrast, the subtle differences in Instrument category structure that we observed across English, Dutch, and German may be just the tip of the iceberg once a larger and more diverse set of languages has been tested. We can envisage, for example, languages where instrumentality is largely mechanistic rather than agent-oriented—in such languages, the Targets in the Inanimate Agent events might be fairly prototypical. If the languages of the world carve up instrumental conceptual space in radically different ways, this would suggest that Instrument lacks the broad explanatory value of event categories such as Agent.

We also note that we probed only part of the instrumental conceptual space, and that greater cross-variability might be revealed once a larger range of event types has been tested. Studies such as ours require striking a balance between number of conditions and number of items per condition. Although we tested 10 conditions in this study, we did not test all dimensions of potential relevance to understanding similarity structure in instrumental events. For example, we did not test whether physical concreteness (compare *Jayda used provocative language to get the judges' attention*) or being inanimate (compare *Gigi used her supermodel boyfriend to get into clubs*) is characteristic of Proto-Instruments. Another relevant dimension of conceptual space that is worth testing in future studies is typicality—either the typicality of the Target with respect to the event (e.g., cutting bread with a knife vs. cutting bread with a pair of scissors) or the familiarity of the event as a whole (e.g., cutting bread with a knife is fairly familiar but tapping a box with a pen is fairly unfamiliar). Investigating additional dimensions is important to comprehensively account for instrumental categorization across languages.

To further confirm our observations, future studies could also explore a wider range of events for the conditions tested herein. Our design featured five videos from each condition—it is possible that if we had chosen different sets of five videos (e.g., someone cutting paper with scissors rather than someone slicing bread with a knife), we would have found that Tool videos were more prototypical than Contact videos (or vice versa). In practice, however, we have no reason to think that the videos we tested are a nonrepresentative sample of each type of event, but this is something that could be explored in more detail. For example, a future study could test for prototypicality structure focusing on the Tool, Contact, and Body Part conditions, including a wider range of items.

4.2. The Instrument category in language versus conceptual representation

Thematic roles have been argued to be part of innate core knowledge due to the importance of these categories in shaping cognitive development, language development, language emergence, and spontaneous event interpretation, as described in the Introduction. Thematic roles are also argued to be domain-general, with homologous structure across linguistic and conceptual domains (see Rissman & Majid, 2019; Ünal, Ji, & Papafragou, 2020 for review). In our studies, we used linguistic data—that is, whether a particular lexical item can be used in a description of an event—to make inferences about conceptual structure. Further studies are needed to test whether the category structure that emerged from this linguistic evidence parallels category structure as demonstrated in other, nonlinguistic measures.

One source of evidence is behavior in tasks probing event cognition. For example, change-detection studies have demonstrated a conceptual asymmetry between Goals and Sources (Lakusta & Landau, 2012; Regier & Zheng, 2007) and between Goals and Instruments (Ünal et al., 2021). Individuals that constitute more prototypical agents given linguistic diagnostics have been shown to be more prototypical agents in nonlinguistic categorization tasks (Rissman & Lupyan, 2021). Similar methods could be used to assess whether the Instrumental category structure observed here is domain-general—for example, whether memory is more robust for the Targets in Tool events than for the Targets in Indirect Action events. Another option is to ask participants to read sentences with Targets filling different roles (e.g., *the chef used a knife to cut the bread* vs. *the worker used a ladder to paint the ceiling*) and ask whether recognition of pictures of the Targets is more robust when the Target is a tool.

An additional source of evidence about the status of instruments in conceptual representation is whether speakers spontaneously mention the instrument in event descriptions. In the current study, we asked how lexical items partition the space of instrumental events—answering this question required eliciting descriptions of the Targets. Studying whether people spontaneously mention the Target provides a window into the conceptual prominence of the Target (or, at least, whether the speaker thinks their addressee needs to know about the Target). As noted in footnote 2 and shown in Appendix B, when English speakers described the videos without being prompted with red circles, they frequently omitted the Target for some events. For the video of a woman eating broccoli with a fork, for example, 52% of participants did not mention the fork. Such data can reflect whether people represent the instrument as an integral part of the event.

Studies examining child cognitive and linguistic development could provide further insight as to the prominence and stability of an Instrument category. Infants as young as four months are sensitive to the functions of tools in events (Stavans & Baillargeon, 2018). This finding is consistent with the tool-prototype hypothesis but also allows the possibility that other types of instruments, such as body parts, are equally prototypical. Additional research is needed to distinguish these possibilities.

5. Conclusion

Previous studies testing prototype theories of thematic roles have largely focused on Agent and Patient roles. This is the first study to our knowledge that tests whether the Instrumental role also has prototype structure, and systematically examines the nature of that prototype. Across English, Dutch, and German, despite small differences, we found strong evidence for an underlying conceptual structure that is shared across languages, supporting prototype analyses of thematic roles. We found the prototypes in each language were largely overlapping, although not identical. We suggest the prototypical Instrument in English, Dutch, and German is the direct extension of an intentional agent, supporting the idea that instrumentality is inextricably linked to agency. Our results are compatible with the hypothesis that there is a universal Proto-Instrument role and sets the stage for testing this hypothesis across a larger, more diverse set of languages

Acknowledgments

This research was supported by NSF-PAC #1734260, a Radboud Excellence Initiative post-doctoral fellowship awarded to Lilia Rissman, the Radboud University Center for Language Studies, and the University of York Department of Psychology. Thank you to Ronny Bujok and Marieke van Wel for assistance with data collection and transcription and thank you to all research participants.

Notes

1 The notation "#'"indicates that a sentence does not felicitously convey a specific intended meaning. Sentences marked with "#" are not necessarily ill-formed syntactically.

2 In this online study, the mean rate of producing the Target was 81% across the eight instrumental conditions (i.e., excluding the Put-Theme and Give-Theme conditions). Mean rates of producing the target were low in some crucial conditions (76% in the Tool condition and 66% in the Body condition).

3 It is possible the inclusion of red circles biased participants' descriptions, for example, the red circles may have encouraged participants to describe the Target entities through nominal phrases rather than denominal verbs (e.g., *she broke the plate with a hammer* vs. *she hammered the plate*). To test this possibility, we compared the English descriptions from Study 1 to the descriptions from the English task described in footnote 2, where there were no red circles in the stimuli. To test whether red circles influenced descriptions, we analyzed the data from the "no circles" study using correspondence analysis and hierarchical clustering as described in Section 2.5.1. The resulting clustering of the videos was parallel to what we find in the original English dataset reported in Study 1 (for a side-by-side comparison of the clustering solutions for these two studies, see Supporting Information at https://osf.io/3r28k/). This suggests the red circles do not bias participants to describe videos in a systematically different way than if there had been no circles. We also tested the specific hypothesis that the red circles discouraged the use of denominal verbs such as *hammer*. Across the 45 videos with red circles, we judged that in English it is possible to describe the Target through a denominal verb for three videos: *hammer* for [tool_break], *paint* for [locatum_paint] and *elbow* for [body_elbow]. Comparing the percentage of trials where these verbs were produced between Study 1 versus the "no circles" study, we found denominal verbs were used at comparable rates ([tool_break]: 0% versus 0%; [locatum_paint]: 88% versus 92%; [body_elbow]: 5% versus 0%). This also confirms the red circles did not bias descriptions. No red circles were present for Means of Transit videos.

4 We coded the periphrastic verb *use* (e.g., *the woman used a hammer to smash the plate*) as a distinct term from the adverbial modifier *using*, as these terms appear to be felicitous for different sets of events. For example, *Max wrote a letter using the hour before lunch* appears degraded in its felicity relative to *Max used the hour before lunch to write a letter*. *Use* and *using* have different syntactic positions — if the description were *the woman was using a hammer to smash the plate*, the term would be *use*.

5 These 18 videos were: [tool_break], [tool_cut], [tool_eat], [tool_grind], [tool_kill] , [contact_beat], [contact_tap], [contact_tie], [contact_touch], [contact_wipe], [body_foot], [body_hand], [body_hands], [body_elbow], [accid_broom], [accid_wand], [locatum_fill], and [locatum_paint].
6 We excluded [locatum_fill] from this analysis in each language and excluded [give_soda] and [give_ball] from the German analysis (see Section 2.5.2).

Conflict of Interest

The authors declare no conflict of interest.

References

Abbott, J. T., Griffiths, T. L., & Regier, T. (2016). Focal colors across languages are representative members of color categories. *Proceedings of the National Academy of Sciences*, *113*(40), 11178–11183. https://doi.org/10.1073/pnas.1513298113

Ackerman, F., & Moore, J. (2001). *Proto-properties and grammatical encoding*. CSLI.

Atkinson, Q. D., & Gray, R. D. (2006). How old is the Indo-European language family? Illumination or more moths to the flame. In P. Forster & C. Renfrew (Eds.), *Phylogenetic methods and the prehistory of languages* (pp. 91–109). McDonald Institute for Archaeological Research.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using S4 classes. R package version 1.1-7. http://CRAN.R-project.org/package=lme4.

Beavers, J. (2011). On affectedness. *Natural Language & Linguistic Theory*, *29*, 335–370.

Berlin, B., & Kay, P. (1969). *Basic colour terms* (Vol. *19*): University of California Press.

Bickel, B., Zakharko, T., Bierkandt, L., & Witzlack-Makarevich, A. (2014). Semantic role clustering: An empirical assessment of semantic role types in non-default case assignment. *Studies in Language*, *38*(3), 485–511. https://doi.org/10.1075/sl.38.3.03bic

Bouckaert, R., Lemey, P., Dunn, M., Greenhill, S. J., Alekseyenko, A. V., Drummond, A. J., Gray, R. D., Suchard, M. A., & Atkinson, Q. D. (2012). Mapping the origins and expansion of the Indo-European language family. *Science*, *337*(6097), 957–960. https://doi.org/10.1126/science.1219669

Brown, D. E. (2004). Human universals, human nature & human culture. *Daedalus*, *133*(4), 47–54.

Brown, P. M., & Dell, G. S. (1987). Adapting production to comprehension - The explicit mention of instruments. *Cognitive Psychology*, *19*(4), 441–472.

Brown, R. W., & Lenneberg, E. H. (1954). A study in language and cognition. *The Journal of Abnormal and Social Psychology*, *49*(3), 454–462.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, *10*(1), 395–411. https://doi.org/10.18637/jss.v080.i01

Carey, S. (2009). *The origin of concepts*. Oxford University Press.

Chomsky, N. (1972). *Studies on semantics in generative grammar*. Mouton.

Comrie, B. (2013). Alignment of case marking of full noun phrases. In M. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online*. Max Planck Institute for Evolutionary Anthropology.

Croft, W. (1991). *Syntactic categories and grammatical relations: the cognitive organization of information*. University of Chicago Press.

Croft, W. (2003). *Typology and universals* (2nd ed.). Cambridge University Press.

Davidoff, J., Davies, I., & Roberson, D. (1999). Colour categories in a stone-age tribe. *Nature*, *398*(6724), 203–204.

Dowty, D. (1989). On the semantic content of the notion of 'thematic role'. In G. Chierchia, B. Partee, & R. Turner (Eds.), *Properties, types and meaning* (Vol. *2*, pp. 69–129). Springer.

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, *67*(3), 547–619. http://www.jstor.org/stable/415037

Dryer, M. (2013). Order of subject, object and verb. In M. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online*. Max Planck Institute for Evolutionary Anthropology.

Ergin, R., Meir, I., Aran, D. I., Padden, C., & Jackendoff, R. (2018). The development of argument structure in central taurus sign language. *Sign Language Studies*, *18*(4), 612–639.

Fillmore, C. J. (1968). The case for case. In E. W. Bach & R. T. Harms (Eds.), *Universals in linguistic theory* (pp. 1–88). Holt, Rinehart and Winston.

Flaherty, M. E. (2014). *The emergence of argument structural devices in Nicaraguan Sign Language*: The University of Chicago.

Galili, T. (2015). dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics*, *31*(22), 3718–3720.

Gärdenfors, P. (2014). *The geometry of meaning: Semantics based on conceptual spaces*. MIT Press.

Geeraerts, D. (2010). *Theories of lexical semantics*. Oxford University Press.

Gibson, E., Futrell, R., Jara-Ettinger, J., Mahowald, K., Bergen, L., Ratnasingam, S., Gibson, M., Piantadosi, S. T., & Conway, B. R. (2017). Color naming across languages reflects color use. *Proceedings of the National Academy of Sciences*, *114*(40), 10785.

Goddard, C., Wierzbicka, A., & Wong, J. (2016). "Walking" and "running" in English and German: The conceptual semantics of verbs of human locomotion. *Review of Cognitive Linguistics*, *14*(2), 303–336. https://doi.org/10.1075/rcl.14.2.03god

Goldin-Meadow, S., Brentari, D., Coppola, M., Horton, L., & Senghas, A. (2015). Watching language grow in the manual modality: Nominals, predicates, and handshapes. *Cognition*, *136*(0), 381–395. http://doi.org/10.1016/j.cognition.2014.11.029

Greenacre, M. J. (2007). *Correspondence analysis in practice* (Vol. *2*). Chapman & Hall/CRC.

Grigoroglou, M., & Papafragou, A. (2019). Children's (and adults') production adjustments to generic and particular listener needs. *Cognitive Science*, *43*(10), e12790. https://doi.org/10.1111/cogs.12790

Grimm, S. (2011). Semantics of case. *Morphology*, *21*(3), 515–544. https://doi.org/10.1007/s11525-010-9176-z

Gruber, J. (1965). *Studies in lexical relations*. (Dissertation/Thesis). Massachusetts Institute of Technology. Dept. of Modern Languages,

Hafri, A., Papafragou, A., & Trueswell, J. C. (2013). Getting the gist of events: Recognition of two-participant actions from brief displays. *Journal of Experimental Psychology: General*, *142*(3), 880–905. https://doi.org/10.1037/a0030045

Hafri, A., Trueswell, J. C., & Strickland, B. (2018). Encoding of event roles from visual scenes is rapid, spontaneous, and interacts with higher-level visual processing. *Cognition*, *175*(JUNE), 36–52. https://doi.org/10.1016/j.cognition.2018.02.011

Haidle, M. N. (2014). Building a bridge—an archeologist's perspective on the evolution of causal cognition. *Frontiers in Psychology*, *5*(1472). https://doi.org/10.3389/fpsyg.2014.01472

Hampton, J. A. (2006). Concepts as prototypes. *The Psychology of Learning and Motivation*, *46*, 79–113. https://doi.org/10.1016/S0079-7421(06)46003-5

Hartmann, I., Haspelmath, M., & Cysouw, M. (2014). Identifying semantic role clusters and alignment types via microrole coexpression tendencies. *Studies in Language*, *38*(3), 463–484. https://doi.org/10.1075/sl.38.3.02har

Haspelmath, M. (2003). The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. In M. Tomasello (Ed.), *The new psychology of language: Cognitive and functional approaches to language structure* (Vol. *2*, pp. 211–242). Erlbaum.

Heider, E. R. (1972). Universals in color naming and memory. *Journal of Experimental Psychology*, *93*(1), 10–20.

Hopper, P. J., & Thompson, S. A. (1980). Transitivity in grammar and discourse. *Language*, *56*(2), 251–299. http://www.jstor.org/stable/413757

Husson, F., Lê, S., & Pagès, J. (2017). *Exploratory multivariate analysis by example using R*. Chapman and Hall/CRC.

Iwata, S. (2008). *Locative alternation: a lexical-constructional approach*. John Benjamins Publishing Company.

Jackendoff, R. (1983). *Semantics and cognition* (Vol. *8*). MIT press.

Jackendoff, R. (1987). The Status of thematic relations in linguistic theory. *Linguistic Inquiry*, *18*(3), 369–411. https://doi.org/10.2307/4178548

Jackendoff, R. (1990). *Semantic structures*. MIT Press.

Janda, L. A. (2013). *A geography of case semantics: The Czech dative and the Russian instrumental*. Mouton de Gruyter.

Josserand, M., Meeussen, E., Majid, A., & Dediu, D. (2021). Environment and culture shape both the colour lexicon and the genetics of colour perception. *Scientific Reports*, *11*(1), 19095. https://doi.org/10.1038/s41598-021-98550-3

Kako, E. (2006). Thematic role properties of subjects and objects. *Cognition*, *101*(1), 1–42. http://doi.org/10.1016/j.cognition.2005.08.002

Kim, J. S., Elli, G. V., & Bedny, M. (2019). Knowledge of animal appearance among sighted and blind adults. *Proceedings of the National Academy of Sciences*, *116*(23), 11213–11222. https://doi.org/10.1073/pnas.1900952116

Koenig, J.-P., & Davis, A. R. (2006). The KEY to lexical semantic representations. *Journal of Linguistics*, *42*(1), 71–108.

Koenig, J.-P., Mauner, G., Bienvenue, B., & Conklin, K. (2008). What with? The anatomy of a (Proto)-role. *Journal of Semantics*, *25*(2), 175–220.

Kopecka, A., & Narasimhan, B. (Eds.). (2012). *Events of putting and taking: a crosslinguistic perspective*. John Benjamins Publishing.

Lakoff, G. (1968). Instrumental adverbs and the concept of deep structure. *Foundations of Language*, *4*(1), 4–29. http://www.jstor.org/stable/25000311

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press.

Lakusta, L., & Landau, B. (2012). Language and memory for motion events: Origins of the asymmetry between source and goal paths. *Cognitive Science*, *36*(3), 517–544. https://doi.org/10.1111/j.1551-6709.2011.01220.x

Lakusta, L., Spinelli, D., & Garcia, K. (2017). The relationship between pre-verbal event representations and semantic structures: The case of goal and source paths. *Cognition*, *164*, 174–187. https://doi.org/10.1016/j.cognition.2017.04.003

Lê, S., Josse, J., & Husson, F. (2008). FactoMineR: An R package for multivariate analysis. *2008*, *25*(1), 18. https://doi.org/10.18637/jss.v025.i01

Levin, Beth (1993). *English verb classes and alternations: a preliminary investigation*. Chicago, IL: University of Chicago Press.

Levin, B., & Rappaport-Hovav, M. (2005). *Argument realization*. Cambridge University Press.

Lockridge, C. B., & Brennan, S. E. (2002). Addressees' needs influence speakers' early syntactic choices. *Psychonomic bulletin & review*, *9*(3), 550–557. https://doi.org/10.3758/BF03196312

Luján, E. R. (2010). Semantic maps and word formation: Agents, instruments, and related semantic roles. *Linguistic Discovery*, *8*(1), 162–175.

Luján, E. R., & Ruiz Abad, C. (2014). Semantic roles and word formation: instrument and location in Ancient Greek. In S. Luraghi & H. Narrog (Eds.), *Perspectives on semantic roles* (pp. 241–269). John Benjamins.

Luraghi, S. (2001). Some remarks on instrument, comitative, and agent in Indo-European. *STUF - Language Typology and Universals*, *54*(4), 385–401. https://doi.org/10.1524/stuf.2001.54.4.385

Luraghi, S. (2003). *On the meaning of prepositions and cases: The expression of semantic roles in Ancient Greek*. John Benjamins Publishing.

Majid, A., Boster, J. S., & Bowerman, M. (2008). The cross-linguistic categorization of everyday events: A study of cutting and breaking. *Cognition*, *109*(2), 235–250. https://doi.org/10.1016/j.cognition.2008.08.009

Majid, A., & Burenhult, N. (2014). Odors are expressible in language, as long as you speak the right language. *Cognition*, *130*(2), 266–270. https://doi.org/10.1016/j.cognition.2013.11.004

Majid, A., Gullberg, M., van Staden, M., & Bowerman, M. (2007). How similar are semantic categories in closely related languages? A comparison of cutting and breaking in four Germanic languages. *Cognitive Linguistics*, *18*(2), 179–194. https://doi.org/10.1515/COG.2007.007

Majid, A., Roberts, S. G., Cilissen, L., Emmorey, K., Nicodemus, B., O'Grady, L., Woll, B., LeLan, B., de Sousa, H., Cansler, B. L., Shayan, S., de Vos, C., Senft, G., Enfield, N. J., Woll, B., Razak, R. A., Fedden, S., Tufvesson, S., Dingemanse, M., Ozturk, O., Brown, P., Hill, C., Le Guen, O., Hirtzel, V., van Gijn, R., Sicoli, M. A., & Levinson, S. C. (2018). Differential coding of perception in the world's languages. *Proceedings of the National Academy of Sciences*, *115*(45) 11369-11376.

Malchukov, A., Haspelmath, M., & Comrie, B. (2010). Ditransitive constructions: A typological overview. In A. Malchukov, M. Haspelmath, & B. Comrie (Eds.), *Studies in ditransitive constructions: A comparative handbook* (pp. 1–64). Walter de Gruyter.

Malt, B. C., Ameel, E., Imai, M., Gennari, S. P., Saji, N., & Majid, A. (2014). Human locomotion in languages: Constraints on moving and meaning. *Journal of Memory and Language*, *74*, 107–123. https://doi.org/10.1016/j.jml.2013.08.003

Murphy, G. L. (2002). *The big book of concepts*. The MIT Press.

Narrog, H., & Ito, S. (2007). Re-constructing semantic maps: the comitative-instrumental area. *STUF – Sprachtypologie und Universalienforschung*, *60*, 273–292. https://doi.org/10.1524/stuf.2007.60.4.273

Neldner, K., Reindl, E., Tennie, C., Grant, J., Tomaselli, K., & Nielsen, M. (2020). A cross-cultural investigation of young children's spontaneous invention of tool use behaviours. *Royal Society Open Science*, *7*(5), 192240. https://doi.org/10.1098/rsos.192240

Newman, J. (Ed.). (1998). *The linguistics of giving*. John Benjamins.

Nilsen, D. L. F. (1973). *The instrumental case in English; syntactic and semantic considerations*. Mouton.

Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*. MIT Press.

Plotnik, J. M., & Clayton, N. S. (2015). Convergent cognitive evolution across animal taxa: comparisons of chimpanzees, corvids and elephants. In E. Margolis & S. Laurence (Eds.), *The conceptual mind: New directions in the study of concepts* (pp. 29–56).

Primus, B. (1999). *Cases and thematic roles: Ergative, accusative and active* (Vol. *393*). Niemeyer.

R Core Team. (2022). *R: A language and environment for statistical computing*. https://www.R-project.org/

Regier, T., Kay, P., & Cook, R. S. (2005). Focal colors are universal after all. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(23), 8386–8391. https://doi.org/10.1073/pnas.0503281102

Regier, T., & Zheng, M. (2007). Attention to endpoints: A Cross-linguistic constraint on spatial meaning. *Cognitive Science*, *31*(4), 705–719. https://doi.org/10.1080/15326900701399954

Reisinger, D., Rudinger, R., Ferraro, F., Harman, C., Rawlins, K., & Van Durme, B. (2015). Semantic Proto-Roles. *Transactions of the Association for Computational Linguistics*, *3*, 475–488.

Rice, S. (1996). Prepositional prototypes. In M. Pütz & R. Dirven (Eds.), *The construal of space in language and thought* (pp. 135–165). Mouton de Gruyter.

Rice, S., & Kabata, K. (2007). Crosslinguistic grammaticalization patterns of the allative. *Linguistic Typology*, *11*(3), 451–514. https://doi.org/10.1515/LINGTY.2007.031

Rissman, L. (2013). Periphrastic use: a modal account of instrumentality. In Y. Fainleib, N. LaCara, & Y. Park (Eds.), *Proceedings of the 41st annual meeting of the north east linguistics society* (Vol. *2*, pp. 137–150). GLSA Publications.

Rissman, L., & Goldin-Meadow, S. (2017). The development of causal structure without a language model. *Language Learning and Development*, *13*(3), 286–299. https://doi.org/10.1080/15475441.2016.1254633

Rissman, L., & Lupyan, G. (2021). A dissociation between conceptual prominence and explicit category learning: Evidence from Agent and Patient event roles. *Journal of Experimental Psychology: General*. Advance online publication. https://doi.org/10.1037/xge0001146

Rissman, L., & Majid, A. (2019). Thematic roles: Core knowledge or linguistic construct? *Psychonomic Bulletin & Review*, *26*(6), 1850–1869. https://doi.org/10.3758/s13423-019-01634-5

Rissman, L., & Rawlins, K. (2017). Ingredients of instrumental meaning. *Journal of Semantics*, *34*(3), 507–537. https://doi.org/10.1093/jos/ffx003

Rissman, L., Rawlins, K., & Landau, B. (2015). Using instruments to understand argument structure: Evidence for gradient representation. *Cognition*, *142*(0), 266–290. http://doi.org/10.1016/j.cognition.2015.05.015

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General; Journal of Experimental Psychology: General*, *104*(3), 192.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Lawrence Erlbaum.

Schlesinger, I. M. (1989). Instruments as agents: On the nature of semantic relations. *Journal of Linguistics*, *25*(1), 189–210.

Schlesinger, I. M. (1995). *Cognitive space and linguistic case*. Cambridge University Press.

Seed, A., & Byrne, R. (2010). Animal tool-use. *Current Biology*, *20*(23), R1032-R1039. https://doi.org/10.1016/j.cub.2010.09.042

Siewierska, A. (2013). Alignment of verbal person marking. In M. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online*. Max Planck Institute for Evolutionary Anthropology.

Simms, L. J., Zelazny, K., Williams, T. F., & Bernstein, L. (2019). Does the number of response options matter? Psychometric perspectives using personality questionnaire data. *Psychological assessment*, *31*(4), 557–566.

Simpson, E. H. (1949). Measurement of diversity. *Nature*, *163*(4148), 688-688.

Skelton, A. E., Catchpole, G., Abbott, J. T., Bosten, J. M., & Franklin, A. (2017). Biological origins of color categorization. *Proceedings of the National Academy of Sciences*, *114*(21), 5545–5550.

Stavans, M., & Baillargeon, R. (2018). Four-month-old infants individuate and track simple tools following functional demonstrations. *Developmental Science*, *21*(1), e12500. https://doi.org/10.1111/desc.12500

Stolz, T. (2001). Comitatives vs. instrumentals vs. agents. In T. Stolz (Ed.), *Aspects of typology and universals* (Vol. *1*, pp. 153–174). Akademie Verlag.

Strickland, B. (2017). Language reflects "core" cognition: A New theory about the origin of cross-linguistic regularities. *Cognitive Science*, *41*, 70–101. https://doi.org/10.1111/cogs.12332

Talmy, L. (1976). Semantic causative types. In M. Shibatani (Ed.), *The grammar of causative constructions* (Vol. *6*, pp. 43–116). Academic Press.

Tatone, D., Geraci, A., & Csibra, G. (2015). Giving and taking: Representational building blocks of active resource-transfer events in human infants. *Cognition*, *137*(0), 47–62. http://doi.org/10.1016/j.cognition.2014.12.007

Taylor, J. R. (2003). *Linguistic categorization: prototypes in linguistic theory*. Oxford University Press.

Ünal, E., Ji, Y., & Papafragou, A. (2020). From event representation to linguistic meaning. *Topics in Cognitive Science*. https://doi.org/10.1111/tops.12475

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

Vaesen, K. (2012). The cognitive bases of human tool use. *Behavioral and Brain Sciences*, *35*(04), 203–218. https://doi.org/10.1017/S0140525x11001452

White, A. S., Rawlins, K., & Van Durme, B. (2017). The semantic Proto-Role linking model. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, *2*, 92–98.

Wilson, F., Papafragou, A., Bunger, A., & Trueswell, J. (2011). Rapid extraction of event participants in caused motion events. *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, *33*, 1206–1211.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, *104*(19), 7780–7785.

Yin, J., & Csibra, G. (2015). Concept-based word learning in human infants. *Psychological Science*, *26*(8), 1316–1324. https://doi.org/10.1177/0956797615588753

Appendix A

Condition	Video	Video Description	Target	Video Length (s)	Intention Norm Score (*M*)	Change Norm Score (*M*)
Tool	tool_break	A woman breaking a plate with a hammer	Hammer	3	4.9	4.9
	tool_cut	A woman slicing a baguette with a knife	Knife	9	4.8	4.7
	tool_eat	A woman eating broccoli with a fork	Fork	8	4.7	4.7
	tool_grind	A woman grinding pumpkin seeds with mortar & pestle	Mortar and/or pestle	9	4.9	4.9
	tool_kill	A woman killing a ladybug with a shoe	Shoe	8	4.8	5.0
Contact	contact_beat	A woman beats a towel laid over a rope with a plastic dustpan	Dustpan	6	4.9	2.2
	contact_tap	A woman taps a small metal box with a pen	Pen	2	4.8	1.2
	contact_tie	A woman ties another woman to a tree using rope	Rope	17	4.9	2.1
	contact_touch	A woman touches a cat on its head with a stick	Stick	5	4.7	1.8
	contact_wipe	A woman wipes a table with a yellow rag	Rag	4	4.7	1.8
Body Part	body_elbow	A man intentionally smashes a small cupcake with his elbow	Elbow	2	4.8	4.8
	body_foot	A man passes a scarf to another woman using his foot	Foot	5	4.5	2.9
	body_hand	A man intentionally knocks over a music stand with his hand	Hand	3	4.8	3.9
	body_hands	A woman intentionally knocks over a stack of cups with her hands	Hands	3	4.5	4.3
	body_leg	A man breaks a stick over his leg	Leg/thigh	2	5.0	4.6

Condition	Video	Video Description	Target	Video Length (s)	Intention Norm Score (*M*)	Change Norm Score (*M*)
Accidental Agent	accid_balls	A woman tries to juggle three balls, she fails to catch one, which then knocks over a plastic bottle	Balls	3	1.4	3.0
	accid_broom	A woman is sweeping the floor with a broom, she accidentally knocks over a plastic bottle with the broom	Broom	8	1.4	2.8
	accid_cake	A man is drinking from a mug, he accidentally places the mug down on top of a small cupcake which is in front of him	Mug	6	1.2	4.7
	accid_matzah	A woman is tossing a ball back and forth between her hands, she drops it and it falls and smashes a piece of cracker in front of her	Ball	6	1.2	4.7
	accid_wand	A man is conducting with a wand, he accidentally knocks over a music stand with the wand	Wand	5	1.1	3.6
Means of Transit	transit_bike	Someone bikes from Antwerp to Brussels	Bike/cycling	20	NA	NA
	transit_boat	Someone takes a ferry from Port Angeles to Victoria, Canada	Ferry	20	NA	NA
	transit_car	Someone drives from Amsterdam to Paris	Car/driving	23	NA	NA
	transit_fly	Someone flies from Rome to Moscow	Plane/flying	17	NA	NA
	transit_train	Someone takes a train from London to Edinburgh	Train	18	NA	NA

Condition	Video	Video Description	Target	Video Length (s)	Intention Norm Score (*M*)	Change Norm Score (*M*)
Locatum	locatum_cover	A woman covering a baby doll with a towel	Towel	4	4.8	2.2
	locatum_cram	A woman stuffing balled up socks into a travel accessories case	Socks	11	4.5	2.8
	locatum_fill	A woman pouring orange juice from a pitcher into a glass, filling up the glass	Orange juice	11	4.8	2.4
	locatum_load	A woman loading groceries into a plastic groceries cart	Groceries	11	4.6	2.8
	locatum_paint	A woman painting a white sheet of paper completely blue	Blue paint	28	4.9	4.6
Indirect Action	indirect_board	A woman chops a carrot on a chopping board	Chopping board	8	4.8	4.9
	indirect_chair	A woman climbs on a chair in order to reach a plant on top of a bookshelf	Chair	4	4.9	2.5
	indirect_ladder	A woman climbs a ladder in order to open a window	Ladder	9	4.8	2.8
	indirect_notepad	A woman draws a heart on a notepad with a pen	Notepad	12	4.9	3.3
	indirect_wash	A woman washes spinach in a colander under a faucet	Colander	5	4.7	2.9
Inanimate Agent	inanim_ball	A piece of train track falls over, which hits a blue cylinder, which falls and hits a blue ball, which rolls	Blue cylinder	2	1.2	2.9
	inanim_blocks	A large piece of train track falls, which hits a medium piece of train track, which hits a small piece of train track that falls, in a domino effect	Medium piece of track	2	1.4	3.4
	inanim_chain	A blue object swings back and forth on a chain, ultimately hitting a block, which falls and hits a red car, which rolls forward	Block	3	1.4	2.8

Condition	Video	Video Description	Target	Video Length (s)	Intention Norm Score (*M*)	Change Norm Score (*M*)
	inanim_train	A white train rolls down a ramp, hitting a red car, which rolls forward and hits an orange car, which rolls forward	Red car	2	2.4	2.8
	inanim_tree	A white train rolls down a ramp, hitting a red van, which rolls forward and hits a toy tree, which falls	Red van	2	1.9	3.4
Put-Theme	put_ball	A man drops an apple into a black bag	Apple	2	4.8	2.4
	put_block	A woman pushes a ball against a wooden block	Ball	1	4.8	1.0
	put_box	A man puts a box on a bookshelf	Box	3	4.8	1.7
	put_chips	A woman moves a bag of chips from a table to a chair	Chips	2	4.8	1.4
	put_rope	A woman puts some rope over the branch of a tree	Rope	2	4.8	1.4
Give-Theme	give_apple	A woman throws an apple to a man	Apple	1	4.7	1.8
	give_ball	A man kicks a ball to another man	Ball	2	4.8	2.1
	give_mug	A woman gives a mug to another woman	Mug	3	4.8	1.8
	give_rag	A man gives a rag to a woman	Rag	4	4.5	2.0
	give_soda	A man takes a soda from a woman	Soda	3	4.6	1.8

Note. For the intention norm, participants rated the extent to which the agent chose to perform the action. For the change norm, participants rated the extent to which the patient changed as a result of the action.

Appendix B

Mean percentage of included trials in Study 1 for each video for each language

Condition	Video	English	Dutch	German	English (No Circles)
Tool	tool_break	1.00	1.00	1.00	1.00
	tool_cut	0.95	0.95	0.92	0.48
	tool_eat	1.00	0.97	0.97	0.48
	tool_grind	0.95	0.87	0.92	0.84
	tool_kill	1.00	1.00	0.97	1.00
Contact	contact_beat	1.00	0.97	1.00	1.00
	contact_tap	1.00	0.97	1.00	0.96
	contact_tie	0.91	0.95	0.89	0.68
	contact_touch	1.00	1.00	1.00	0.88
	contact_wipe	0.98	0.97	1.00	0.72
Body Part	body_elbow	1.00	1.00	0.94	1.00
	body_foot	1.00	1.00	1.00	1.00
	body_hand	0.81	0.79	0.89	0.36
	body_hands	0.72	0.69	0.78	0
	body_leg	0.95	0.90	0.94	0.92
Accidental Agent	accid_balls	0.33	0.26	0.19	0.36
	accid_broom	0.26	0.21	0.17	0.24
	accid_cake	0.93	0.97	0.83	0.88
	accid_matzah	0.60	0.59	0.27	0.60
	accid_wand	0.47	0.64	0.43	0.28
Means of Transit	transit_bike	1.00	1.00	0.97	1.00
	transit_boat	0.88	1.00	0.89	0.88
	transit_car	1.00	0.92	0.94	0.96
	transit_fly	0.98	1.00	0.94	0.92
	transit_train	1.00	1.00	0.97	0.96
Locatum	locatum_cover	1.00	1.00	1.00	0.96
	locatum_cram	1.00	1.00	1.00	1.00
	locatum_fill	0.95	0.97	0.97	1.00
	locatum_load	1.00	1.00	1.00	1.00
	locatum_paint	1.00	1.00	0.97	1.00
Indirect Action	indirect_board	0.74	0.79	0.72	0.40
	indirect_chair	1.00	1.00	1.00	1.00
	indirect_ladder	1.00	1.00	1.00	1.00
	indirect_notepad	0.98	0.98	0.92	0.88
	indirect_wash	0.88	0.90	0.92	0.68
Inanimate Agent	inanim_ball	0.91	0.82	0.86	0.80
	inanim_blocks	0.93	0.64	0.42	0.80
	inanim_chain	1.00	1.00	0.97	0.92
	inanim_train	1.00	1.00	0.97	0.96
	inanim_tree	1.00	0.97	1.00	1.00

Condition	**Video**	**English**	**Dutch**	**German**	**English (No Circles)**
Put-Theme	put_ball	1.00	1.00	1.00	1.00
	put_block	1.00	1.00	1.00	1.00
	put_box	1.00	1.00	1.00	1.00
	put_chips	1.00	1.00	1.00	1.00
	put_rope	1.00	1.00	1.00	1.00
Give-Theme	give_apple	1.00	1.00	1.00	0.96
	give_ball	1.00	0.97	1.00	1.00
	give_mug	1.00	1.00	1.00	1.00
	give_rag	1.00	1.00	1.00	1.00
	give_soda	1.00	1.00	1.00	1.00

Note. The column "English (no circles)" refers to an additional study (described in Section 2.2) where English speakers described the videos without red circle prompts.