Measuring Meaning: Alignment and Misalignment Across Indices of Verb Semantics

Lilia Rissman (lrissman@wisc.edu)
Department of Psychology, 1202 W. Johnson Street
Madison, WI 53706 USA

Abstract

Developing accurate models of word meaning requires good empirical evidence about what words mean. I investigate how English verbs encode relationships between event participants, focusing on how verbs encode instruments (e.g., does *slice* specify that a tool must be used for slicing?). I compared two commonly used indices of verb meaning: linguistic judgments and sentence completions. Although these two indices were moderately correlated for a small sample of verbs, they were only weakly correlated when a larger sample of verbs was tested. These results indicate that the particular context of a task can strongly influence how meaning affects behavior. Dominant models of verb meaning fail to fully account for the results (either a logical entailment model or a cue-based model).

Keywords: semantics; verbs; thematic roles; language production; argument structure; instruments

Introduction

Words perform interpersonal magic: they allow communication between language users through shared understanding of what words mean. Determining the content of the meaning representations that are shared is a fiendishly difficult problem in cognitive science. In some theories of lexical semantics, words are stable bundles of entailments: properties that are true of all sentences regardless of the context that a word appears in (Levin & Rappaport-Hovav, 2013). In alternate theories, words do not themselves have meaning but meaning results through an interaction between a word and the entire context of its use (Elman, 2009).

In this paper, I investigate how word meaning is represented by asking how verbs specify relations between event participants (a level of meaning often referred to as *argument structure*). For example, the verb *eat* can be analyzed as encoding a relation between two arguments: the individual who eats and the entity that is eaten. In linguistic theory, one proposition entails another iff the second proposition is true in all the contexts where the first proposition is true. The relationship between a verb and its arguments constitutes strong evidence that verb meaning can be modeled in terms of entailments. For example, it is hard to imagine how an event of eating could fail to involve someone who eats: if *eating happened* is true, it follows that *someone ate* is also true.

In the literature on argument structure, arguments are often contrasted with *adjuncts*: constituents that modify an event but are not arguments of the verb (see Koenig, Mauner, & Bienvenue, 2003; Vater, 1978). For example, in *Remi ate the broccoli for his little brother,* the brother is an adjunct, because eating itself does not require that an additional person benefit from the eating. In other words, adjuncts are event participants that are not entailed by the verb. In this paper, the phrase *event participant* refers broadly to any component of an event: in *Remi ate the broccoli for his little brother in the kitchen,* for example, Remi, the broccoli, the brother, and the kitchen are all event participants.

The argument/adjunct distinction has long been noted to be problematic because for many verbs, it is unclear what the arguments are (see Rissman, Rawlins, & Landau, 2015 for review). Consider the verb *sweep*, for example—do sentences with *sweep* entail that an instrument (such as a broom) is used, or is an instrument merely a typical part of a sweeping event? The difficulty in determining which event participants are arguments of a verb undermines the proposal that a verb's meaning is a stable bundle of entailments that constrains every context of use.

One reason why this question is difficult to resolve is that a range of linguistic diagnostics and psycholinguistic tasks have been used to identify a verb's meaning, and different types of data often provide different glimpses into a verb's representation (Willits, Amato, & MacDonald, 2015). Put another way, different diagnostics of argumenthood do not always align (Vater, 1978). Here I compare two common methods for assessing which event participants are entailed by a verb: linguistic judgments and sentence completion data. In the former, participants reflect on verb meaning, whereas in the latter, participants provide a continuation given a prompt such as *Tyrell ate the banana___*. I ask whether previous analyses of how verbs represent instruments—analyses which are based on judgment data—are also supported by sentence completion data, a more implicit measure of verb semantics. I then ask whether the results from these two tasks support or cast doubt on theories in which argument structure is modeled in terms of entailment.

Verb Semantics and Instruments

For many verbs, including *eat*, determining argument structure is seemingly straightforward—eating requires someone who eats (an agent) and something that is eaten (a patient). These inherent participants are expressed in the privileged syntactic positions of Subject and Object in

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society.*

English, supporting their status as arguments of *eat* (Rappaport-Hovav & Levin, 1998).[1]

For instrumental participants such as the fork in *Remi ate the broccoli with a fork*, argument status is less clear. Psycholinguistic tasks show that some verbs activate information about instruments: for example, in a self-paced reading task, adults were faster at interpreting a phrase as instrumental when it was accompanied by a verb such as *stab, stir*, or *poke* than by a verb such as *eat, kill*, or *serve* (Koenig et al., 2003). Given such data, Koenig and colleagues argue that instruments are arguments for the former set of verbs. Whether these verbs in fact *entail* the presence of an instrument is uncertain, however, in part because of the possibility of body part instruments (see Koenig, Mauner, Bienvenue, & Conklin, 2008). For example, if I poke someone with my finger, my finger is part of me (the agent). This could be interpreted as an event with only an agent and a patient, but if my finger is interpreted as serving an instrumental role, it would be an event with an agent, a patient, and an instrument. The argument status of instruments is also unclear because as far as syntactic diagnostics, instrumental *with*-phrases pattern as adjuncts regardless of verbal meaning (Rissman et al., 2015).

In view of this difficulty, Rissman et al. (2015) assessed verbal encoding of instruments through a semantic generalization task. English-speaking adults were instructed that verbs have "arguments," for example that *steal* has three arguments, someone who steals, something that is stolen, and someone who is stolen from. After this training, participants were asked to generalize this notion of "argument" to verbs not previously encountered. In one experiment, participants read sentences such as *Martha POKED something [with a fork] [yesterday]* and had to judge either that one of the bracketed phrases was an "argument" of the verb or that neither bracketed phrase was an "argument." In a second experiment, participants were shown a verb in isolation and were asked to list the "arguments" of the verb. Judgments were collected for 24 verbs. This study produced three main findings: 1) participants were more likely to judge that the instrument was a "argument" for verbs such as *stab, stir*, or *poke* than for verbs such as *eat, kill*, or *serve,* 2) there was a gradient cline of judgments from the most instrumental verb (*slice*) to the least instrumental verb (*eat*), and 3) instrumental judgments were semantically organized, with incision verbs (*slice, cut, chop, stab*) being judged as the most instrumental (see Koenig et al., 2008 for further discussion of incision verbs). Barbu and Toivonen (2016) conducted a similar judgment study and found comparable results: instruments were judged to be more conceptually prominent for some verbs than for others.

Given these findings, Rissman et al. (2015) propose that for verbs such as *slice, cut, chop*, and *stab*, which encode the physical property of incision, instrumentality is a stable, inherent component of these verbs' meanings. At the same time, Rissman and colleagues argue that instrumentality is a gradient feature of verbal semantics, a representation which is difficult to model in terms of lexical entailments.

Approach

In the judgment studies reported by Rissman et al. (2015), participants were implicitly asked to compute similarity over different meanings (i.e., "from what I know about *steal*, what does that tell me about *slice*?"). This type of behavior suggests a particular perspective on how verbs encode instruments—namely, that abstract event features, such as incision, lead to activation of instrument knowledge. In this paper, I ask whether language production behavior suggests the same perspective on verb meaning. Sentence completion tasks are frequently used to assess how a verb biases people to reason about causes vs. consequences (e.g., fragments such as *John hugged Mary___* often elicit causal completions such as *because she was sad*) (Majid, Sanford, & Pickering, 2007). Sentence completion data are also taken to reflect a verb's inherent meaning and its argument structure (Koenig et al., 2003). For example, Koenig, Mauner, and Bienvenue (2002) found that when adults were given sentence fragments such as *the farmer split the logs* and told to provide a continuation, they more often provided instrumental completions (e.g., *with an axe*) for verbs like *stab, stir*, or *poke* than for verbs such as *eat, kill*, or *serve.* As described by Koenig et al. (2003), "participant information that is lexically encoded is retrieved upon recognition of a word. Because this information is activated, it is more likely to be used to continue a sentence" (82). Not all studies, however, have found a positive relationship between judgments and language production: Barbu and Toivonen (2016) found that verbs such as *write, scrub,* and *cut*, which frequently elicited the judgment that an instrument was conceptually prominent, almost never elicited instruments in a sentence completion task. These conflicting results call for a comparison of judgment and production data across a wider range of verbs than previously studied.

In three studies, I ask whether verbs that are judged to have instrumental meanings are also more likely to elicit instrumental completions given prompts such as *Mike sliced the bread_____*. I ask whether two previously-observed properties of instrumental judgments also characterize sentence completions: verbs being organized in terms of abstract semantic properties such as incision, and verbs falling on a gradient cline from most to least instrumental.

[1] I describe here a "projectionist" approach to argument structure, where a verb's arguments project onto syntactic structure (Rappaport-Hovav & Levin, 1998). In "non-projectionist" theories, argument structure is determined syntactically through functional projections (Borer, 2005). For either type of theory, it is a live question how verbs specify relations between event participants (i.e., through entailment or some other representational format). As I am not using syntactic argument realization evidence to make inferences about verb meaning, the current work does not presuppose that projectionist theories are correct.

Study 1

Method

Participants I tested 48 adult native English-speaking participants on Amazon Mechanical Turk (F = 34; age range = 21-72; mean age = 42). Participants received $0.50**.** In all studies, participants self-reported being native speakers of English but were not necessarily monolingual.

Design and Materials In each trial, participants viewed a sentence fragment with the form [Subject V+PAST *the* Object] (e.g., *Lucy stirred the potion* __). Participants were instructed to type in a possible completion of this sentence. 60 verbs were tested, including the 24 verbs tested by Rissman et al. (2015): *touch, hit, beat, poke, stab, cut, chop, slice, write, draw, dig, stir, eat, drink, break, open, kill, attack, paint, grow, move, lift, clean*, and *wash*. Data for the remaining 36 verbs are not included in this analysis.

Across the entire stimuli set, each verb appeared with one of six direct objects (e.g., *the potion*). These direct objects were selected to be semantically compatible with the verb but to constitute a range of common and uncommon exemplars. For *stir*, for example, the six direct objects were *potion, soup, cocktail, mud, melted chocolate,* and *scrambled eggs*. Participants viewed each verb once, that is, with a single direct object. Participants thus completed 60 trials, and each participant received a unique random order of the stimuli.

Procedure Participants were asked to provide the first sentence completion that came to mind. They were given the example fragment *Sean wants an iPad* ___ and were told that possible completions might include *for his birthday, to watch movies, because he likes Apple products*, and *more than he wants an iPhone*. Participants were instructed to make their responses varied, that is, to not just type "yesterday" for every sentence**.**

Coding Responses were coded as instrumental if they were introduced by *with* or *using* and conveyed the means by which an action was completed (e.g., Rebecca sliced the baguette___ *with the only knife that she had*, Jay painted the picture___ *using watercolors*).

Results

Across the 24 verbs, the mean rate of producing instrumental completions ranged from 0% (*grow*) to 50% (*hit*). To assess whether judgment and production tasks reflect instrumentality in similar ways, I computed Pearson's correlations between the instrument completion rate for each verb and the instrument judgment rates from the two tasks reported by Rissman et al. (2015)—I label these tasks "Sentence" and "Verb Alone." Figure 1 shows *r* and *p*-values for each of these correlations. The two judgment tasks were strongly correlated ($r = .80$). Crucially, rates of producing instrument completions were moderately and significantly correlated with instrument judgments for both tasks (Sentence $r = .60$; Verb Alone, $r = .62$).

I modeled the probability of producing an instrument completion used mixed effects logistic regression and the *lme4* package for R (Bates, Maechler, Bolker, & Walker, 2014). Models included random intercepts for participants and direct objects and judgment predictors were scaled. Judgments in the Sentence task predicted instrument completions as both a linear and a quadratic predictor (linear: $b = .98$, $CI_{95} = [.67, 1.29]$, $z = 6.11$, $p < .001$; quadratic: $b = -.45$, $CI_{95} = [-.76, -.14]$, $z = -2.99$, $p < .01$). Judgments in the Verb Alone task predicted instrument completions as a linear predictor and marginally as a quadratic predictor (linear: $b = .77$, $CI_{95} = [.52, 1.02]$, $z = 6.05$, $p < .001$; quadratic: $b = -.21$, $CI_{95} = [-.43, .006]$, $z = -1.90$, $p = .058$).

Discussion

The positive correlations and model results observed for Study 1 suggest that verb meanings have similar effects on a linguistic judgment task and a language production task. The significant quadratic predictors indicate that instrument completions are least likely for verbs on the high and low ends of the judgment continuum, a point I will return to in the General Discussion.

So far, I have considered a relatively small sample of verbs ($N = 24$). In Studies 2 and 3, I test whether the positive relationships observed between instrument judgments and instrument completions in Study 1 is present across a larger sample of verbs.

Study 2

I adapted the judgment tasks from Rissman et al. (2015), creating a method appropriate for a larger number of verbs. Rather than train participants to report judgments about "arguments," I instructed participants about strongly and weakly instrumental verbs and asked participants to generalize this instruction to new verbs.

Method

Participants I tested 33 adult native English speakers on Amazon Mechanical Turk (F = 10).[2] An additional nine speakers were tested but were excluded due to failure on control trials. Participants received $4.

Design Speakers reported judgments about 136 English verbs. The verbs were selected by asking 16 English speakers to describe 67 pictures of common instrumental events (e.g., raking leaves, eating ice cream with a spoon). The most common verbs from this sample were selected. These 136 verbs included the 24 instrumental verbs from Study 1, as well as 32 denominal verbs such as *hammer*.

[2] Due to experimenter error, age information for Study 2 participants was not collected.

Materials and Procedure Participants were instructed to report judgments on how verbs highlight entities. They were told that the verb *chase* highlights two entities, someone who chases and someone/thing that gets chased. They were then told that some verbs highlight tools, e.g. that *shred* "highlights a tool that is used for shredding (like a cheese grater)." Participants were also told that some verbs, like *smudge*, are compatible with tools but don't highlight them. In this initial instruction, participants were given four verbs that I judged to be strongly instrumental (*shred, prod, whip, whack*) and four verbs that I judged to be weakly instrumental (*smudge, revive, rotate, defrost*).

Participants then completed ten practice trials with five strong verbs (*sever, whisk, dice, scribble, plow*) and five weak verbs (*harvest, purify, inspect, lecture, shatter*). Participants were asked whether each of these 10 verbs highlights an instrument, with feedback given on each trial. All strong verbs in the instruction phase were classified as such by Koenig et al. (2008). The weak verbs were selected to match the frequency of the strong verbs in the Corpus of Contemporary American English (Davies, 2008-). In total, participants encountered 18 instructional or "seed" verbs.

Following this instruction, participants viewed each of the 136 test verbs in the context of a sentence such as *John SAW something with binoculars*. Participants were asked whether the verb in capital letters highlights a tool or not. Example tools were given to help clarify the intended sense of the verb (e.g., seeing with eyes rather than sawing with a saw). Three example sentences were presented per trial and all three example tools were typical for the verb.

The 18 seed verbs were also tested during the experimental phase as an attention check. Participants needed to answer correctly on 13 out of 18 seed verbs for their data to be included (p = .03 on a binomial test). Including seed verbs, participants judged 154 verbs in total. Each participant judged each verb once, and verbs were viewed in a random order.

Results

Some verbs were always judged to highlight a tool (*chop, drill, nail, shovel, stab, whisk*). Other verbs were never judged to highlight a tool (*check, collect, grow, look, perform, test*). The remaining verbs spanned the full range between these two extremes. For *burn, drum, juice, pump* and *take notes*, for example, 52% of participants judged that these verbs highlight a tool. Not surprisingly, denominal verbs such as *hammer* were highly likely to be judged as highlighting a tool (82%). Incision verbs such as *slice, chop*, and *mince* were also often judged to highlight a tool (89%).

The judgments from Study 2 and the judgment tasks from Rissman et al. (2015) were strongly correlated for the 24 verbs tested in common (Sentence r = .80; Verb Alone, r = .86; see Figure 1). The Study 2 judgments and Study 1 completions were moderately correlated (r = .43).

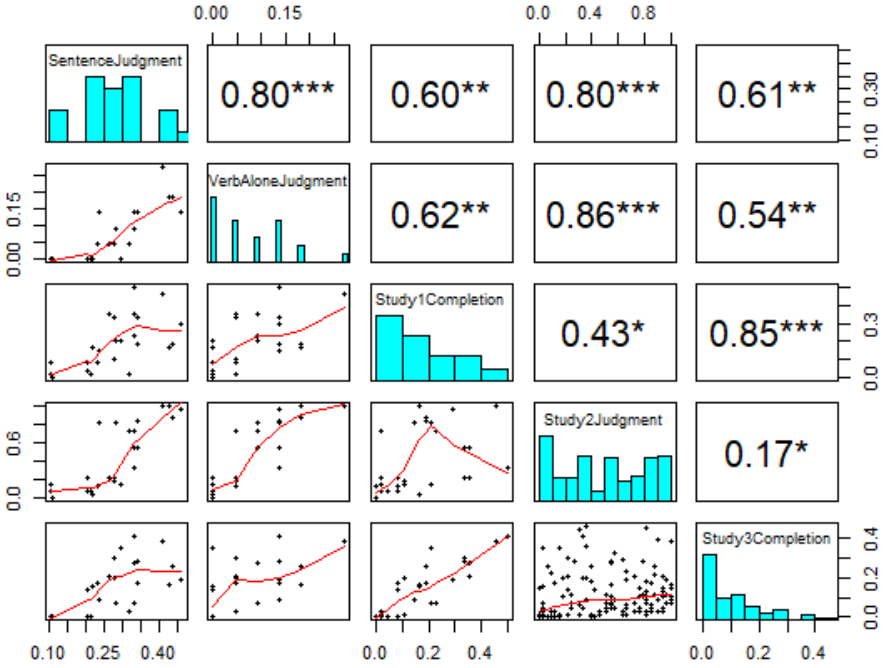

Figure 1. Relationships between each meaning measure. Points indicate individual verbs. Distributions of each measure are shown on the diagonal. Red lines are loess smoothed lines of best fit. Greater SentenceJudgment, VerbAloneJudgment and Study2Judgment values indicate stronger instrumental judgments. Greater Study1Completion and Study3Completion values indicate higher frequency of an instrument completion being produced. SentenceJudgment and VerbAloneJudgment data were originally reported in Rissman et al. (2015). * = p < .05; ** = p < .01; *** = p < .001.

Discussion

The three judgment tasks were strongly aligned with each other, supporting the validity of these tasks for assessing verbs' instrumental meaning. This is a notable finding, as there are several design and procedural differences between the tasks from Rissman et al. (2015) and the task in Study 2.

As the Study 1 completions aligned with the judgment data in Rissman et al. (2015), so did the Study 2 judgments align with the Study 1 completions (albeit more weakly). This suggests that verb meaning influences semantic categorization and language production in similar ways. The crucial test of this interpretation follows in Study 3, where I analyze sentence completion data for the larger set of verbs from Study 2.

Study 3

Method

Participants I tested 136 adult native English speakers on Amazon Mechanical Turk (F = 60, age range = 20-69; mean age = 39). An additional eight speakers were tested but were excluded for producing repetitive responses (N = 6; e.g., writing only *today* or *again* for each trial), for producing incoherent responses (N = 1), or for self-reporting as not being a native English speaker (N = 1). Participants received $2.50.

Design and Materials. As in Study 1**,** participants viewed sentence fragments with the form [Subject V+PAST *the* Object]. Completions were collected for 172 verbs. These included 132 of the 136 verbs tested in Study 2. The four excluded verbs (e.g., *use*) were not compatible with the syntactic frame of the sentence fragments. Data for the 40 additional verbs are not analyzed in this paper. Each participant provided sentence completions for 94 verbs: 66 instrument verbs and 28 other verbs. Strong and weak verbs were balanced across each participant's sample of 66 instrument verbs. Each participant viewed each verb once and saw a unique, randomly generated sample of the stimuli.

As in Study 1, each verb appeared with six different direct objects (e.g., *Lucy stirred the potion*____). In Study 3, I used a different procedure to generate these direct objects than in Study 1: 83 native English speakers provided sentence completions for fragments with only a subject, verb, and definite determiner (e.g., *Lucy stirred the* _____). Among the object completions that the speakers produced in this separate study, I chose the most frequent direct objects for the sentence fragments for Study 3. For *stir*, for example, the six direct objects were *batter, chili, pot, rice, stew,* and *soup*.

Procedure & Coding The procedure and scheme for coding completions were the same as in Study 1.

Results

Across all 132 verbs, the mean rate of producing instrumental completions ranged from 0% (18 verbs including *eat, move*, *grow*, and *hold*) to 46% (*strike*). For the 24 verbs common to both Studies 1 and 3, rates of

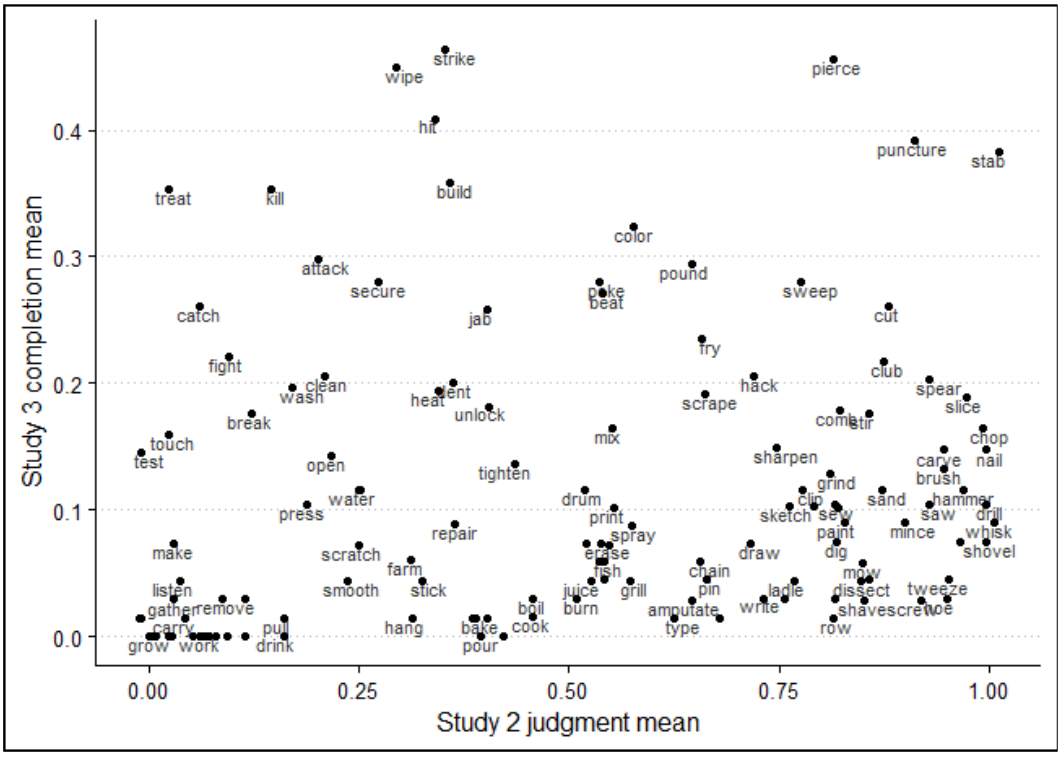

Figure 2**.** Rate of judging that a verb highlights a tool (Study 2) vs. rate of producing an instrument completion (Study 3). Each point indicates an individual verb. To avoid overplotting, not all points are labeled.

producing instrumental completions were highly correlated (r = .85; see Figure 1). Surprisingly, however, the judgments from Study 2 and the completions from Study 3 were only weakly (but significantly) correlated (r = .17). Figure 2 shows a comparison of the judgments and completion rates for each verb. The six verbs that most often elicited instrumental completions were *strike, pierce, wipe, hit, puncture*, and *stab*. While denominal and incision verbs had high instrument judgments, they elicited instrument completions relatively infrequently (9%; 19% of trials, respectively).

To assess whether judgments predict completions, I modeled the probability of producing an instrument completion used mixed effects logistic regression as in Study 1. Judgments from Study 2 predicted instrument completions as both a linear and a quadratic predictor (linear: b = .25, CI_{95} = [.14, .34], z = 4.80, p < .001; quadratic: b = -.23, CI_{95} = [-.32, -.14], z = -4.90, p < .001).

Discussion

The findings from Study 3 replicate the findings from Studies 1 and 2 insofar as instrument judgments and completions were significantly correlated. Nonetheless, the strength of this relationship was much weaker when a larger and more semantically diverse set of verbs was tested.

General Discussion

In this study I investigated how English verbs encode instrumental participants, towards the broader project of understanding whether verbs' instrumental meaning can be modeled in terms of entailment. As different psycholinguistic tasks are known to reflect meaning in different ways, I asked whether the particular semantic organization that is reflected in judgments is also reflected through a language production task, namely sentence completions. All the judgment/completion pairings in Figure 1 resulted in statistically significant correlations, indicating that these two tasks reflect verbal meaning in shared ways. At the same time, the positive relationship between judgments and completions was found to be much weaker when a larger sample of verbs was tested.

A notable finding from Studies 1 and 3 was the significance of the quadratic predictors: that instrument completions were least likely for verbs at the high and low end of the judgment continuum. This likely indicates an interaction between the type of meaning reflected in the judgments and the pragmatics of the production task. Verbs with low instrument judgment ratings tended not to elicit instrument completions, consistent with the linking hypothesis that completions indicate degree of activation (Koenig et al., 2003). Nonetheless, verbs with high instrument ratings also frequently failed to elicit instrument completions, presumably because instrumental meaning was easy to infer and therefore did not need to be produced. This negative correlation emerged not only for denominal verbs such as *rake*—incision verbs were also infrequently followed by instrument completions. These findings indicate a dissociation between comprehension and production: when an instrument can be strongly inferred during sentence comprehension, producers can use the opportunity to provide alternate information that is not strongly inferred.

Turning to the argument status of instruments, Studies 1-3 replicate the gradient cline across verbs observed by Rissman et al. (2015). That is, I observed no clustering of verbs into strongly-instrumental and weakly-instrumental categories (see Figure 2). In addition, instrumentality as measured through judgments was only weakly aligned with instrumentality as measured through sentence completions. Both of these results present a challenge to a formal model where a verb either does or does not entail the presence of an instrument (i.e., where an instrument either is or is not an argument).

Although I observed a gradient pattern across verbs, the pattern was not a chaotic one — both Studies 2 and 3 demonstrated semantic organization concerning which verbs were most likely to elicit instrumental judgments/ completions (even if the semantic organizations were not the same). In Study 2, denominal and incision verbs were most instrumental. In Study 3, verbs of forceful contact such as *strike*, *pierce*, and *hit* were among the most likely to elicit instrument completions. Koenig et al. (2008) analyze this category of verbs as semantically requiring instrument. One interpretation of Study 3 is that these verbs semantically activate instrumental meaning, but not so strongly or specifically that the instrument can be omitted altogether.

The observation that verbs patterned together in semantically well-defined subgroups is important because this suggests components of verb meaning which are stable and abstract. There is ample evidence that some types of word meanings are not well-modeled in terms of entailments — the English preposition *over*, for example, seems to span a chain of meanings rather than lexicalize a core set of entailments (Taylor, 2003). Drawing on such data, a prominent theoretical alternative to an entailment-based approach is a cue-based approach, where words do not "have" meaning in and of themselves but rather act as cues to meaning, in conjunction with a specific context (Elman, 2009). The finding in Studies 2-3 that both judgment and production measures were sensitive to abstract properties of verb meaning, such as incision and forceful contact, is difficult to reconcile with a cue-based approach in which words do not have stable meanings.

Ultimately, the question of whether instruments are arguments is best set aside in favor of more nuanced questions about how words are represented in the mind — how do we represent certain properties, such as instrumentality, in terms of relative degrees of prominence, while also capturing the fact that word meanings are not wholly untethered? And how does word meaning interact with the different pragmatic needs of particular contexts and tasks? This paper points to the need for cognitive scientists to develop alternative models of meaning that can account for the full range of data presented here.

Acknowledgments

This research was supported by NSF-PAC #1734260, NIDCD RO1 DC000491, and a Radboud Excellence Initiative postdoctoral fellowship awarded to Lilia Rissman. Thank you to SarahClaire Ingebritsen for her hard work and diligence in coding this dataset. The suggestions of the anonymous CogSci members who reviewed this paper are much appreciated. Thank you to all study participants.

References

Barbu, R.-M., & Toivonen, I. (2016). Event participants and linguistic arguments. In *Proceedings of the 38th Annual Meeting of the Cognitive Science Society* (pp. 1961-1966).

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

Borer, H. (2005). *Structuring sense: Volume 2. The normal course of events*. Oxford; New York: Oxford University Press.

Davies, M. (2008-). The Corpus of Contemporary American English: 450 million words, 1990-present. from Available online at http://corpus.byu.edu/coca/.

Elman, J. L. (2009). On the Meaning of Words and Dinosaur Bones: Lexical Knowledge Without a Lexicon. *Cognitive Science, 33*(4), 547-582. doi:10.1111/j.1551-6709.2009.01023.x

Koenig, J.-P., Mauner, G., & Bienvenue, B. (2002). Class Specificity and the Lexical Encoding of Participant Information. *Brain and Language, 81*(1-3), 224-235. doi:http://dx.doi.org/10.1006/brln.2001.2519

Koenig, J.-P., Mauner, G., & Bienvenue, B. (2003). Arguments for adjuncts. *Cognition, 89*(2), 67-103. doi:https://doi.org/10.1016/S0010-0277(03)00082-9

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.

Levin, B., & Rappaport-Hovav, M. (2013). Lexicalized meaning and manner/result complementarity. In *Studies in the Composition and Decomposition of Event Predicates* (pp. 49-70): Springer.

Majid, A., Sanford, A. J., & Pickering, M. J. (2007). The linguistic description of minimal social scenarios affects the extent of causal inference making. *Journal of Experimental Social Psychology, 43*(6), 918-932. doi:https://doi.org/10.1016/j.jesp.2006.10.016

Rappaport-Hovav, M., & Levin, B. (1998). Building verb meanings. In M. Butt & W. Geuder (Eds.), *The projection of arguments: Lexical and compositional factors* (pp. 97-134). Stanford, CA: CSLI publications.

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

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

Vater, H. (1978). On the possibility of distinguishing between complements and adjuncts. In *Valence, semantic case, and grammatical relations*. Amsterdam: John Benjamins.

Willits, J. A., Amato, M. S., & MacDonald, M. C. (2015). Language knowledge and event knowledge in language use. *Cognitive psychology, 78*, 1-27.