

(Re)constructing Meaning in NLP

Sean Trott

University of California, San Diego
sttrott@ucsd.edu

Nancy Chang

Google

ncchang@google.com

Tiago Timponi Torrent

Federal University of Juiz de Fora
tiago.torrent@ufjf.edu.br

Nathan Schneider

Georgetown University

nathan.schneider@georgetown.edu

Abstract

Human speakers have an extensive toolkit of ways to express themselves. In this paper, we engage with an idea largely absent from discussions of meaning in natural language understanding—namely, that the way something is expressed reflects different ways of conceptualizing or *construing* the information being conveyed. We first define this phenomenon more precisely, drawing on considerable prior work in theoretical cognitive semantics and psycholinguistics. We then survey some dimensions of construed meaning and show how insights from construal could inform theoretical and practical work in NLP.

1 Introduction

Natural language is a versatile tool for allowing humans to express all manner of communicative intents, from simple descriptions of the entities and situations in their direct experience to elaborate rhetorical flights of fancy. Many NLP applications, such as information extraction, question answering, summarization, and dialogue systems, have restricted their scope to what one might call objective information content—relatively uncontroversial facts that systems can infer from an utterance, store in a database and reason about.

While it is tempting to equate such information with the meaning of an utterance, a large body of literature in linguistics and psycholinguistics argues that an utterance conveys much more than a simple set of facts: it carries with it a halo of intimations arising from the speaker’s choices, including considerations of perspective, emphasis, and framing. That is, *linguistic choices subtly color meaning*; far from merely conveying objective facts, they reflect how speakers conceptualize meaning and affect listeners’ interpretations in predictable ways.

Take, for example, this metaphor-rich portrayal of a newborn as a tyrant over her parental subjects:

- (1) Nora’s arrival brought a regime change. Life under her adorable tyranny was filled with squawking, swaddling and ceaseless sleep-input-output cycles. We were relieved when she relaxed her tiny iron grip.

This report of new parenthood describes a major life change along with everyday caregiver routines, but its emphasis is on the parents’ experience of being suppressed (*under*) and controlled (*grip*) by a creature who is cast, variously, as a tyrant (*regime*), a bird (*squawk*), and a relentless machine (*sleep-input-output cycles, iron grip*)—albeit a (subjectively) adorable one.

The power of linguistic choices to shape understanding is also evident in more mundane (and well-studied) examples:

- (2) a. Chuck *bought* a car *from* Jerry.
Jerry *sold* a car *to* Chuck.
Jerry *paid* Chuck *for* the car.
b. I work *at* Microsoft.
I work *for* Microsoft.
c. The statue *stands* in the plaza.
The statue *is standing* in the plaza.

Each set includes sentences that convey roughly the same facts—i.e. they could describe the same scenario—but nonetheless differ in various respects. The familiar *framing* differences between *buy/sell/pay* (2a) focus attention on different participants and subevents in a commercial transaction. (2b) involves a subtler difference in emphasis, where the choice of *at* highlights the location of the work, while *for* evokes how that work benefits the employer. Grammatical marking can also shift event connotations, as illustrated by the stative vs. temporary contrast in (2c).

Such distinctions illustrate the general phenomenon of **construal**, which we claim has been neglected in NLP. We believe that a proper recog-

nition of construal would provide a unified framework for addressing a wide range of issues involving meaning and linguistic variation, opening the way to systems that more closely approximate (actually) natural language.

This paper surveys the theoretical and empirical landscape related to construal phenomena and makes the case for its relevance to NLP. After clarifying the terms adopted here (§2), we lay out a few key dimensions of construed meaning (§3) and then elaborate on some mechanisms of construal (§4). A trio of case studies illustrate how different types of construal can challenge NLP systems (§5). We end with some conclusions and suggestions for how to begin addressing these challenges (§6).

2 Meaning and construal

Our view of *construal* and its close companion *meaning* is rooted in both frame-based and cognitive semantic traditions. The notion that words and other linguistic units evoke background scenes along with specific perspectives on those scenes is captured by Fillmore’s (1977) slogan, MEANINGS ARE RELATIVIZED TO SCENES. This idea has deeper consequences than merely assigning different semantic roles to examples like (2a). As Langacker (1993, p. 460) observes, “any given situation can be viewed in multiple if not infinitely many ways. Starting from the same basic conceptual content... we can form an endless variety of specific conceptions by making alternate choices in regard to the many dimensions of construal.”

This view of linguistic meaning—which we might call inherently *multivalent*—is more flexible than in many theoretical and computational treatments, particularly truth-conditional approaches that liken meanings to facts in a database. The visual domain offers a more informative analog: a photographic or artistic rendering of a scene can vary in vantage point, viewing distance, objects in sight or in focus, color and lighting choices, etc. (Langacker, 1993; Talmy, 1988). Context matters, too: a painting hanging on a preschool wall may be received differently if displayed in a museum. Just as there is no one objective, context-independent depiction of a scene, there are many valid ways to present an idea through language.

We thus extend Fillmore’s slogan to include all kinds of conceptual content (beyond scenes); the broader communicative context; and the effect of choices made as part of the construal process:

MEANINGS ARE RELATIVIZED TO CONTENT, CONTEXT AND CONSTRUAL.

Below we elaborate on how each of these interrelated factors affects construed meaning.

Conceptual content. We assume that linguistic units can evoke and combine all kinds of conceptual content, including open-ended world knowledge (entities, actions, events, relations, etc.) as well as more schematic structures often associated with grammar and function words. Crucially, concepts must also be amenable to certain kinds of transformation (e.g., shifts in perspective or granularity) as part of construal; see below.¹

Communicative context. We take meaning to encompass scene-level entities and events, discourse-level information about the interlocutors and their communicative intents, and other phenomena straddling the (fuzzy) semantic-pragmatic boundary, related to attention (e.g., profiling and perspective) and conditions of usage falling under what Fillmore (1985) dubbed “U-Semantics” (in contrast to truth-oriented “T-Semantics”).²

Contextual factors (e.g., the interlocutors’ identity, beliefs, goals, conceptual repertoire, cultural backgrounds) can radically alter construed meaning. On this view, meaning is not arbitrarily subjective, or merely intersubjective; it is also constrained by all aspects of the communicative context.

Construal. We define **construal** as a dynamic process of *meaning construction*, in which speakers and hearers encode and decode, respectively, some intended meaning in a given communicative context. To do so, they draw on their repertoire of linguistic and conceptual structures, composing and transforming them to build coherent interpretations consistent with the speaker’s lexical, grammatical, and other expressive choices.³

We take construal to be fundamental to all language use, though how *much* construal and what

¹We are not here concerned with precisely how concepts are represented or learned, since we believe the insights related to construal apply broadly across theoretical frameworks.

²For example, only U-Semantics can explain why “the children are *on* the bus” is preferred over “the children are *in* the bus” if the bus is in transit, despite referring to the same spatial relationship.

³Both speakers and hearers engage in construal: speakers, in choosing how to present the idea, experience or other content they wish to convey; hearers, in reconstructing that intended meaning. Words like ‘analysis’ and ‘interpretation’ should thus be understood as applying to meaning construction by either interlocutor. (We do not focus here on the many differences between comprehension and production.)

kinds of construal vary across interpretations.⁴ In the simplest cases, the relevant components fit neatly together (à la compositional semantics). But many (or even most) utterances involve a myriad of disparate structures—conceptual, linguistic, and contextual—that may need to be transformed, (re)categorized, or otherwise massaged to be integrated into a single coherent whole.

This conceptual flexibility is not arbitrary: the space of combinatorial options is delimited by **construal operations** defined with respect to certain privileged **construal dimensions**. A number of dimensions and operations have been proposed, many motivated by general cognitive processes; we will review some of these in §3, and illustrate how they are engaged during language use in §4.

This inclusive, flexible view of meaning has broad implications for a wide variety of linguistic phenomena, and many parallels in prior work—far too many to address exhaustively here. We restrict our current scope in several ways: (1) While some aspects of context will be mentioned below, we do not address many phenomena related to pragmatic inference (e.g. politeness, indirect requests). (2) Though many construal dimensions are relevant cross-linguistically, we will not address typological patterns in the lexical, grammatical, and cultural conventions that influence construal. (3) We highlight construal phenomena that are psycholinguistically attested and/or relevant to NLP research.

3 Dimensions of construed meaning

Several (partial) taxonomies of construal dimensions have been proposed in the cognitive linguistics literature (Langacker, 1993; Talmy, 1988; Croft and Wood, 2000; Taylor, 1995; Casad, 1995); see Croft and Cruse (2004) for an overview. We will not attempt to reconcile their many differences in terminology and organization, but instead present selected dimensions most relevant for NLP.

3.1 Perspective

Languages have many ways of describing scenes from a specific PERSPECTIVE (or *vantage point*). The spatial domain provides clear examples: a cup might be described as being *left* or *right* of some other object, depending on whose perspective is adopted; or explicitly marked as being *on my/your*

⁴Conventionality plays an important role here: initially creative expressions may require less construal as they become entrenched and their meanings more efficiently accessed.

her/Sue’s left. Likewise, the same motion event can be described relative to differing deictic centers (e.g., the *arrival* in (1) can also be viewed as a departure from the hospital).

Perspective can extend beyond the spatial domain. The use of past tense in (1) indicates the speaker’s retrospective viewpoint. Differences in opinion, belief state or background have also been treated as perspective shifting.

Talmy’s (1988) taxonomy defines a broader version of PERSPECTIVE that includes distribution of attention. Descriptions of a static scene can adopt a dynamic perspective, evoking the experience of moving through the scene (“There is a house every now and then through the valley”); these descriptions can be even more explicit, as with **fictive motion** (“The road runs through the valley”) (Talmy, 1996; Matlock, 2004b).

Psycholinguistic evidence. Grammatical person can affect which perspective a comprehender adopts when reading about an event (Brunyé et al., 2009) and which actions they are most likely to remember (Ditman et al., 2010). Fictive motion can also influence the way comprehenders conceptualize a static scene (Matlock, 2004a,b).

Relevant NLP research. Perspective is crucial for understanding spatial language, e.g. for robotics (§5.2) and other kinds of situated language. Work on grounding referents from natural language descriptions has incorporated visual perspective as another source of information about the intended referent (Devin and Alami, 2016; Ros et al., 2010; Trafton et al., 2005).

3.2 Prominence

PROMINENCE (or *salience*) refers to the relative attention focused on different elements in a scene (Langacker, 1993; Talmy, 1988). Languages have various devices for highlighting, or *profiling*, some elements over others (or leaving them implicit). For example, verbs like those in (2a) differ in which elements in a larger scene are preferentially expressed. Similarly, many spatial and temporal adpositions involve an asymmetric profiling of one entity relative to another; thus “the painting is above the piano” and “the piano is below the painting” describe the same situation but differ in focus.

Verbal and constructional alternations also manipulate prominence: The active/passive pair “Microsoft employs me” and “I am employed by Microsoft” differ in profiling the employer and

speaker, respectively. Similarly, transitive “I rolled the ball” vs. intransitive “The ball rolled” differ in whether the ball-roller is even mentioned.

Languages also differ systematically in how motion events are most idiomatically expressed, in particular in whether the main verb encodes (and foregrounds) the manner (English *run*) or path (Spanish *entrar*) of motion.

Psycholinguistic evidence. A speaker’s decisions about which features to encode in the main verb versus a satellite can influence which events comprehenders find most similar (Billman and Krych, 1998) and which features they tend to remember (Gennari et al., 2002).

In other work, Fausey and Boroditsky (2010) found that descriptions of an accidental event using a transitive construction (“She had ignited the napkin”) led participants to assign more blame to the actor involved, and even demand higher financial penalties, than descriptions using non-agentive constructions (“The napkin had ignited”).

In language production, there are a number of factors influencing which construction a speaker chooses (e.g., current items in discourse focus (Bresnan et al., 2007), lexical and syntactic priming (Pickering and Ferreira, 2008)).

Relevant NLP research. Recovering implicit information is widely studied in NLP, and deciding which information to express is key to NLG and summarization. We mention three examples exploring how choices of form lend prominence to certain facets of meaning in ways that strongly resonate with our claims about construal.

Greene and Resnik (2009) show that syntactic framing—e.g. active (*Prisoner murders guard*) vs. passive (*Guard is murdered*)—is relevant to detecting speaker sentiment about violent events.

Hwang et al. (2017) present an annotation scheme for capturing adpositional meaning construal (as in (2b)). Rather than disambiguate the adposition with a single label, they separately annotate an adposition’s role with respect to a scene (e.g. employment) and the aspect of meaning brought into prominence by the adposition itself (e.g., benefactive *for* vs. locative *at*). This more flexibly accounts for meaning extensions and resolves some annotator difficulties.

Rohde et al. (2018) studied the construction of discourse coherence by asking participants to insert a conjunction (*and, or, but, so, because, before*) where none was originally present, before an

explicit discourse adverbial (e.g. *in other words*). They found that some contexts licensed multiple alternative conjunctions, each expressing a different coherence relation—i.e., distinct implicit relations can be inferred from the same passage. This speaks to the challenge of fully annotating discourse coherence relations and underscores the role of both linguistic and contextual cues in coherence.

3.3 Resolution

Concepts can be described at many levels of RESOLUTION—from highly detailed to more schematic. We include here both *specificity* (e.g., *pug* < *dog* < *animal* < *being*) and *granularity* (e.g., viewing a forest at the level of individual leaves vs. branches vs. trees). Lexical items and larger expressions can evoke and combine concepts at varying levels of detail (“The gymnast triumphantly landed upright” vs. “A person did something”).

Psycholinguistic evidence. Resolution is related to **basic-level categories** (Rosch et al., 1976; Lakoff, 1987; Hajibayova, 2013), the most culturally and cognitively salient levels of a folk taxonomy. Speakers tend to use basic-level terms for reference (e.g., *tree* vs. *entity/birch*), and basic-level categories are more easily and quickly accessed by comprehenders (Mervis and Rosch, 1981; Rosch et al., 1976).

Importantly, however, what counts as basic-level depends on the speaker’s domain expertise (Tanaka and Taylor, 1991). Speakers may deviate from basic-level terms under certain circumstances, e.g., when a more specific term is needed for disambiguation (Graf et al., 2016). Conceptualization is thus a flexible process that varies across both individual cognizers (e.g., as a function of their world knowledge) and specific communicative contexts.

Relevant NLP research. Resolution is already recognized as important for applications such as text summarization and dialogue generation (Louis and Nenkova, 2012; Li and Nenkova, 2015; Ko et al., 2019a; Li et al., 2016; Ko et al., 2019b), e.g., in improving human judgments of informativity and relevance (Ko et al., 2019b). Also relevant is work on knowledge representation in the form of inheritance-based ontologies and lexica (e.g., FrameNet (Fillmore and Baker, 2009), ConceptNet (Liu and Singh, 2004)).

3.4 Configuration

CONFIGURATION refers to internal-structural properties of entities, groups of entities, and events,

indicating their schematic “shape” and “texture”: multiplicity (or *plexity*), homogeneity, boundedness, part-whole relations, etc. (Langacker, 1993; Talmy, 2000). To borrow an example from Croft (2012), a visitor to New England can describe stunning autumn *leaves* or *foliage*. Though both words indicate a multiplex perception, they exhibit a grammatical difference: the (plural) **count** noun *leaves* suggests articulated boundaries of multiple individuals, whereas the **mass** noun *foliage* suggests a more impressionistic, homogeneous rendering.

This dimension includes many distinctions and phenomena related to **aspect** (Vendler, 1967; Comrie, 1976), including whether an event is seen as discrete (*sneeze*) or continuous (*read*); involves a change of state (*leave* vs. *have*); has a defined endpoint (*read* vs. *read a book*); etc. Lexical and grammatical markers of configuration properties interact in complex ways; see discussion of count/mass and aspectual coercion in §4.

Psycholinguistic evidence. Differences in grammatical aspect can modulate how events are conceptualized (Matlock, 2011). Stories written in imperfective aspect are remembered better; participants are also more likely to believe that the events in these stories are still happening (Magliano and Schleich, 2000) and build richer mental simulations of these events (Bergen and Wheeler, 2010). In turn, these differences in conceptualization have downstream consequences, ranging from judgments about an event’s complexity (Wampler and Wittenberg, 2019) to predictions about the consequences of a political candidate’s behavior on reelection (Fausey and Matlock, 2011).

The mass/count distinction has attested psychological implications, including differences in word recognition time (Gillon et al., 1999) (see Fieder et al. (2014) for a review).

Relevant NLP research. Configurational properties are closely linked to well-studied challenges at the syntax-semantic interface, in particular nominal and aspectual *coercion* effects (§4). Several approaches explicitly model coercion operations based on event structure representations (Moens and Steedman, 1988; Passonneau, 1988; Pulman, 1997; Chang et al., 1998), while others explore statistical learning of aspectual classes and features (Siegel and McKeown, 2000; Mathew and Katz, 2009; Friedrich and Palmer, 2014). Lexical resources have also been developed for aspectual annotation (Donatelli et al., 2018) and the count/

mass distinction (Schiehlen and Spranger, 2006; Kiss et al., 2017).

3.5 Metaphor

The dimension of METAPHOR is broadly concerned with *cross-domain comparison*, in which speakers “conceptualize two distinct structures in relation to one another” (Langacker, 1993, p. 450). Metaphors have been analyzed as structured mappings that allow a *target* domain to be conceptualized in terms of a *source* domain (Lakoff and Johnson, 1980).

Metaphors pervade language use, and exhibit highly systematic, extensible structure. For example, in English, events are often construed either as locations in space or as objects moving through space. Our experience of time is thus often described in terms of either motion toward future events (“we’re approaching the end of the year”), or the future moving toward us (“the deadline is barreling towards us”) (Boroditsky, 2000, 2001; Hendricks and Boroditsky, 2017; Núñez and Sweetser, 2006). Metaphor plays a role in our linguistic characterization of many other domains as well (Lakoff and Johnson, 1980).

Psycholinguistic evidence. Different metaphors can shape a comprehender’s representation about the same event or concept in radically different ways. Thibodeau and Boroditsky (2011) found that describing a city’s crime problem as a *beast* or as a *virus* elicited markedly different suggestions about how best to address the problem, e.g., whether participants tended to endorse enforcement- or reform-based solutions. Similar effects of metaphor on event conceptualization have been found across other domains, such as cancer (Hauser and Schwarz, 2015; Hendricks et al., 2018) and climate change (Flusberg et al., 2017) (see Thibodeau et al. (2017) for a thorough review).

Relevant NLP research. Considerable NLP work has addressed the challenge of metaphor detection and understanding (Narayanan, 1999; Shutova et al., 2010, 2013; Shutova, 2015). This work has made use of both statistical, bottom-up approaches to language modeling (Gutiérrez et al., 2016; Shutova et al., 2013), as well as knowledge bases such as MetaNet (Dodge et al., 2015; Stickles et al., 2014; David and Dancygier, 2017).

3.6 Summary

The selective review of construal dimensions presented here is intended to be illustrative, not exhaustive or definitive. Returning to the visual anal-

ogy, we can see these dimensions as primarily concerned with how (and what part of) a conceptual “scene” is *perceived* (PERSPECTIVE, PROMINENCE); the choice or *categorization* of which schematic structures are present (CONFIGURATION and METAPHOR); or both (RESOLUTION).

We have omitted another high-level categorization dimension, SCHEMATIZATION, which includes concepts related to force dynamics, image schemas, and other experientially grounded schemas well discussed in the literature (Talmy, 2000). We have also not addressed pragmatic inference related to politeness (Brown and Levinson, 1987), indirect requests (Clark, 1979), and other aspects of communicative intent. Additionally, some phenomena are challenging to categorize within the dimensions listed here; a more complete analysis would include evidentiality (Chafe and Nichols, 1986), modality (Mortelmans, 2007), light verb constructions (Wittenberg and Levy, 2017; Wittenberg et al., 2014), and more. Nonetheless, we hope this partial taxonomy provides a helpful entry point to relevant prior work and starting point for further alignment.

4 Construal in action

How might construal work in practice? We have emphasized so far the *flexibility* afforded by the dimensions in §3. But we must also explain why some words and concepts make easier bedfellows than others. This section presents a thumbnail sketch of how the construal process copes with apparent mismatches, where it is the collective *constraints* of the input structures that guide the search for coherence.

We focus on comprehension (similar processes apply in production), and assume some mechanism for proposing interpretations consisting of a set of conceptual structures and associated **compatibility constraints**. Compatibility constraints are analogous to various kinds of binding constraints proposed in the literature (variable binding, role-filler bindings, unification bindings, and the like): they are indicators that two structures should be conceptualized as a single unit. But compatibility is softer and more permissive than identity or type-compatibility, in that it can also be satisfied with the help of *construal operations*. Some operations effect relatively subtle shifts in meaning; others have more dramatic effects, including changes to truth-conditional aspects of meaning.

Below we illustrate how some example linguistic

phenomena fit into the sketch just presented and mention connections to prior lines of work.

Count/mass coercion. English nouns are flexible in their **count/mass** status (see §3.4). Atypical marking for number or definiteness can cause a shift, or **coercion**, in boundedness: plural or indefinite marking on mass nouns (*a lemonade*, *two lemonades*) yields a bounded interpretation (cups or bottles of lemonade). Conversely, count nouns with no determiner are coerced to an undifferentiated mass, via a phenomenon known as **grinding** (“there was mosquito all over the windshield”) (Pelletier and Schubert, 1989, 2003; Copestake and Briscoe, 1995). Here we see evidence of the outsize influence of tiny grammatical markers on manipulating lexical defaults in the construal process.

Aspectual composition. Aspect is a prime arena for studying how multiple factors conspire to shape event construal. Verbs are associated with default aspectual classes that can be coerced under pressure from conflicting cues, where details of event structure systematically constrain possible coercions and their inferential consequences (Moens and Steedman, 1988; Talmy, 1988).

In fact, aspectual coercion can be reanalyzed in terms of construal dimensions. For example, *durative* modifiers (e.g. *for an hour*) prefer to combine with **atelic** processes (lacking a defined endpoint, as in 3a) on which to impose a bound (analogous to count/mass coercion) and duration. Combination with any other aspectual class triggers different operations to satisfy that preference:

- (3) a. He {slept / ran} for an hour.
- b. He sneezed for an hour.
- c. He read the book for an hour.
- d. He left for an hour.

A single sneeze, being a discrete event unlikely to last an hour, undergoes **ITERATION** into a series of sneezes (3b), illustrating a change in plicity (§3.4); while the book-reading in in (3c) is simply viewed as unfinished (cf. “He read the book”). The departure in (3d) is a discrete event, but unlike sneezing, it also results in a state change that is reversible and therefore boundable (cf. the iterative reading of “He broke the glass for an hour”, the non-permanent reading of 2c). Its coercion thus features multiple operations: a **PROMINENCE** shift to profile the result state of being gone; and then a **BOUNDING** that also reverses state, implying a return (Chang et al., 1998).

Constructional coercion. The flagship example cited in the construction grammar literature (4a) has also been analyzed as a kind of coercion, serving to resolve conflicts between lexical and grammatical meaning (Goldberg, 1995, 2019):

(4) a. She sneezed the napkin off the table.
 b. She {pushed / blew / sneezed / ?slept} the napkin off the table.

Here, the verb *sneeze*, though not typically transitive or causal, appears in a **Caused Motion** argument structure construction, which pairs oblique-transitive syntax with a caused motion scene. The resulting conflict between its conventional meaning and its putative causal role is resolvable, however, by a commonsense inference that sneezing expels air, which can plausibly cause the napkin’s motion (cf. Forbes and Choi, 2017).

This coercion, also described as *role fusion*, differs from the previous examples in manipulating the PROMINENCE of a latent component of meaning. Coercion doesn’t always succeed, however: presumably sneezing could only move a boulder with contextual support, and sleeping has a less plausibly forceful reading. In fact, construal depends on the interaction of many factors, including degree of conventionality (where *push* and *blow* are prototypical caused motion verbs), embodied and world knowledge (the relative forces of *sneeze* and *sleep* to napkin weight), and context.⁵

There is extensive psycholinguistic evidence of constructional coercion and the many factors influencing ease of construal (see Goldberg (2003, 2019) for reviews). Some of these phenomena have been analyzed within computational implementations of construction grammar (Bergen and Chang, 2005; Bryant, 2008; Bergen and Chang, 2013; Dodge and Petrucc, 2014; Steels, 2017; Steels and Feldman, 2017; Matos et al., 2017), and have also been incorporated in corpus annotation schemes (Bonial et al., 2011; Hwang et al., 2014; Lyngfelt et al., 2018).

Metonymy and metaphor. Metonymy and metaphor are associated with semantic mismatches

⁵A related theory is Dowty’s (1991) *semantic proto-roles* account, which links the grammatical subject/object asymmetry to two clusters of semantic features that are more agent-like (e.g., animacy) or patient-like (e.g., affectedness), respectively; associations between these *proto-roles* and grammatical subjects and objects are attested in comprehension (Kako, 2006; Pyykkönen et al., 2010) and have been investigated computationally (Reisinger et al., 2015; Rudinger et al., 2018).

that trigger construal operations. A possible analysis of *tiny iron grip* from (1) illustrates both.

First, the modifiers *tiny* and *iron* expect a physical entity, but *grip* is a (nominalized) action. This conflict triggers a profile shift (PROMINENCE) to the grip’s effector (a hand), effectively licensing a metonymy. A further conflict arises between the hand and its description as *iron* (unlikely to be literal unless the protagonist is of robotic lineage). A structural alignment (METAPHOR) then maps the iron’s strength to the grip’s force, which in turn maps to the degree of dictatorial control.⁶

We observe that multiple construal operations can occur in sequence; that a conceptual or linguistic element may afford more than one construal within the same analysis (*grip* as both a hand and metaphorical control); and that aspects of common sense, world knowledge, and culture (though not the focus of the present work) inevitably constrain construal options.

5 Case studies

We turn to a few illustrations of how the pervasive effects of construal can arise in applied settings.

5.1 Case study 1: Conversational assistants

Even simple tasks like rescheduling a meeting pose many challenges to dialogue systems, in both understanding users’ intents and formulating natural responses. Consider the following exchange:

U-1: When is my 1-1 with Chuck?
 A-2: 4 PM today, in 15 minutes.
 U-3: Is there another slot soon?
 A-4: Not today, should I check tomorrow?
 U-5: Let’s push it to his tomorrow evening.
 A-6: Rescheduled 1-1 with Chuck for 2 PM tomorrow, 6 PM in Brazil.

The agent’s first response (A-2) demonstrates sensitivity to PERSPECTIVE by providing a relative time. Interpreting “another slot soon” in the user’s follow-up (U-3) requires both understanding that *another* is implicitly defined in contrast to the existing slot (relying on PROMINENCE) and then inferring the appropriate RESOLUTION meant by *soon* (on the scale of hours, rather than minutes or seconds). The agent’s succinct response in (A-4) exploits PROMINENCE yet again, both by eliding reference to the sought-after open meeting slot with

⁶Alternatively, *iron grip* could be treated as an entrenched idiom with a readily accessible construal that *tiny* can modify.

Chuck, and by using “tomorrow” (the direct object of “check”) as a metonymic shorthand for the joint constraints of the user’s and Chuck’s calendars.

The next user turn (U-5) employs METAPHOR in its construal of an event as a physical object, capable of being pushed. The metaphorical destination (“his tomorrow evening”) requires consideration of differing time zones (PERSPECTIVE), as made explicit in the final agent turn (A-6).

Interactions between situational context and the kinds of compatibility constraints discussed in §4 can also affect a dialogue system’s best response. A user asking a fitness tracking app “How long have I been running?” while panting around a track may be referring to the current run, but the same question asked while sitting at home is more likely wondering how long they’ve been *habitually* running. A successful response requires the integration of the constraints from (at least): the verb *running*, whose progressive marking is associated with ongoing processes, but ambiguous between a single run and a series of runs (CONFIGURATION); the present-perfect *have been V-ing*, which implies an internal view (PERSPECTIVE); and the situational context (is the user currently running?).

5.2 Case study 2: Human-robot interaction

Situated interactions between humans and robots require the integration of language with other modalities (e.g., visual or haptic).⁷ Clearly, any spatially grounded referring expressions must be tailored to the interlocutors’ PERSPECTIVE (whether shared or not) (Kunze et al., 2017).

Focus of attention (PROMINENCE) is especially important for systems that must interpret procedural language. Recipes, for example, are notoriously telegraphic, with rampant omissions of information that a human cook could easily infer in context (Ruppenhofer and Michaelis, 2010; Malmaud et al., 2014). Consider (5):

(5) *In* a medium bowl, cream *together* the sugar and butter. Beat *in* the eggs, one at a time, then stir *in* the vanilla.

The italicized words provide crucial constraints that would help a cook (human or robot) track the evolving spatial relations. The first *in* establishes

⁷Indeed, the needs of human-robot interaction have motivated extensions to Abstract Meaning Representation (Bănarescu et al., 2013) beyond predicate-argument structure and entities to capture tense and aspect, spatial information, and speech acts (Bonial et al., 2019).

the bowl as the reference point for the creaming action, whose result—the mixture of sugar and butter *together*—becomes the implicit landmark for the subsequent beating *in* of eggs and vanilla.

Systems following instructions also require a means of segmenting continuous sensorimotor data and linking it to discrete linguistic categories (Regnieri et al., 2013; Yagcioglu et al., 2018) (cf. the *symbol grounding problem* (Harnad, 1990)). This mapping may depend on flexibly adjusting RESOLUTION and CONFIGURATION based on linguistic cues (e.g., *cut/dice/slice/sliver the apple*).

5.3 Case study 3: Paraphrase generation

Despite many advances, paraphrase generation systems remain far from human performance. One vexing issue is the lack of evaluation metrics that correlate with human judgments for tasks like paraphrase, image captioning, and textual entailment (see, e.g., Bhagat and Hovy, 2013; Pavlick and Kwiatkowski, 2019; Wang et al., 2019b).

In particular, it is unclear how closely a good paraphrase should hew to *all* aspects of the source sentence. For example, should active/passive descriptions of the same scene, or the sets of sentences in (2), be considered meaning-equivalent? Or take the putative paraphrase below:

(6) a. The teacher sat on the student’s left.
b. Next to the children was a mammal.

These could plausibly describe the same scene; should their differences across multiple dimensions (PERSPECTIVE, PROMINENCE, RESOLUTION) be rewarded or penalized for this diversity?

A first step out of this quandary is to recognize construal dimensions and operations as a source of linguistic variability. Paraphrase generation and other semantically oriented tasks could incorporate these into system design and evaluation in task-specific ways.

6 Discussion

Throughout this paper, we have emphasized the flexible and multivalent nature of linguistic meaning, as evidenced by the **construal** phenomena described here. The effects of construal are ubiquitous: from conventional to creative language use, through morphemes and metaphors. Indeed, even the smallest forms can, like tiny tyrants, exert a transformative force on their surroundings, inducing anything from a subtle shift in emphasis to a

radical reconceptualization.

As illustrated in §5, this flexibility of language use poses a challenge for NLP practitioners. Yet crucially—and fortunately—construal is not random: **variations in linguistic form correspond systematically to differences in construal**. The dimensions of construal and their associated operations (§3 and §4) offer principled constraints that render the search for coherence more tractable.

How, then, should we proceed? Our goal is for construal dimensions such as those highlighted in §3 to be incorporated into any research program aspiring to human-level linguistic behavior. Below, we describe several concrete recommendations for how to do this.

More meaningful metrics. Taking construal seriously means rethinking how NLP tasks are designed and evaluated. Construal dimensions can provide a rubric for assessing tasks, datasets, and meaning representations (Abend and Rappoport, 2017) for which meaningful distinctions they make or require. (E.g.: Does it capture the level of RESOLUTION at which entities and events are described? Does it represent METAPHOR? Is it sensitive to the PROMINENCE of different event participants?)

Such questions might also help guard against unintended biases like those recently found in NLP evaluations and systems (e.g., Caliskan et al., 2017; Gururangan et al., 2018). Popular NLU benchmarks (like SuperGLUE; Wang et al., 2019a) should be critically examined for potential *construal biases*, and contrasts should be introduced deliberately to probe whether systems are modeling lexical choices, grammatical choices, and meaning in the desired way (Naik et al., 2018; Kaushik et al., 2020; McCoy et al., 2019; Gardner et al., 2020).

As a broader suggestion, datasets should move away from a one-size-fits-all attitude based on gold annotations. Ideally, evaluation metrics should take into account not only partial structure matches, but also similarity to alternate construals.

Cognitive connections. The many connections between construal and the rest of cognition highlight the need for further interdisciplinary engagements in the study of construal.

The psycholinguistics literature is a particularly rich source of construal-related data and human language benchmarks. Psycholinguistic data could also be used to probe neural language models (Futrell et al., 2018; Linzen and Leonard, 2018; van Schijndel and Linzen, 2018; Ettinger, 2020).

How well do such models capture the phenomena reviewed in §3, and where do they fall short?

A fuller account of the constellation of factors involved in construal should also take seriously the grounded, situated nature of language use (Har-nad, 1990; Kiros et al., 2018; Bender and Koller, 2020; Bisk et al., 2020). Frameworks motivated by the linguistic insights mentioned in §2 (such as the work on *computational construction grammar* referenced in §4) and by growing evidence of *embodied simulations* as the basis for meaning (Narayanan, 1999; Bergen and Chang, 2005; Feldman, 2006; Bergen, 2012; Tamari et al., 2020) are especially relevant lines of inquiry.

Much work remains to flesh out the construal dimensions, operations and phenomena preliminarily identified in §3 and §4, especially in connecting to typological, sociolinguistic, developmental, and neural constraints on conceptualization. We believe a concerted effort across the language sciences would provide valuable guidance for developing better NL systems and resources.

7 Conclusion

As the saying goes, the camera doesn’t lie—but it may tell us only a version of the truth. The same goes for language.

Some of the phenomena we have described may seem, at first glance, either too subtle to bother with or too daunting to tackle. But we believe it is both timely and necessary, as language technologies grow in scope and prominence, to seek a more robust treatment of meaning. We hope that a deeper appreciation of the role of construal in language use will spur progress toward systems that more closely approximate human linguistic intelligence.

Acknowledgments

We are grateful to Lucia Donatelli, Nick Hay, Aurelie Herbelot, Jena Hwang, Jakob Prange, Susanne Riehemann, Hannah Rohde, Rachel Rudinger, and anonymous reviewers for many helpful suggestions; and to the ACL 2020 organizers for planning a special theme, *Taking Stock of Where We’ve Been and Where We’re Going*. Special thanks to Nora Chang-Hay for finally relaxing her tiny iron grip.

This research was supported in part by NSF award IIS-1812778. The FrameNet Brasil Lab is funded by CAPES grants 88887.125411/2016-00 and 88887.144043/2017-00.

References

Omri Abend and Ari Rappoport. 2017. [The state of the art in semantic representation](#). In *Proc. of ACL*, pages 77–89, Vancouver, Canada.

Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. [Abstract Meaning Representation for sembanking](#). In *Proc. of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 178–186, Sofia, Bulgaria.

Emily M. Bender and Alexander Koller. 2020. [Climbing towards NLU: On meaning, form, and understanding in the age of data](#). In *Proc. of ACL*.

Benjamin Bergen and Nancy Chang. 2013. [Embodied Construction Grammar](#). In Thomas Hoffmann and Graeme Trousdale, editors, *The Oxford Handbook of Construction Grammar*, pages 168–190. Oxford University Press, New York.

Benjamin Bergen and Kathryn Wheeler. 2010. [Grammatical aspect and mental simulation](#). *Brain and Language*, 112(3):150–158.

Benjamin K. Bergen. 2012. *Louder Than Words: The New Science of How the Mind Makes Meaning*. Perseus Books Group, New York.

Benjamin K. Bergen and Nancy Chang. 2005. [Embodied Construction Grammar in simulation-based language understanding](#). In Jan-Ola Östman and Mirjam Fried, editors, *Construction grammars: cognitive grounding and theoretical extensions*, pages 147–190. John Benjamins, Amsterdam.

Rahul Bhagat and Eduard Hovy. 2013. [What is a paraphrase?](#) *Computational Linguistics*, 39(3):463–472.

Dorrit Billman and Meredyth Krych. 1998. [Path and manner verbs in action: Effects of “skipping” or “exiting” on event memory](#). In *Proc. of CogSci*, volume 20, pages 156–161, Madison, Wisconsin.

Yonatan Bisk, Ari Holtzman, Jesse Thomason, Jacob Andreas, Yoshua Bengio, Joyce Chai, Mirella Lapata, Angeliki Lazaridou, Jonathan May, Aleksandr Nisnevich, Nicolas Pinto, and Joseph Turian. 2020. [Experience grounds language](#). *arXiv:2004.10151 [cs]*.

Claire Bonial, Susan Windisch Brown, Jena D. Hwang, Christopher Parisien, Martha Palmer, and Suzanne Stevenson. 2011. [Incorporating coercive constructions into a verb lexicon](#). In *Proc. of the ACL 2011 Workshop on Relational Models of Semantics*, pages 72–80, Portland, Oregon, USA.

Claire Bonial, Lucia Donatelli, Stephanie M. Lukin, Stephen Tratz, Ron Artstein, David Traum, and Clare Voss. 2019. [Augmenting Abstract Meaning Representation for human-robot dialogue](#). In *Proc. of the First International Workshop on Designing Meaning Representations*, pages 199–210, Florence, Italy.

Lera Boroditsky. 2000. [Metaphoric structuring: Understanding time through spatial metaphors](#). *Cognition*, 75(1):1–28.

Lera Boroditsky. 2001. [Does language shape thought?: Mandarin and English speakers’ conceptions of time](#). *Cognitive Psychology*, 43(1):1–22.

Joan Bresnan, Anna Cueni, Tatiana Nikitina, and R. Harald Baayen. 2007. [Predicting the dative alternation](#). In Gerlof Bouma, Irene Kraemer, and Joost Zwarts, editors, *Cognitive foundations of interpretation*, pages 69–94. KNAW, Amsterdam.

Penelope Brown and Stephen C. Levinson. 1987. *Politeness: Some universals in language usage*, volume 4. Cambridge University Press.

Tad T. Brunyé, Tali Ditman, Caroline R. Mahoney, Jason S. Augustyn, and Holly A. Taylor. 2009. [When you and I share perspectives: Pronouns modulate perspective taking during narrative comprehension](#). *Psychological Science*, 20(1):27–32.

John Bryant. 2008. *Best-fit constructional analysis*. Ph.D. dissertation, University of California, Berkeley, Berkeley, California.

Aylin Caliskan, Joanna J. Bryson, and Arvind Narayanan. 2017. [Semantics derived automatically from language corpora contain human-like biases](#). *Science*, 356(6334):183–186.

Eugene Casad. 1995. [Seeing it in more than one way](#). In John R. Taylor, editor, *Language and the Cognitive Construal of the World*, pages 23–49. Mouton de Gruyter, Berlin.

Wallace L. Chafe and Johanna Nichols. 1986. *Evidentiality: The linguistic coding of epistemology*, volume 20. Ablex Publishing Corporation, Norwood, NJ.

Nancy Chang, Daniel Gildea, and Srinivas Narayanan. 1998. [A dynamic model of aspectual composition](#). In *Proc. of CogSci*, pages 226–231, Madison, WI, USA.

Herbert H. Clark. 1979. [Responding to indirect speech acts](#). *Cognitive Psychology*, 11(4):430–477.

Bernard Comrie. 1976. *Aspect: An introduction to the study of verbal aspect and related problems*, volume 2. Cambridge University Press, New York.

Ann Copestake and Ted Briscoe. 1995. [Semi-productive polysemy and sense extension](#). *Journal of Semantics*, 12(1):15–67.

William Croft. 2012. *Verbs: Aspect and Causal Structure*. Oxford University Press, Oxford, UK.

William Croft and D. Alan Cruse. 2004. Conceptualization and construal operations. In *Cognitive Linguistics*, chapter 3. Cambridge University Press.

William Croft and Esther J. Wood. 2000. Construal operations in linguistics and artificial intelligence. In Liliana Albertazzi, editor, *Meaning and Cognition: A multidisciplinary approach*, pages 51–78. John Benjamins, Amsterdam.

Oana David and Barbara Dancygier. 2017. Computational approaches to metaphor: the case of MetaNet. In Barbara Dancygier, editor, *The Cambridge Handbook of Cognitive Linguistics*, pages 574–589. Cambridge University Press, Cambridge.

Sandra Devin and Rachid Alami. 2016. An implemented theory of mind to improve human-robot shared plans execution. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 319–326. IEEE.

Tali Ditman, Tad T. Brunyé, Caroline R. Mahoney, and Holly A. Taylor. 2010. Simulating an enactment effect: Pronouns guide action simulation during narrative comprehension. *Cognition*, 115(1):172–178.

Ellen Dodge, Jisup Hong, and Elise Stickles. 2015. MetaNet: Deep semantic automatic metaphor analysis. In *Proc. of the Third Workshop on Metaphor in NLP*, pages 40–49, Denver, Colorado, USA.

Ellen K. Dodge and Miriam R. L. Petrucc. 2014. Representing caused motion in Embodied Construction Grammar. In *Proc. of the ACL 2014 Workshop on Semantic Parsing*, pages 39–44, Baltimore, MD.

Lucia Donatelli, Michael Regan, William Croft, and Nathan Schneider. 2018. Annotation of tense and aspect semantics for sentential AMR. In *Proc. of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018)*, pages 96–108, Santa Fe, New Mexico, USA.

David R. Dowty. 1991. Thematic proto-roles and argument selection. *Language*, 67(3):547–619.

Allyson Ettinger. 2020. What BERT is not: Lessons from a new suite of psycholinguistic diagnostics for language models. *Transactions of the Association for Computational Linguistics*, 8:34–48.

Caitlin M. Fausey and Lera Boroditsky. 2010. Subtle linguistic cues influence perceived blame and financial liability. *Psychonomic Bulletin & Review*, 17(5):644–650.

Caitlin M. Fausey and Teenie Matlock. 2011. Can grammar win elections? *Political Psychology*, 32(4):563–574.

Jerome A. Feldman. 2006. *From molecule to metaphor: a neural theory of language*. MIT Press, Cambridge, MA.

Nora Fieder, Lyndsey Nickels, and Britta Biedermann. 2014. Representation and processing of mass and count nouns: A review. *Frontiers in Psychology*, 5:589.

Charles J. Fillmore. 1977. The case for case reopened. In Peter Cole and Jerrold M. Sadock, editors, *Syntax and Semantics, vol. 8: Grammatical Relations*, pages 59–81. Academic Press, New York.

Charles J. Fillmore. 1985. Frames and the semantics of understanding. *Quaderni di Semantica*, 6(2):222–254.

Charles J. Fillmore and Collin Baker. 2009. A frames approach to semantic analysis. In Bernd Heine and Heiko Narrog, editors, *The Oxford Handbook of Linguistic Analysis*, pages 791–816. Oxford University Press, Oxford, UK.

Stephen J. Flusberg, Teenie Matlock, and Paul H. Thibodeau. 2017. Metaphors for the war (or race) against climate change. *Environmental Communication*, 11(6):769–783.

Maxwell Forbes and Yejin Choi. 2017. Verb Physics: relative physical knowledge of actions and objects. In *Proc. of ACL*, pages 266–276, Vancouver, Canada.

Annemarie Friedrich and Alexis Palmer. 2014. Automatic prediction of aspectual class of verbs in context. In *Proc. of ACL*, pages 517–523, Baltimore, Maryland, USA.

Richard Futrell, Ethan Wilcox, Takashi Morita, and Roger Levy. 2018. RNNs as psycholinguistic subjects: Syntactic state and grammatical dependency. *arXiv preprint arXiv:1809.01329*.

Matt Gardner, Yoav Artzi, Victoria Basmova, Jonathan Berant, Ben Bogin, Sihao Chen, Pradeep Dasigi, Dheeru Dua, Yanai Elazar, Ananth Gottumukkala, Nitish Gupta, Hanna Hajishirzi, Gabriel Ilharco, Daniel Khashabi, Kevin Lin, Jiangming Liu, Nelson F. Liu, Phoebe Mulcaire, Qiang Ning, Sameer Singh, Noah A. Smith, Sanjay Subramanian, Reut Tsarfaty, Eric Wallace, Ally Zhang, and Ben Zhou. 2020. Evaluating NLP models via contrast sets. *arXiv:2004.02709 [cs]*.

Silvia P. Gennari, Steven A. Sloman, Barbara C. Malt, and W. Tecumseh Fitch. 2002. Motion events in language and cognition. *Cognition*, 83(1):49–79.

Brendan Gillon, Eva Kehayia, and Vanessa Taler. 1999. The mass/count distinction: Evidence from on-line psycholinguistic performance. *Brain and Language*, 68(1-2):205–211.

Adele E. Goldberg. 1995. *Constructions: A construction grammar approach to argument structure*. University of Chicago Press, Chicago.

Adele E. Goldberg. 2003. Constructions: A new theoretical approach to language. *Trends in Cognitive Sciences*, 7(5):219–224.

Adele E. Goldberg. 2019. *Explain Me This: Creativity, Competition, and the Partial Productivity of Constructions*. Princeton University Press, Princeton.

Caroline Graf, Judith Degen, Robert X.D. Hawkins, and Noah D. Goodman. 2016. **Animal, dog, or dalmatian? Level of abstraction in nominal referring expressions**. In *Proc. of CogSci*, pages 2261–2266, Philadelphia, PA.

Stephan Greene and Philip Resnik. 2009. **More than words: syntactic packaging and implicit sentiment**. In *Proc. of NAACL-HLT*, pages 503–511, Boulder, Colorado.

Suchin Gururangan, Swabha Swayamdipta, Omer Levy, Roy Schwartz, Samuel Bowman, and Noah A. Smith. 2018. **Annotation artifacts in natural language inference data**. In *Proc. of NAACL-HLT*, pages 107–112, New Orleans, Louisiana.

E. Dario Gutiérrez, Ekaterina Shutova, Tyler Marghetis, and Benjamin Bergen. 2016. **Literal and metaphorical senses in compositional distributional semantic models**. In *Proc. of ACL*, pages 183–193, Berlin, Germany.

Lala Hajibayova. 2013. **Basic-level categories: A review**. *Journal of Information Science*, 39(5):676–687.

Stevan Harnad. 1990. **The symbol grounding problem**. *Physica D: Nonlinear Phenomena*, 42(1):335 – 346.

David J. Hauser and Norbert Schwarz. 2015. **The war on prevention: Bellicose cancer metaphors hurt (some) prevention intentions**. *Personality and Social Psychology Bulletin*, 41(1):66–77.

Rose K. Hendricks and Lera Boroditsky. 2017. **New space–time metaphors foster new nonlinguistic representations**. *Topics in Cognitive Science*, 9(3):800–818.

Rose K. Hendricks, Zsófia Demjén, Elena Semino, and Lera Boroditsky. 2018. **Emotional implications of metaphor: Consequences of metaphor framing for mindset about cancer**. *Metaphor and Symbol*, 33(4):267–279.

Jena D. Hwang, Archna Bhatia, Na-Rae Han, Tim O’Gorman, Vivek Srikumar, and Nathan Schneider. 2017. **Double trouble: the problem of construal in semantic annotation of adpositions**. In *Proc. of *SEM*, pages 178–188, Vancouver, Canada.

Jena D. Hwang, Annie Zaenen, and Martha Palmer. 2014. **Criteria for identifying and annotating caused motion constructions in corpus data**. In *Proc. of LREC*, pages 1297–1304, Reykjavík, Iceland.

Edward Kako. 2006. **Thematic role properties of subjects and objects**. *Cognition*, 101(1):1–42.

Divyansh Kaushik, Eduard Hovy, and Zachary Lipton. 2020. **Learning the difference that makes a difference with counterfactually-augmented data**. In *Proc. of ICLR*.

Jamie Kiros, William Chan, and Geoffrey Hinton. 2018. **Illustrative language understanding: large-scale visual grounding with image search**. In *Proc. of ACL*, pages 922–933, Melbourne, Australia.

Tibor Kiss, Francis Jeffry Pelletier, Halima Husić, and Johanna Poppek. 2017. **Issues of mass and count: Dealing with ‘dual-life’ nouns**. In *Proc. of *SEM*, pages 189–198, Vancouver, Canada.

Wei-Jen Ko, Greg Durrett, and Junyi Jessy Li. 2019a. **Domain agnostic real-valued specificity prediction**. In *Proc. of AAAI*, volume 33, pages 6610–6617.

Wei-Jen Ko, Greg Durrett, and Junyi Jessy Li. 2019b. **Linguistically-informed specificity and semantic plausibility for dialogue generation**. In *Proc. of NAACL-HLT*, pages 3456–3466, Minneapolis, Minnesota.

Lars Kunze, Tom Williams, Nick Hawes, and Matthias Scheutz. 2017. **Spatial referring expression generation for hri: Algorithms and evaluation framework**. In *2017 AAAI Fall Symposium Series*, pages 27–35, Palo Alto, CA.

George Lakoff. 1987. *Women, fire, and dangerous things: what categories reveal about the mind*. University of Chicago Press, Chicago.

George Lakoff and Mark Johnson. 1980. *Metaphors We Live By*. University of Chicago Press, Chicago.

Ronald W. Langacker. 1993. **Universals of construal**. In *Proc. of Berkeley Linguistics Society*, volume 19, pages 447–463.

Junyi Jessy Li and Ani Nenkova. 2015. **Fast and accurate prediction of sentence specificity**. In *Proc. of AAAI*, pages 2281–2287, Austin, Texas.

Junyi Jessy Li, Bridget O’Daniel, Yi Wu, Wenli Zhao, and Ani Nenkova. 2016. **Improving the annotation of sentence specificity**. In *Proc. of LREC*, pages 3921–3927, Portorož, Slovenia.

Tal Linzen and Brian Leonard. 2018. **Distinct patterns of syntactic agreement errors in recurrent networks and humans**. In *Proc. of CogSci*, pages 690–695, Madison, WI.

Hugo Liu and Push Singh. 2004. **ConceptNet—a practical commonsense reasoning tool-kit**. *BT Technology Journal*, 22(4):211–226.

Annie Louis and Ani Nenkova. 2012. **A corpus of general and specific sentences from news**. In *Proc. of LREC*, pages 1818–1821, Istanbul, Turkey.

Benjamin Lyngfelt, Lars Borin, Kyoko Ohara, and Tiago Timponi Torrent. 2018. *Constructicography: Constructicon development across languages*. John Benjamins, Amsterdam.

Joseph P. Magliano and Michelle C. Schleich. 2000. **Verb aspect and situation models.** *Discourse Processes*, 29(2):83–112.

Jonathan Malmaud, Earl Wagner, Nancy Chang, and Kevin Murphy. 2014. **Cooking with semantics.** In *Proc. of the ACL 2014 Workshop on Semantic Parsing*, pages 33–38, Baltimore, MD.

Thomas A. Mathew and E. Graham Katz. 2009. **Supervised categorization for habitual versus episodic sentences.** In *Sixth Midwest Computational Linguistics Colloquium*, Bloomington, Indiana.

Teenie Matlock. 2004a. The conceptual motivation of fictive motion. In *Studies in Linguistic Motivation*, pages 221–248. Mouton de Gruyter, Berlin.

Teenie Matlock. 2004b. **Fictive motion as cognitive simulation.** *Memory & Cognition*, 32(8):1389–1400.

Teenie Matlock. 2011. **The conceptual motivation of aspect.** In Klaus-Uwe Panther and Gunter Radden, editors, *Motivation in Grammar and the Lexicon*, pages 133–148. John Benjamins Publishing, Amsterdam.

Ely Matos, Tiago Torrent, Vânia Almeida, Adrieli Lavola, Ludmila Lage, Natália Marcao, and Tatiane Tavares. 2017. **Constructional analysis using constrained spreading activation in a FrameNet-based structured connectionist model.** In *Computational Construction Grammar and Natural Language Understanding: Papers from the 2017 AAAI Spring Symposium*, pages 222–229, Stanford, California. AAAI Press.

Tom McCoy, Ellie Pavlick, and Tal Linzen. 2019. **Right for the wrong reasons: Diagnosing syntactic heuristics in natural language inference.** In *Proc. of ACL*, pages 3428–3448, Florence, Italy.

Carolyn B. Mervis and Eleanor Rosch. 1981. **Categorization of natural objects.** *Annual Review of Psychology*, 32(1):89–115.

Marc Moens and Mark Steedman. 1988. **Temporal ontology and temporal reference.** *Computational Linguistics*, 14(2):15–28.

Tanja Mortelmann. 2007. **Modality in cognitive linguistics.** In *The Oxford Handbook of Cognitive Linguistics*. Oxford University Press.

Aakanksha Naik, Abhilasha Ravichander, Norman Sadeh, Carolyn Rose, and Graham Neubig. 2018. **Stress test evaluation for natural language inference.** In *Proc. of COLING*, pages 2340–2353, Santa Fe, New Mexico, USA.

Srinivas Narayanan. 1999. **Moving right along: a computational model of metaphoric reasoning about events.** In *Proc. of AAAI*, pages 121–128, Orlando, Florida.

Rafael E. Núñez and Eve Sweetser. 2006. **With the future behind them: convergent evidence from Aymara language and gesture in the crosslinguistic comparison of spatial construals of time.** *Cognitive Science*, 30(3):401–450.

Rebecca J. Passonneau. 1988. **A computational model of the semantics of tense and aspect.** *Computational Linguistics*, 14(2):44–60.

Ellie Pavlick and Tom Kwiatkowski. 2019. **Inherent disagreements in human textual inferences.** *Transactions of the Association for Computational Linguistics*, 7:677–694.

Francis Jeffry Pelletier and Lenhart K. Schubert. 1989. **Mass expressions.** In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic: Volume IV: Topics in the Philosophy of Language*, Synthese Library, pages 327–407. Springer Netherlands, Dordrecht.

Francis Jeffry Pelletier and Lenhart K. Schubert. 2003. **Mass expressions.** In D. M. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, Handbook of Philosophical Logic, pages 249–335. Springer Netherlands, Dordrecht.

Martin J. Pickering and Victor S. Ferreira. 2008. **Structural priming: A critical review.** *Psychological Bulletin*, 134(3):427.

Stephen Pulman. 1997. **Aspectual shift as type coercion.** *Transactions of the Philological Society*, 95.

Pirita Pyykkönen, Danielle Matthews, and Juhani Järvikivi. 2010. **Three-year-olds are sensitive to semantic prominence during online language comprehension: A visual world study of pronoun resolution.** *Language and Cognitive Processes*, 25(1):115–129.

Michaela Regneri, Marcus Rohrbach, Dominikus Wetzel, Stefan Thater, Bernt Schiele, and Manfred Pinkal. 2013. **Grounding action descriptions in videos.** *Transactions of the Association for Computational Linguistics*, 1:25–36.

Drew Reisinger, Rachel Rudinger, Francis Ferraro, Craig Harman, Kyle Rawlins, and Benjamin Van Durme. 2015. **Semantic proto-roles.** *Transactions of the Association for Computational Linguistics*, 3:475–488.

Hannah Rohde, Alexander Johnson, Nathan Schneider, and Bonnie Webber. 2018. **Discourse coherence: concurrent explicit and implicit relations.** In *Proc. of ACL*, pages 2257–2267, Melbourne, Australia.

Raquel Ros, Séverin Lemaignan, E. Akin Sisbot, Rachid Alami, Jasmin Steinwender, Katharina Hamann, and Felix Warneken. 2010. **Which one? grounding the referent based on efficient human-robot interaction.** In *19th International Symposium in Robot and Human Interactive Communication*, pages 570–575. IEEE.

Eleanor Rosch, Carolyn B. Mervis, Wayne Gray, David Johnson, and Penny Boyes-Braem. 1976. **Basic objects in natural categories.** *Cognitive Psychology*, 8(3):382–439.

Rachel Rudinger, Adam Teichert, Ryan Culkin, Sheng Zhang, and Benjamin Van Durme. 2018. **Neural-Davidsonian semantic proto-role labeling.** In *Proc. of EMNLP*, pages 944–955, Brussels, Belgium.

Josef Ruppenhofer and Laura A. Michaelis. 2010. **A constructional account of genre-based argument omissions.** *Constructions and Frames*, 2(2):158–184.

Michael Schiehlen and Kristina Spranger. 2006. **The mass-count distinction: acquisition and disambiguation.** In *Proc. of LREC*, pages 265–270, Genoa, Italy.

Marten van Schijndel and Tal Linzen. 2018. **Modeling garden path effects without explicit hierarchical syntax.** In *Proc. of CogSci*, pages 2603–2608, Madison, WI.

Ekaterina Shutova. 2015. **Design and evaluation of metaphor processing systems.** *Computational Linguistics*, 41(4):579–623.

Ekaterina Shutova, Lin Sun, and Anna Korhonen. 2010. **Metaphor identification using verb and noun clustering.** In *Proc. of Coling*, pages 1002–1010, Beijing, China.

Ekaterina Shutova, Simone Teufel, and Anna Korhonen. 2013. **Statistical metaphor processing.** *Computational Linguistics*, 39(2):301–353.

Eric V. Siegel and Kathleen R. McKeown. 2000. **Learning methods to combine linguistic indicators: Improving aspectual classification and revealing linguistic insights.** *Computational Linguistics*, 26(4):595–628.

Luc Steels. 2017. **Basics of Fluid Construction Grammar.** *Constructions and Frames*, 9(2):178–255.

Luc Steels and Jerome Feldman, editors. 2017. *Computational Construction Grammar and Natural Language Understanding: Papers from the 2017 AAAI Spring Symposium*. AAAI Press, Stanford, California.

Elise Stickles, Ellen Dodge, and Jisup Hong. 2014. A construction-driven, MetaNet-based approach to metaphor extraction and corpus analysis. Presented at Conceptual Structure, Discourse, and Language (CSDL 12), Santa Barbara, California.

Leonard Talmy. 1988. Grammatical construal. In Brygida Rudzka-Ostyn, editor, *Topics in Cognitive Linguistics*, pages 165–205. John Benjamins, Amsterdam.

Leonard Talmy. 1996. Fictive motion in language and “ception”. In Paul Bloom, Mary A. Peterson, Lynn Nadel, and Merrill F. Garrett, editors, *Language and space*, pages 211–276. The MIT Press, Cambridge, MA.

Leonard Talmy. 2000. *Toward a cognitive semantics: concept structuring systems*. MIT Press, Cambridge, MA.

Ronen Tamari, Chen Shani, Tom Hope, Miriam R. L. Petrucci, Omri Abend, and Dafna Shahaf. 2020. **Language (re)modelling: Towards embodied language understanding.** In *Proc. of ACL*.

James W. Tanaka and Marjorie Taylor. 1991. **Object categories and expertise: Is the basic level in the eye of the beholder?** *Cognitive Psychology*, 23(3):457–482.

John R. Taylor. 1995. Introduction: On construing the world. In John R. Taylor, editor, *Language and the Cognitive Construal of the World*, pages 1–22. Mouton de Gruyter, Berlin.

Paul H. Thibodeau and Lera Boroditsky. 2011. **Metaphors we think with: the role of metaphor in reasoning.** *PLoS ONE*, 6(2):e16782.

Paul H. Thibodeau, Rose K. Hendricks, and Lera Boroditsky. 2017. **How linguistic metaphor scaffolds reasoning.** *Trends in Cognitive Sciences*, 21(11):852–863.

J. Gregory Trafton, Nicholas L. Cassimatis, Magdalena D. Bugajska, Derek P. Brock, Farilee E. Mintz, and Alan C. Schultz. 2005. **Enabling effective human-robot interaction using perspective-taking in robots.** *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 35(4):460–470.

Zeno Vendler. 1967. *Linguistics in Philosophy*. Cornell University Press, Ithaca, NY.

Joshua Wampler and Eva Wittenberg. 2019. Doing thus and so: Event referential expressions and referent complexity. Presented at California Meeting on Psycholinguistics (CAMP) 3, UC Santa Cruz.

Alex Wang, Yada Pruksachatkun, Nikita Nangia, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel R. Bowman. 2019a. **SuperGLUE: A stickier benchmark for general-purpose language understanding systems.** In *Proc. of NeurIPS*, pages 3266–3280, Vancouver, Canada.

Su Wang, Rahul Gupta, Nancy Chang, and Jason Baldridge. 2019b. **A task in a suit and a tie: paraphrase generation with semantic augmentation.** In *Proc. of AAAI*, volume 33, pages 7176–7183, Honolulu, Hawaii.

Eva Wittenberg, Ray Jackendoff, Gina Kuperberg, Martin Paczynski, Jesse Snedeker, and Heike Wiese. 2014. The processing and representation of light

verb constructions. In Asaf Bachrach, Isabelle Roy, and Linnaea Stockall, editors, *Structuring the argument*, pages 61–80. John Benjamins, Amsterdam.

Eva Wittenberg and Roger Levy. 2017. **If you want a quick kiss, make it count: How choice of syntactic construction affects event construal.** *Journal of Memory and Language*, 94:254–271.

Semih Yagcioglu, Aykut Erdem, Erkut Erdem, and Nazli Ikizler-Cinbis. 2018. **RecipeQA: a challenge dataset for multimodal comprehension of cooking recipes.** In *Proc. of EMNLP*, pages 1358–1368, Brussels, Belgium.