Language (Re)modelling: Towards Embodied Language Understanding

Ronen Tamari[†]   **Chen Shani**[†]   **Tom Hope**[⋆*]
Miriam R. L. Petruck[‡]   **Omri Abend**[†]   **Dafna Shahaf**[†]

[†]The Hebrew University of Jerusalem
[⋆]Allen Institute for Artificial Intelligence
[*]Paul G. Allen School of Computer Science & Engineering, University of Washington
[‡]International Computer Science Institute, Berkeley, CA
{ronent,chenxshani,oabend,dshahaf}@cs.huji.ac.il
tomh@allenai.org   miriamp@icsi.berkeley.edu

Abstract

While natural language understanding (NLU) is advancing rapidly, today's technology differs from human-like language understanding in fundamental ways, notably in its inferior efficiency, interpretability, and generalization. This work proposes an approach to representation and learning based on the tenets of embodied cognitive linguistics (ECL). According to ECL, natural language is inherently executable (like programming languages), driven by mental simulation and metaphoric mappings over hierarchical compositions of structures and schemata learned through embodied interaction. This position paper argues that the use of grounding by metaphoric inference and simulation will greatly benefit NLU systems, and proposes a system architecture along with a roadmap towards realizing this vision.

1 Introduction

> "Not those speaking the same language, but those sharing the same feeling understand each other." – *Jalal ad-Din Rumi*

While current NLU systems "speak" human language by learning strong statistical models, they do not possess anything like the rich mental representations that people utilize for language understanding. Indeed, despite the tremendous progress in NLU, recent work shows that today's state-of-the-art (SOTA) systems differ from human-like language understanding in crucial ways, in particular in their generalization, grounding, reasoning, and explainability capabilities (Glockner et al., 2018; McCoy et al., 2019a,b; Nie et al., 2019; Yogatama et al., 2019; Lake et al., 2019).

Question-answering (QA) is currently one of the predominant methods of training deep-learning models for general, open-domain language understanding (Gardner et al., 2019b). While QA is a versatile, broadly-applicable framework, recent studies have shown it to be fraught with pitfalls (Gardner et al., 2019a; Mudrakarta et al., 2018). A recent workshop on QA for reading comprehension suggested that "There is growing realization that the traditional supervised learning paradigm is broken [...] – we're fitting artifacts" (Gardner, 2019).

In many respects, the problems of NLU mirror those of artificial intelligence (AI) research in general. Lake et al.'s (2017a) seminal work identified a significant common factor at the root of problems in general AI. The current deep-learning paradigm is a statistical pattern-recognition approach predominantly applied to relatively narrow *task-specific prediction*. In contrast, human cognition supports a wide range of inferences (planning, action, explaining, etc.), hinting at a view of intelligence focused on *model-building*, specifically, mental models: rich, structured, manipulable, and explainable representations useful for performing in dynamic, uncertain environments. This distinction motivates the quest for a new **cognitively-inspired model-building** learning paradigm for general AI, which has inspired fruitful subsequent research and discussion (e.g., Lake et al. (2017b)).

The observation that NLU and general AI share a common central problem (task-specific prediction-based learning), and the growing realization that deeper text understanding requires building mental models (Gardner et al., 2019a; Forbes et al., 2019), motivate the search for an *NLU analog* of the cognitively-inspired model building paradigm.

Amid recent position papers highlighting significant differences between human language understanding and current NLU systems (McClelland et al., 2019; Bisk et al., 2020), here we take a more focused look at mental models; challenges arising due to their embodied nature, their importance in general NLU, and how we might begin integrating them into current approaches.

Mainstream NLU work, be it entirely distributional, such as BERT (Devlin et al., 2018), or also involving symbolic knowledge representation (Liu et al., 2019a; Bosselut et al., 2019), seldom addresses mental models directly. Crucially, such approaches lack the interactive *worlds* within which mental models[1] are learned jointly through language and embodied action. The most closely related lines of work to the present proposal are grounded approaches, which feature worlds in the form of interactive environments, and address mapping text to programs (executable semantic parses) (e.g., Gauthier and Mordatch, 2016; Liang, 2016; Kiela et al., 2016; Chevalier-Boisvert et al., 2019). However, while well-aligned with a model-building paradigm, typically such approaches have been limited to short or synthetic literal language and narrow domains assuming predefined environments. Embodied approaches to *general* NLU, as advocated here, are few and far between. Mostly, examples fall under the construction grammar framework (Steels and de Beule, 2006; Bergen and Chang, 2005). However, despite their intellectual merit, they were not operationalized to scale readily for mainstream applications (see §3).

This position paper argues that executable semantic parsing and grounded approaches to NLU constitute a first step in a much larger program, whose outline is set forth, for *general* language understanding through **embodied cognitive linguistics (ECL)**. Following much cognitive science research (see §3, §4), this paper posits that (1) **execution or simulation** is a central part of semantics, essential for addressing some of the persistent difficulties in text understanding, and (2) **metaphoric inference capabilities** are central to knowledge representation, and facilitate grounded understanding of general language. Importantly, capacities for both simulation and metaphor are emergent, borne of embodied interaction within an external world.

Our contributions are: we analyze inherent limitations of SOTA statistical language models applied to NLU and propose a framework to address these limitations. The novelty of this approach stems from bringing together ideas from the cognitive science literature, the general-AI community, and NLU. This framework constitutes a path to generalize current execution-based methods towards more general language understanding.

---

[1]Typically, mental models are construed as "world simulators"; see §3.

The world contains 2 crates. Each crate contains 4 boxes. Oranges and apples are objects. Each box may contain up to 5 objects. Objects can be moved from one box to another. Objects can be removed from boxes or crates. There are two apples in the first box in the first crate. There is one orange and one apple in the second box of the second crate. First, the apples were transfered from the first box of the first crate to the first box of the second crate. Next, all apples were removed from the second crate.

Initial World State

Figure 1: Open-domain challenge – a world with boxes, crates and objects.

This paper proposes a system architecture and a roadmap towards implementing the vision outlined here, suggesting preliminary directions for future work (learned world models, incorporating interaction into datasets). We believe that this framework will facilitate consolidation with multiple related lines of research across the different communities.

2 Challenges for Current NLU Systems

This section presents concrete example problems demonstrating inherent limitations in SOTA NLU.

2.1 Open-domain Literal Language Simulation

Fig. 1 includes a short story about a world with crates, boxes, and objects inside them. It is a short and simple narrative, far from capturing the full-blown complexity of natural language. Following Gardner et al. (2019a), we assume that a system understands the story if it can correctly answer arbitrary questions about it. To do so requires basic commonsense and mathematical reasoning, referent grounding, tracking events, handling declarative knowledge, and more.

The task is similar to narrative comprehension tasks in datasets such as bAbI (Bordes et al., 2015) and SCONE (Long et al., 2016), and could be solved given large amounts of annotated training data. But, the goal here is different, specifically, to develop models that, like humans, can understand such language **on-the-fly** (like zero-shot learning).

QA approaches. Current QA systems, used in an off-the-shelf manner, do not generalize well to tasks on which they have not been trained; NLU models are known to be brittle even to slight changes in style and vocabulary (Gardner et al.,

2020; Keysers et al., 2020). The closest QA setting is the DROP challenge (Dua et al., 2019), requiring reading comprehension and basic numerical reasoning over paragraphs. As a simple sanity check, we tested a near-SOTA model and baseline[2] on this example, asking questions about the initial and final state. The models were notably better answering questions about the initial state than about the final state. This result is perhaps expected, as the answers to questions about the initial state are closer to the input text. Answering questions about later states is more challenging. A key missing component of these systems is the ability to **simulate** the effects of actions, especially commonsense effects (e.g., moving a container moves the elements in it).

Executable semantic parsing approaches. The problem of Fig. 1 could also naturally be cast as an executable semantic parsing (ex. SP) task. Similar tasks already exist, for example, the "Alchemy" sub-task of the SCONE dataset features beakers of chemicals that are mixed, poured, and drained. Executable approaches can leverage simulation to learn structured world models, but are limited by hard-coded, domain-specific executors; adding tasks requires substantial manual effort.

For humans, through largely subconscious metaphorical inference (related to transfer and meta-learning in general AI (Lake et al., 2017a)), it is obvious that both SCONE and Fig. 1 share much the same structure. This similarity allows for effortless generalization, effectively re-purposing a relatively simple executor (for literal language) flexibly across many tasks.

2.2 Non-literal Language

The previous challenge involved literal language, amenable to symbolic execution. However, non-literal language is pervasive in everyday speech (Lakoff and Johnson, 1980). Consider the example in Fig. 2: the phrase "head of the French Army" is non-literal, implying that the army can be treated as a human body. The execution semantics of verbs like "attacked" and "defend" are also non-literal; they are highly contextual, requiring interpretation beyond word-sense disambiguation alone. "Russian hackers attacked the Pentagon networks" or "The senator attacked the media" entail very different *simulations*. This ambiguity is challenging for non-neural (symbolic) simulation-based approaches. Humans compose a structured

---

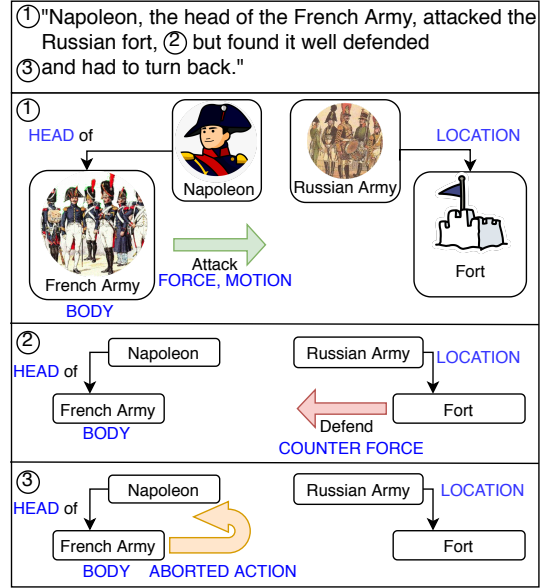[2]Segal et al. (2019) and Dua et al. (2019), respectively.

Figure 2: Non-literal language challenge. To understand this sentence, humans rely on metaphoric inference over embodied concepts (in blue, also called schema; see §3). For example, here "attack" evokes a FORCE or MOTION schema, used to construct a mental model of the scene via mental simulation (§4).

mental model from the language through schemata and mental simulation, as discussed in §3,§4.

To summarize, the limitations outlined above motivate the attempt to extend the capability of simulation to general linguistic inputs. Doing so would enable the construction of grounded, manipulable, and interpretable representations from text. Two desiderata follow from the challenges: (1) more flexible utilization of symbolic executors by exploiting shared (analogical) structures between texts (§2.1), and (2) learned, neural executors for non-literal language comprehension (§2.2).

3 Embodied Cognitive Linguistics: A Model Building Paradigm

Turning to cognitive science for inspiration, we focus on embodied cognitive linguistics (ECL), an important paradigm directly addressing both desiderata. This section presents a brief overview and key tenets of ECL, specifically the theoretical foundations Lakoff and Johnson (1980) and Feldman and Narayanan (2004) developed. Most contemporary cognitive accounts of language incorporate concepts from ECL to some degree. A full review is out of scope of this work; see Gärdenfors (2014) and §4,§5 for discussion in the NLU context.

Early cognitive theories assumed a disembod-

ied, symbolic representation of knowledge (Lewis, 1976; Kintsch and Van Dijk, 1978), separate from the brain's modal systems (vision, motor control, etc.). In contrast, the embodied cognition (EC) view, based on widespread empirical findings, focuses on the role of the body in cognition. In this view, knowledge is stored using multimodal representations (mental imagery, memories, etc.) that arise from embodied experience and action in the world (Barsalou, 2008; Proffitt, 2006). ECL postulates that linguistic representations and other, higher-level cognitive functions are deeply grounded in neural modal systems (Lakoff and Johnson, 1980; Barsalou, 2008). This view is compelling, as it addresses the grounding problem (Harnad, 1990) by linking between high-level symbolic constituents of mental representations and experience or action in the physical world (Varela et al., 2017). Note that embodiment is far from an end-all for language comprehension: for example, social and cultural aspects too are crucial (Arbib et al., 2014). Still, ECL laid important conceptual foundations also underlying subsequent accounts:

- **Embodied schemata:** Pre-linguistic structures formed from bodily interactions and recurring experience, such as CONTAINMENT, PART-WHOLE, FORCE, MOVEMENT (Langacker, 1987; Talmy, 1985, 1983).
- **Metaphoric inference:**[3] The process by which new information may be inferred via structural similarities to a better-understood instantiated system (Lakoff and Johnson, 1980; Gallese and Lakoff, 2005; Day and Gentner, 2007). For example, "I have an example IN mind" suggests that the abstract concept *mind* is mapped to the more concrete domain of *containers*.
- **Mental simulation.** The reenactment of perceptual, motor, and introspective states acquired during experience with the world, body, and mind. In EC, diverse simulation mechanisms (also called mental or forward models (Rumelhart et al., 1986; Grush, 2004)) support a wide spectrum of cognitive activities, including language and decision making (Barsalou, 2008).

We believe that ECL is a useful paradigm for addressing the challenges of §2, as it articulates the role of analogy and mental simulation in NLU. The following two ECL hypotheses summarize them (Lakoff and Johnson, 1980; Feldman and Narayanan, 2004):

Hypothesis 1 (Simulation): *Humans understand the meaning of language by mentally simulating its content. Language in context evokes a simulation structured by embodied schemata and metaphoric mappings, utilizing the same neural structures for action and perception in the environment. Understanding involves inferring and running the best fitting simulation.*

Hypothesis 2 (Metaphoric Representation): *Human concepts are expressible through hierarchical, compositional, metaphoric mappings over a limited vocabulary of embodied schema. Abstract concepts are expressed using more literal concepts.*

Early ECL Implementations. Early attempts to implement ECL in actual language understanding systems were founded on Narayanan (1997)'s x-schema simulation framework and Embodied Construction Grammar (Bergen and Chang, 2005). While notable for approaching challenging problems involving mental simulation, and complex, metaphoric language, early implementation efforts were not operationalized to scale to mainstream applications (Lakoff and Narayanan, 2010). These works also focused on a particular type of simulation (sensorimotor), understood as only one mechanism of many used in language understanding (Stolk et al., 2016).

FrameNet (Ruppenhofer et al., 2016) and MetaNet (David and Dodge, 2014) are closely related projects in that each provides an extensive collection of schemata used in everyday and metaphoric language comprehension, respectively, via the concept of a semantic frame (Fillmore, 1985). However, neither incorporates simulation semantics, as needed for a full realization of the ECL vision (Chang et al., 2002).

4 Linking ECL to NLU and Embodied AI Research

We propose a unifying view of ECL, bringing it closer to contemporary cognitive science and deep learning approaches. This section presents notations and motivating intuitions, further developing the computational framework in §5,§6. The proposal centers around the view of natural language as a kind of *neural programming language* (Lupyan and Bergen, 2016), or higher-level cognitive control system for systematically querying and inducing changes in the mental and physical states of

---

[3]Also called analogical reasoning, we use "metaphorical" and "analogical" interchangeably.

Concept	Symbolic	ECL	Embodied AI
Primitives	Basic data structures, operators, variables...	Schemata: MOVE, CONTAINER, PART-WHOLE...	Deep neural world & action representations (learned through interaction)
Knowledge Organization	a) Composition, inheritence b) Libraries	a) Hierarchical, compositional metaphoric mappings b) Compiled Knowledge	
Executable Unit	Instruction	Semantic parse	$\tilde{a}$
Execution Trace	Intermediate program states	Mental models	$\tilde{T}(\tilde{s}, \tilde{a})$
Simulation	Executor	Emulator[†]	$\tilde{T}$
Semantic parsing / grounding	Parser to executable symbolic program	Parser to executable neural program	O^{-1}, π

Table 1: *Natural language as a neural programming language* conceptualization, with correspondence between symbolic programming, ECL, and embodied AI, using standard POMDP notation. Tilde notation refers to internal counterparts of T, s, a used in mental simulation.
[†]Also called mental simulation (Bergen and Chang, 2005), we adopt *emulator* (Glenberg, 2008) to conform with contemporary cognitive science accounts.

recipients (Elman, 2004; Stolk et al., 2016; Borghi et al., 2018). This approach builds on the ECL hypotheses and suggests a broader view of mental simulation, one that is readily amenable to the same computational formulation as current embodied AI and executable semantic parsing approaches.

Preliminaries. At the core of embodied approaches is the Partially Observable Markov Decision Process (POMDP; Kaelbling et al., 1998). It governs the relations between states (s), actions (a), observations (o), and rewards (r). Of particular interest are the recognition $O^{-1} : O \rightarrow S$, policy $\pi : S \rightarrow A$, and transition $T : S \times A \rightarrow S$ functions. Focusing on mental simulation rather than actual external action, we assume a degree of equivalence between external and internal representations (Rumlehart et al., 1986; Hamrick, 2019). We consider internal mental states and actions ($\tilde{s}, \tilde{a}$), effecting change to mental models via a learned **neural emulator** $\tilde{T}$ (Grush, 2004). Finally, language is considered a form of action (Glenberg, 2008) via external and internal utterances (i.e., semantic parses).

Connecting symbolic & embodied language understanding. Table 1 presents a structured version of the neural programming language conceptualization. Importantly, this view highlights the important commonalities and differences between ECL and both **symbolic programming languages**, as well as **embodied neural mechanisms**, for perception and action. We illustrate these relations more explicitly through a comparison between ECL and executable semantic parsing (Table 1, bottom).

Executable semantic parsing. Involves parsing a novel linguistic input o into a symbolic program **a**, whose execution[4] yields a desired goal state: $T\left(O^{-1}(o), \mathbf{a}\right) = s^*$. Executable semantic parsing focuses on action in an external, symbolic environment T, and typically doesn't address $\tilde{T}$, e.g., mapping a natural language question o directly to an executable query **a** on an SQL engine T.

ECL semantic parsing. Shares the same structure as executable semantic parsing, with the important distinction that simulation is enacted via internal neural representations: $\tilde{T}\left(O^{-1}(o), \tilde{\mathbf{a}}\right) = \tilde{s}^*$. The fully neural formulation enables grounded understanding of non-literal language, demonstrated here for the Fig. 2 example. **Metaphoric inference (hyp. 2)** facilitates parsing a novel linguistic input o into **internal, structured, neural** state representations $\tilde{s}, \tilde{a}$. Accordingly, the utterance u="Napoleon, the head of the French Army" might be parsed to an internal state $\tilde{s}$ composed of a PART-WHOLE schema as shown in the figure. The phrase "attacked the Russian fort" could be grounded to a parse $\tilde{a}$ driving simulation over MOTION and FORCE schemata. The requirement that $\tilde{s}$ and $\tilde{a}$ should afford **mental simulation (hyp. 1)** by the neural world emulator $\tilde{T}$ marks an important difference from current neural word embeddings, one that contributes to deeper language understanding; in the resulting mental model $\tilde{T}(\tilde{s}, \tilde{a})$, Napoleon and the French Army likely moved *together* due to the PART-WHOLE relation between them. This inference is non-trivial since it requires implicit

---

[4]Slightly abusing notation, we apply T iteratively on a sequence of actions $\mathbf{a} = (a_0, ..., a_{L-1})$.

knowledge (heads and bodies often move together). Indeed, a SOTA NLI model[5] considers it "very likely" that the Fig. 2 sentence contradicts the entailment that "The French Army moved towards the fort but did not enter it." To summarize:

- Executable semantic parsing approaches address grounding **literal** language to **symbolic** primitives; and metaphoric inference suggests a mechanism for grounding **general** language using **neural** primitives (schemata).
- Executable semantic parsing approaches utilize **hard-coded, external symbolic executors**, whereas ECL highlights the role of **learned neural world emulators**, as in current embodied research AI efforts (see §7.2).

5 Proposal for an Embodied Language Understanding Model

Formalizing the view characterized above suggests a novel computational model of language understanding. While current statistical models focus on the *linguistic signal*, research shows that most of the relevant information required for understanding a linguistic message is not present in the words (Stolk et al., 2016; David et al., 2016). Accordingly, the ECL view suggests shifting the focus to the *mental models* that communicators use, and the neural mechanisms used to construct them, e.g., mental simulation.

What follows here adapts a relevant cognitive-inspired framework from general AI to the present NLU setting (§5.1), and discusses computational challenges (§5.2). Note that similar insights have been applied to multi-agent communication problems (Andreas et al., 2017), but their application to general NLU has been limited.

5.1 Formal Framework

The recently introduced Consciousness Prior (CP; Bengio, 2017) is a framework to represent the mental model of a single agent, through the notion of *abstract state representations*.[6] Here, an abstract state corresponds with $\tilde{s}$ (§4), a low-dimensional, structured, interpretable state encoding, useful for planning, communication, and predicting upcoming observations (François-Lavet et al., 2019). One example is a dynamic knowledge graph embedding to represent a scene (Kipf et al., 2020).

---

[5] We use Liu et al. (2019b) with https://demo.allennlp.org/textual-entailment/.

[6] For brevity we omit discussion of deriving abstract states from the full mental state, see Bengio (2017) for details.

We adapt CP to a two-player cooperative linguistic communication setting (Tomasello, 2008). We assume a communicator (A) and recipient (B), as shown in Fig. 3. The computational problem of communicators is a "meeting of minds" (Gärdenfors, 2014), or achieving some **alignment** of their **mental models** (Rumelhart, 1981; Stolk et al., 2016): the communicator A wishes to induce in B some (possibly ordered) set of goal abstract states $\mathcal{G}^*$.

We leave exploration of the communicator side to future work, and focus here on understanding. We assume that A sequentially generates utterances $u_t \in \mathcal{U}$ (we assume equivalence between utterances u and observations o) using an utterance model (Bengio, 2017). Analogously, B uses a *comprehension* model C s.t., $\tilde{s}_t = C(\tilde{s}_{t-1}, u_t)$. We assume that alignment is possible: there exists some sequence of utterances that will induce $\mathcal{G}^*$.

This framework is readily applicable to static text (reading comprehension). For example, in Fig. 1, $\mathcal{G}^*$ would be the sequence of desired states, and each sentence corresponds to an utterance ($u_1 =$"The world contains 2 crates.",...).

5.2 Computational challenges of embodiment

We can now more precisely characterize the challenges that the recipient faces. At the root of the problem is the *embodiment principle* (Lawrence, 2017): human internal representations and computation capacity, as represented by $\tilde{s}$ and $\tilde{T}$, respectively, are many orders of magnitude larger than their linguistic communication "bandwidth". We note that though $\tilde{s}_t$ is only a subspace of the full mental state, following Stolk et al. (2016); Bengio (2017) we assume that it still holds that $\dim(\tilde{s}_t) \gg \dim(u_t)$. The embodiment principle dictates extreme economy in language use (Grice et al., 1975), and results in three major challenges:

Common ground (prior world knowledge). Meaning cannot be spelled out in words but rather must be **evoked** in the listener (Rumelhart, 1981) by assuming and exploiting common ground (Clark and Schaefer, 1989; Tomasello, 2008), i.e., shared structures of mental representations. In other words, to achieve some aligned goal state g^*, the communicators must rely heavily on pre-existing similarities in $\tilde{s}$, $\tilde{a}$, and $\tilde{T}$. Developing computational versions of human world models ($\tilde{T}$) is likely AI-complete or close, but useful middle ground may be attained by partial approximations.
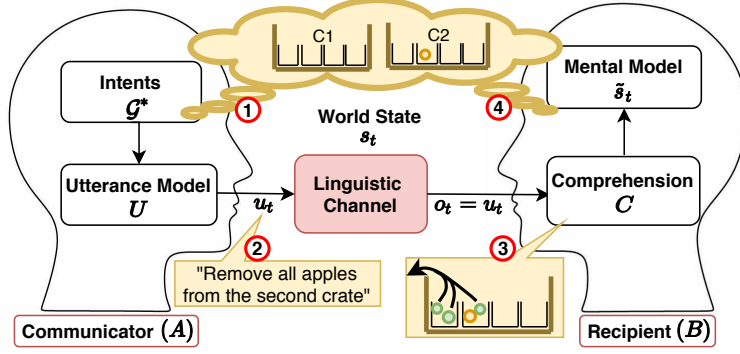
Figure 3: Schema of linguistic communication framework. Communicator's intent (1) is a high dimensional mental state, i.e., remove apples from the second crate. The low capacity of the linguistic channel (2) leaves the burden of understanding primarily on Communicator and Recipient (embodiment principle). The Recipient's goal is to understand (3), i.e., reconstruct the intent by integrating linguistic input, knowledge of the state of the world, and internal knowledge (memories, commonsense). Reconstruction results in a successful alignment (4).

Common ground (discourse). In the context of discourse, new information must be accumulated efficiently to update the mental model (Clark and Schaefer, 1989; Stolk et al., 2016). Consider "Remove all apples from the second crate" (Figure 1). Full comprehension is only possible in the context of a sufficiently accurate mental model. Using our previous notations, the comprehension of u_t depends both on the previous utterances $u_{1:(t-1)}$ and intermediate mental model $\tilde{s}_{t-1}$.

Abstract vs. Literal Language. Interpretation of literal language is relatively straightforward – it is the language first acquired by children, directly related to the physical world. However, much of human language is more abstract, relying on *metaphors* borne of embodiment. The symbolic programming analog fails for utterances like "these elections seem like a circus". Symbolic programming languages cannot handle non-literal interpretations: *how* are `elections` like a `circus`? This is related to *selective analogical inference* (Gentner and Forbus, 2011), closely related to ECL: not everything in the source domain (circus) is mapped to the target (elections). Humans easily perceive the salient metaphoric mappings (`clown`→`candidate`), but this feat remains extremely complex for machines.

6 Architecture Sketch

This section presents a schematic ECL-inspired architecture towards the implementation of the **comprehension model** (C), which addresses the challenges presented in §5.2. Fig. 4 shows the proposed architecture. For simplicity, the focus is on a static

reading comprehension setting, but the architecture supports richer environments as well.

6.1 Environment

The environment provides an "interaction API" to the agent, as well as the reward signal. The supported interaction may vary considerably depending on the task; for reading comprehension, it allows structured access to the text while supporting flexible reading strategies (Yuan et al., 2019). The flexibility is important for long documents, where navigation may be required (Geva and Berant, 2018). For executable semantic parsing, there might be external systems to interact with besides the text, such as a database (Liang et al., 2016).

6.2 Agent

The agent architecture approximates the important ECL functions outlined in §4, and consists of four main modules:

Memory. We distinguish between two forms of memory, the first an episodic, short-term **mental model** – the system's current abstract state representation ($\tilde{s}_t$). The symbolic programming analog is the execution trace of a program, containing the states of relevant working variables at each execution step. Fig. 4 displays the updated mental model, after the removal of the apples. **Compiled knowledge**, or long-term memory, reflects highly familiar object representations, behaviors and schemata, such as common sense, intuitive psychology and physics. The symbolic programming language analogs of this are libraries; largely static, hierarchical and compositional repositories of functions and
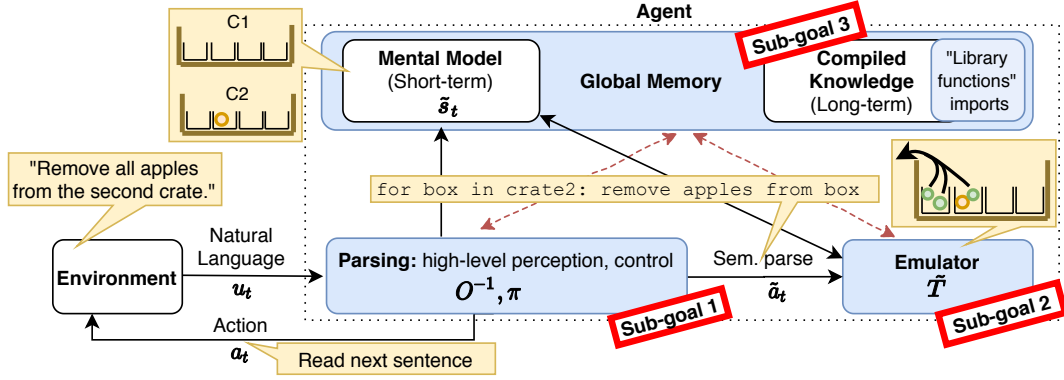
Figure 4: Architecture for comprehender (§5), demonstrated on a symbolic version of the example task of Fig. 1. The agent receives natural language input from the environment. The agent has global memory – short-term, keeping track of the mental model of the world, and long-term, containing compiled knowledge ("library classes and functions"). The parser interprets input to parse $\tilde{a}_t$ enacting mental simulation using emulator. The mental model is then updated, ready for the next input. The sub-goals refer to the order in which components are learned (as opposed to hard-coded) in our proposed roadmap (§7).

classes. In the course of language interpretation, these libraries are "importable": for the symbolic example in Fig. 4, the parser might instantiate a new variable of an imported type (e.g., `crate2 = Container()`). Both types of memory are accessible for all components of the agent.

Parser. Abstraction of higher-level perception, control, reasoning and linguistic functions. Handles interpretation of new linguistic inputs based on prior knowledge and the current mental state. Consonant with the view of analogy-making as a kind of higher-level perception or recognition (Mitchell, 1993), **metaphoric inference** is involved in grounding a novel input u_t into **internal, neural** state representations $\tilde{s}_t, \tilde{a}_t$ affording simulation. See Fig. 4 and Fig. 2 for examples on literal and non-literal language, respectively.

Emulator. Functionally similar to the executor module in executable semantic parsing, but learned, and obviously far greater in scale. This module is an abstraction of neural emulation mechanisms ($\tilde{T}$), representing a wide range of functions, from lower-level motor control and imagery to higher-level models used for planning and theory of mind (Grush, 2004). It operates over the current mental model and semantic parse from the parser. The output is then an updated mental model.

Importantly, the proposed architecture is designed to address the challenges outlined in §5.2; compiled knowledge underlies **human common ground**, the building blocks of $\tilde{s}, \tilde{a}$ and $\tilde{T}$. Memory

and emulation are instrumental for **accumulation in discourse**. The ability to understand **abstract language** involves all modules in the system.

7 Implementation Roadmap

The architecture outlined in §6 is very ambitious; its implementation requires much further research. This section proposes a *roadmap* to this goal, identifying three sub-goals (Fig. 4), presented in order of increasing difficulty. Broadly speaking, the level of difficulty is determined by which components are assumed as *given* in the input (here this also means they are hard-coded in a symbolic programming language), and which must be *learned*.

7.1 Sub-goal 1: learning open-domain simulation

Observing that literal language is close to the embodied primitives level, its interpretation is simpler (than that of non-literal language, see §4). Therefore, in this phase, the emulator and compiled knowledge are hard-coded; here the focus is learning the parser. In other words, this sub-goal focuses on extending executable semantic parsing from relatively narrow domains to handle more general *literal language* on-the-fly, similarly to zero-shot semantic parsing (Givoli and Reichart, 2019).

For the example in §2.1, the parser could be expected to infer the types (boxes as containers, fruits as objects) either by context (Yao et al. (2018) explore a preliminary schema-based approach) or

explicit declarative language, using them to **configure the emulator** to handle the specific required problem setting (Tamari et al., 2020).

As in similar projects exploring embodied understanding (Pustejovsky and Krishnaswamy, 2016; Baldridge et al., 2018), new simulator frameworks must be developed. While full embodiment calls for multiple modalities, the degree to which it is required remains an important open question (Lupyan and Lewis, 2019). Accordingly, and for immediate applicability to purely textual NLU problems we propose also focusing on the simpler setting of interactive text (Nelson, 2005). Recent research on text-based games shows how agents can learn to "program" in such languages (Côté et al., 2019; Ammanabrolu and Riedl, 2019), and how real language understanding problems can be framed as executable semantic parsing using configurable text-based simulators (Tamari et al., 2019).

7.2 Sub-goal 2: learning to simulate

This phase assumes that the compiled knowledge is given (hard-coded), and the parsing and emulator modules are neural (learned). A hard-coded emulator will likely be needed to train a learned emulator. The learned event execution of Narayanan (1997) provides a useful starting point towards computational models capable of such inference. In general, learned simulation is relatively unexplored in the context of natural language, though recent work has explored it in a "blocks-world" instruction following setup (Gaddy and Klein, 2019). Outside of NLU, learning structured world models is a long-studied, fast-growing field in embodied AI research (Schmidhuber, 1990; Ha and Schmidhuber, 2018; Hamrick, 2019; Anand et al., 2019; Kipf et al., 2020), and recently also in learned executors for neural programming (Kant, 2018). We expect much useful cross fertilization with these fields.

7.3 Sub-goal 3: learning compiled knowledge

This phase focuses on the component seemingly hardest to learn – compiled knowledge. Out of scope here is fully neural setting where all components are jointly learned, as in continual learning research (Parisi et al., 2019). Instead, we focus on a simpler setting, in which the compiled knowledge is learned but represented by symbolic code; i.e., learning the static code library underlying the simulation framework. This sub-goal is relevant for training the parser (§7.1) as well as the emulator (§7.2), and can be pursued in parallel to them.

In this setting, learning compiled knowledge is closely related to automated knowledge base construction (Winn et al., 2019) or frame induction from text (QasemiZadeh et al., 2019). Our proposed paradigm suggests enriching classic symbolic knowledge representations (Speer et al., 2017) to executable form (Tamari et al., 2020). Preliminary steps in this direction are seen in inferential knowledge bases such as ATOMIC (Sap et al., 2019), which provides limited execution logic using edges typed with if-then relations.

Alongside FrameNet and MetaNet, others have collected schema and metaphor mappings, by learning them from large corpora (Beigman Klebanov et al., 2016; Gao et al., 2018). Pastra et al. (2011) built a database of concepts directly groundable to sensorimotor representations, primarily for robotics applications.

8 Conclusions

This position paper has proposed an approach to representation and learning based on the tenets of ECL. The proposed architecture, drawing on contemporary cognitive science, aims to address key limitations of current NLU systems through mental simulation and grounded metaphoric inference. We outlined major challenges and suggested a roadmap towards realizing the proposed vision.

Growing empirical evidence shows that language is intricately intertwined with a vast range of other neural processes. Accordingly, this work suggests a symbiotic view of cognitive science, embodied AI, and computational linguistics. By sharing common foundational problems, these fields may better share and co-evolve common solutions. Finally, we believe that attaining deeper language understanding must be a large scale effort, beyond the scope of any one research group. We hope that the paradigm presented here will help provide coherence to such efforts. One of our main goals was to stimulate a discussion; moving forward, we welcome comments, feedback, and suggestions.

References

Prithviraj Ammanabrolu and Mark Riedl. 2019. Playing text-adventure games with graph-based deep reinforcement learning. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 3557–3565, Minneapolis, Minnesota. Association for Computational Linguistics.

Ankesh Anand, Evan Racah, Sherjil Ozair, Yoshua Bengio, Marc-Alexandre Côté, and R. Devon Hjelm. 2019. Unsupervised state representation learning in atari. *Computing Research Repository*, arXiv:1906.08226.

Jacob Andreas, Anca Dragan, and Dan Klein. 2017. Translating neuralese. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 232–242, Vancouver, Canada. Association for Computational Linguistics.

Michael A. Arbib, Brad Gasser, and Victor Barrès. 2014. Language is handy but is it embodied? *Neuropsychologia*, 55(1):57–70.

Jason Baldridge, Tania Bedrax-Weiss, Daphne Luong, Srini Narayanan, Bo Pang, Fernando Pereira, Radu Soricut, Michael Tseng, and Yuan Zhang. 2018. Points, paths, and playscapes: Large-scale spatial language understanding tasks set in the real world. In *Proceedings of the First International Workshop on Spatial Language Understanding*, pages 46–52, New Orleans, Louisiana, USA.

Lawrence W. Barsalou. 2008. Grounded Cognition. *Annual Review of Psychology*, 59(1):617–645.

Beata Beigman Klebanov, Chee Wee Leong, E. Dario Gutierrez, Ekaterina Shutova, and Michael Flor. 2016. Semantic classifications for detection of verb metaphors. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 101–106, Berlin, Germany. Association for Computational Linguistics.

Yoshua Bengio. 2017. The Consciousness Prior. *Computing Research Repository*, arXiv:1709.08568.

Benjamin K Bergen and Nancy Chang. 2005. Embodied construction grammar in simulation-based language understanding. *Construction grammars: Cognitive grounding and theoretical extensions*, 3:147.

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. *Computing Research Repository*, arXiv:2004.10151.

Antoine Bordes, Jason Weston, Sumit Chopra, Tomas Mikolov, Arman Joulin, S Rush, and L Bottou. 2015. Artificial tasks for artificial intelligence. *Facebook AI Research. ICLR–San Diego–May*, 7:2015.

Anna M. Borghi, Laura Barca, Ferdinand Binkofski, Cristiano Castelfranchi, Giovanni Pezzulo, and Luca Tummolini. 2018. Words as social tools: Language, sociality and inner grounding in abstract concepts. *Physics of Life Reviews*, 29:120–153.

Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, and Yejin Choi. 2019. COMET: Commonsense transformers for automatic knowledge graph construction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4762–4779, Florence, Italy. Association for Computational Linguistics.

Nancy Chang, Srini Narayanan, and Miriam R.L. Petruck. 2002. Putting frames in perspective. In *COLING 2002: The 19th International Conference on Computational Linguistics*.

Maxime Chevalier-Boisvert, Dzmitry Bahdanau, Salem Lahlou, Lucas Willems, Chitwan Saharia, Thien Huu Nguyen, and Yoshua Bengio. 2019. BabyAI: First steps towards grounded language learning with a human in the loop. In *International Conference on Learning Representations*.

Herbert H Clark and Edward F Schaefer. 1989. Contributing to discourse. *Cognitive science*, 13(2):259–294.

Marc-Alexandre Côté, Ákos Kádár, Xingdi Yuan, Ben Kybartas, Tavian Barnes, Emery Fine, James Moore, Matthew Hausknecht, Layla El Asri, Mahmoud Adada, and et al. 2019. Textworld: A learning environment for text-based games. *Computer Games*, page 41–75.

Oana David and Ellen Dodge. 2014. Building the metanet metaphor repository: The natural symbiosis of metaphor analysis and construction grammar. In *The 8th International Construction Grammar Conference (ICCG 8), 3-6 September, Osnarbrück, Germany*.

Oana David, George Lakoff, and Elise Stickles. 2016. Cascades in metaphor and grammar. *Constructions and Frames*, 8(2):214–255.

Samuel B. Day and Dedre Gentner. 2007. Nonintentional analogical inference in text comprehension. *Memory & Cognition*, 35(1):39–49.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: pre-training of deep bidirectional transformers for language understanding. *Computing Research Repository*, arXiv:1810.04805.

Dheeru Dua, Yizhong Wang, Pradeep Dasigi, Gabriel Stanovsky, Sameer Singh, and Matt Gardner. 2019.

DROP: A reading comprehension benchmark requiring discrete reasoning over paragraphs. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 2368–2378, Minneapolis, Minnesota. Association for Computational Linguistics.

Jeffrey L Elman. 2004. An alternative view of the mental lexicon. *Trends in cognitive sciences*, 8(7):301–306.

Jerome Feldman and Srinivas Narayanan. 2004. Embodied meaning in a neural theory of language. *Brain and Language*, 89(2):385–392.

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

Maxwell Forbes, Ari Holtzman, and Yejin Choi. 2019. Do neural language representations learn physical commonsense? *Proceedings of the 41st Annual Conference of the Cognitive Science Society*.

Vincent François-Lavet, Yoshua Bengio, Doina Precup, and Joelle Pineau. 2019. Combined reinforcement learning via abstract representations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 3582–3589.

David Gaddy and Dan Klein. 2019. Pre-learning environment representations for data-efficient neural instruction following. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1946–1956, Florence, Italy. Association for Computational Linguistics.

Vittorio Gallese and George Lakoff. 2005. The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology*, 22(3-4):455–479.

Ge Gao, Eunsol Choi, Yejin Choi, and Luke Zettlemoyer. 2018. Neural metaphor detection in context. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 607–613.

P. Gärdenfors. 2014. *The Geometry of Meaning: Semantics Based on Conceptual Spaces*. The MIT Press. MIT Press.

Matt Gardner. 2019. *How will we know when machines can read?*

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. *Computing Research Repository*, arXiv:2004.02709.

Matt Gardner, Jonathan Berant, Hannaneh Hajishirzi, Alon Talmor, and Sewon Min. 2019a. On Making Reading Comprehension More Comprehensive.

Matt Gardner, Jonathan Berant, Hannaneh Hajishirzi, Alon Talmor, and Sewon Min. 2019b. Question answering is a format; when is it useful? *Computing Research Repository*, arXiv:1909.11291.

Jon Gauthier and Igor Mordatch. 2016. A paradigm for situated and goal-driven language learning. *Computing Research Repository*, arXiv:1610.03585.

Dedre Gentner and Kenneth D Forbus. 2011. Computational models of analogy. *Wiley interdisciplinary reviews: cognitive science*, 2(3):266–276.

Mor Geva and Jonathan Berant. 2018. Learning to search in long documents using document structure. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 161–176, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Ofer Givoli and Roi Reichart. 2019. Zero-Shot Semantic Parsing for Instructions. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4454–4464, Stroudsburg, PA, USA. Association for Computational Linguistics.

Arthur M. Glenberg. 2008. Toward the Integration of Bodily States, Language, and Action. In Gun R. Semin and Eliot R. Smith, editors, *Embodied Grounding*, pages 43–70. Cambridge University Press, Cambridge.

Max Glockner, Vered Shwartz, and Yoav Goldberg. 2018. Breaking NLI systems with sentences that require simple lexical inferences. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 650–655, Melbourne, Australia. Association for Computational Linguistics.

H Paul Grice et al. 1975. Logic and conversation.

Rick Grush. 2004. The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences*, 27(3):377–396.

David Ha and Jürgen Schmidhuber. 2018. Recurrent world models facilitate policy evolution. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems 31*, pages 2450–2462. Curran Associates, Inc.

Jessica B Hamrick. 2019. Analogues of mental simulation and imagination in deep learning. *Current Opinion in Behavioral Sciences*.

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

Leslie Pack Kaelbling, Michael L Littman, and Anthony R Cassandra. 1998. Planning and acting in partially observable stochastic domains. *Artificial intelligence*, 101(1-2):99–134.

Neel Kant. 2018. Recent advances in neural program synthesis. *Computing Research Repository*, arXiv:1802.02353.

Daniel Keysers, Nathanael Schärli, Nathan Scales, Hylke Buisman, Daniel Furrer, Sergii Kashubin, Nikola Momchev, Danila Sinopalnikov, Lukasz Stafiniak, Tibor Tihon, Dmitry Tsarkov, Xiao Wang, Marc van Zee, and Olivier Bousquet. 2020. Measuring Compositional Generalization: A Comprehensive Method on Realistic Data. In *International Conference on Learning Representations (ICLR)*, pages 1–38.

Douwe Kiela, Luana Bulat, Anita L Vero, and Stephen Clark. 2016. Virtual embodiment: A scalable long-term strategy for artificial intelligence research. *Computing Research Repository*, arXiv:1610.07432.

Walter Kintsch and Teun A Van Dijk. 1978. Toward a model of text comprehension and production. *Psychological review*, 85(5):363.

Thomas Kipf, Elise van der Pol, and Max Welling. 2020. Contrastive learning of structured world models. In *International Conference on Learning Representations*.

Brenden Lake, Tal Linzen, and Marco Baroni. 2019. Human few-shot learning of compositional instructions. In Ashok Goel, Colleen Seifert, and Christian Freksa, editors, *Proceedings of the 41st Annual Conference of the Cognitive Science Society*, pages 611–616. Cognitive Science Society, Montreal, Canada.

Brenden M. Lake, Tomer D. Ullman, Joshua B. Tenenbaum, and Samuel J. Gershman. 2017a. Building machines that learn and think like people. *Computing Research Repository*, arXiv:arXiv:1604.00289v3.

Brenden M. Lake, Tomer D. Ullman, Joshua B. Tenenbaum, and Samuel J. Gershman. 2017b. Ingredients of intelligence: From classic debates to an engineering roadmap. *Computing Research Repository*, arXiv:1604.00289.

George Lakoff and Mark Johnson. 1980. The metaphorical structure of the human conceptual system. *Cognitive science*, 4(2):195–208.

George Lakoff and Srini Narayanan. 2010. Toward a computational model of narrative. In *2010 AAAI Fall Symposium Series*.

Ron Langacker. 1987. 1991 foundations of cognitive grammar, 2 volumes.

Neil D. Lawrence. 2017. Living together: Mind and machine intelligence. *Computing Research Repository*, arXiv:1705.07996.

David Lewis. 1976. General semantics. In *Montague grammar*, pages 1–50. Elsevier.

Chen Liang, Jonathan Berant, Quoc Le, Kenneth D. Forbus, and Ni Lao. 2016. Neural Symbolic Machines: Learning Semantic Parsers on Freebase with Weak Supervision (Short Version). *Computing Research Repository*, arXiv:1612.01197.

Percy Liang. 2016. Learning executable semantic parsers for natural language understanding. *Communications of the ACM*, 59(9):68–76.

Weijie Liu, Peng Zhou, Zhe Zhao, Zhiruo Wang, Qi Ju, Haotang Deng, and Ping Wang. 2019a. K-bert: Enabling language representation with knowledge graph. *Computing Research Repository*, arXiv:1909.07606.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019b. Roberta: A robustly optimized bert pretraining approach. *Computing Research Repository*, arXiv:1907.11692.

Reginald Long, Panupong Pasupat, and Percy Liang. 2016. Simpler context-dependent logical forms via model projections. *Computing Research Repository*, arXiv:1606.05378.

Gary Lupyan and Benjamin Bergen. 2016. How Language Programs the Mind. *Topics in Cognitive Science*, 8(2):408–424.

Gary Lupyan and Molly Lewis. 2019. From words-as-mappings to words-as-cues: the role of language in semantic knowledge. *Language, Cognition and Neuroscience*, 34(10):1319–1337.

James L. McClelland, Felix Hill, Maja Rudolph, Jason Baldridge, and Hinrich Schütze. 2019. Extending Machine Language Models toward Human-Level Language Understanding. *Computing Research Repository*, arXiv:1912.05877.

R. Thomas McCoy, Junghyun Min, and Tal Linzen. 2019a. Berts of a feather do not generalize together: Large variability in generalization across models with similar test set performance. *Computing Research Repository*, arXiv:1911.02969.

Tom McCoy, Ellie Pavlick, and Tal Linzen. 2019b. Right for the wrong reasons: Diagnosing syntactic heuristics in natural language inference. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3428–3448, Florence, Italy. Association for Computational Linguistics.

Melanie Mitchell. 1993. *Analogy-Making as Perception: A Computer Model*. MIT Press, Cambridge, MA, USA.

Pramod Kaushik Mudrakarta, Ankur Taly, Mukund Sundararajan, and Kedar Dhamdhere. 2018. Did the model understand the question? In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1896–1906, Melbourne, Australia. Association for Computational Linguistics.

Srinivas Narayanan. 1997. *Knowledge-based Action Representations for Metaphor and Aspect (KARMA)*. PhD dissertation, The University of California.

Graham Nelson. 2005. Natural Language, Semantic Analysis and Interactive Fiction. *IF Theory Reader*, (April 2005):141–188.

Yixin Nie, Adina Williams, Emily Dinan, Mohit Bansal, Jason Weston, and Douwe Kiela. 2019. Adversarial NLI: A New Benchmark for Natural Language Understanding. *Computing Research Repository*, arXiv:1910.14599.

German I Parisi, Ronald Kemker, Jose L Part, Christopher Kanan, and Stefan Wermter. 2019. Continual lifelong learning with neural networks: A review. *Neural Networks*.

Katerina Pastra, Eirini Balta, Panagiotis Dimitrakis, and Giorgos Karakatsiotis. 2011. Embodied language processing: a new generation of language technology. In *Workshops at the Twenty-Fifth AAAI Conference on Artificial Intelligence*.

Dennis R Proffitt. 2006. Embodied perception and the economy of action. *Perspectives on psychological science*, 1(2):110–122.

James Pustejovsky and Nikhil Krishnaswamy. 2016. Voxml: A visualization modeling language. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, pages 4606–4613.

Behrang QasemiZadeh, Miriam R. L. Petruck, Regina Stodden, Laura Kallmeyer, and Marie Candito. 2019. SemEval-2019 task 2: Unsupervised lexical frame induction. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 16–30, Minneapolis, Minnesota, USA. Association for Computational Linguistics.

David E Rumelhart. 1981. Understanding understanding. *Memories, thoughts and emotions: Essays in honor of George Mandler*, 257:275.

D. E. Rumlehart, P. Smolensky, J. L. McClelland, and G. E. Hinton. 1986. *Schemata and Sequential Thought Processes in PDP Models*, page 7–57. MIT Press, Cambridge, MA, USA.

Josef Ruppenhofer, Michael Ellsworth, Miriam R. L. Petruck, Christopher R. Johnson, Collin F. Baker, and Jan Scheffczyk. 2016. *FrameNet II: Extended Theory and Practice*. ICSI: Berkeley.

Maarten Sap, Ronan Le Bras, Emily Allaway, Chandra Bhagavatula, Nicholas Lourie, Hannah Rashkin, Brendan Roof, Noah A Smith, and Yejin Choi. 2019. Atomic: An atlas of machine commonsense for if-then reasoning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 3027–3035.

Jürgen Schmidhuber. 1990. Making the world differentiable: On using self-supervised fully recurrent neural networks for dynamic reinforcement learning and planning in non-stationary environments.

Elad Segal, Avia Efrat, Mor Shoham, Amir Globerson, and Jonathan Berant. 2019. A simple and effective model for answering multi-span questions. *Computing Research Repository*, arXiv:1909.13375.

Robert Speer, Joshua Chin, and Catherine Havasi. 2017. Conceptnet 5.5: An open multilingual graph of general knowledge. In *Thirty-First AAAI Conference on Artificial Intelligence*.

Luc Steels and Joachim de Beule. 2006. A (very) brief introduction to fluid construction grammar. In *Proceedings of the Third Workshop on Scalable Natural Language Understanding*, pages 73–80, New York City, New York. Association for Computational Linguistics.

Arjen Stolk, Lennart Verhagen, and Ivan Toni. 2016. Conceptual Alignment: How Brains Achieve Mutual Understanding. *Trends in Cognitive Sciences*, 20(3):180–191.

Leonard Talmy. 1983. How language structures space. In *Spatial orientation*, pages 225–282. Springer.

Leonard Talmy. 1985. Lexicalization patterns: Semantic structure in lexical forms. *Language typology and syntactic description*, 3(99):36–149.

Ronen Tamari, Hiroyuki Shindo, Dafna Shahaf, and Yuji Matsumoto. 2019. Playing by the book: An interactive game approach for action graph extraction from text. In *Proceedings of the Workshop on Extracting Structured Knowledge from Scientific Publications*, pages 62–71, Minneapolis, Minnesota. Association for Computational Linguistics.

Ronen Tamari, Gabriel Stanovsky, Dafna Shahaf, and Reut Tsarfaty. 2020. Ecological semantics: Programming environments for situated language understanding. *Computing Research Repository*, arXiv:2003.04567.

Michael Tomasello. 2008. *Human Cooperative Communication*.

Francisco J Varela, Evan Thompson, and Eleanor Rosch. 2017. *The embodied mind: Cognitive science and human experience*. MIT press.

John Winn, John Guiver, Sam Webster, Yordan Zaykov, Martin Kukla, and Dany Fabian. 2019. Alexandria:

Unsupervised high-precision knowledge base construction using a probabilistic program. In *Automated Knowledge Base Construction (AKBC)*. Best Research Paper Award.

Yiqun Yao, Jiaming Xu, Jing Shi, and Bo Xu. 2018. Learning to activate logic rules for textual reasoning. *Neural Networks*.

Dani Yogatama, Cyprien de Masson d'Autume, Jerome Connor, Tomas Kocisky, Mike Chrzanowski, Lingpeng Kong, Angeliki Lazaridou, Wang Ling, Lei Yu, Chris Dyer, et al. 2019. Learning and evaluating general linguistic intelligence. *Computing Research Repository*, arXiv:1901.11373.

Xingdi Yuan, Jie Fu, Marc-Alexandre Cote, Yi Tay, Christopher Pal, and Adam Trischler. 2019. Interactive machine comprehension with information seeking agents. *Computing Research Repository*, arXiv:1908.10449.