

ECOLOGICAL SEMANTICS: PROGRAMMING ENVIRONMENTS FOR SITUATED LANGUAGE UNDERSTANDING

Ronen Tamari & Dafna Shahaf
The Hebrew University of Jerusalem
{ronent, dshahaf}@cs.huji.ac.il

Gabriel Stanovsky & Reut Tsarfaty
Allen Institute for Artificial Intelligence
{gabis, reutt}@allenai.org

ABSTRACT

Large-scale natural language understanding (NLU) systems have made impressive progress: they can be applied flexibly across a variety of tasks, and employ minimal structural assumptions. However, extensive empirical research has shown this to be a double-edged sword, coming at the cost of shallow understanding: inferior generalization, grounding and explainability. Grounded language learning approaches offer the promise of deeper understanding by situating learning in richer, more structured training environments, but are limited in scale to relatively narrow, predefined domains. How might we enjoy the best of both worlds: grounded, general NLU? Following extensive contemporary cognitive science, we propose treating environments as “first-class citizens” in semantic representations, worthy of research and development in their own right. Importantly, models should also be partners in the creation and configuration of environments, rather than just actors within them, as in existing approaches. To do so, we argue that models must begin to understand and program in the language of affordances (which define possible actions in a given situation) both for online, situated discourse comprehension, as well as large-scale, offline common-sense knowledge mining. To this end we propose an environment-oriented ecological semantics, outlining theoretical and practical approaches towards implementation. We further provide actual demonstrations building upon interactive fiction programming languages.

1 INTRODUCTION

“Ask not what’s inside your head, but what your head’s inside of.” (Mace, 1977)

Recovery of meaning is at the heart of the endeavor to build better natural language understanding (NLU) systems. Semantics researchers study meaning representation, and in particular the relations between language and cognitive representations (Gärdenfors, 2014).

A recurring point of contention in semantics research (Fodor & Pylyshyn, 1988; Mahon & Caramazza, 2008) concerns the degree to which knowledge representation and language comprehension involve a *symbolic* internal language of thought (LoT) (Fodor, 1975) or are *embodied*; i.e., grounded in the brain’s systems for action and perception (Feldman & Narayanan, 2004; Barsalou, 2007).

Current deep-learning methods for large-scale NLU, such as BERT (Devlin et al., 2018), incorporate minimal cognitive biases and assume primarily distributional semantics (Firth, 1957). Extensive empirical research shows this to be a double-edged sword: while affording widespread applicability to a variety of tasks, such methods are limited by impoverished training environments (static datasets, narrow contextual prediction, etc.) and struggle in settings requiring deeper understanding, such as systematic generalization (Lake et al., 2019; McCoy et al., 2019), common-sense (Forbes et al., 2019) and explainability (Gardner et al., 2019).

Contemporary cognitive science can be seen as adopting a more holistic approach; integrating symbolic, embodied and distributional accounts (Lupyan & Lewis, 2019), but also focusing on the crucial *ecological* component (Gibson, 1979; Hasson et al., 2020): cognition emerges from brain-body-environment interaction. Systematic regularities in the interactions play a key role in inducing systematic linguistic (Narayanan, 1997) and knowledge (Davis et al., 2020) representations. These interactional regularities differ in fundamental ways from statistical regularities available to current

arXiv:2003.04567v2 [cs.AI] 24 May 2020

general NLU methods (Hasson et al., 2020), for example including perceptual, spatiotemporal and causal dynamics (Rodd, 2020; Davis et al., 2020).

Situated (grounded) approaches (Mikolov et al., 2015; Liang, 2016) focus on mapping language to executable forms, and highlight the importance of external environments (McClelland et al., 2019); Hill et al. (2020) show the emergence of systemic generalization to be contingent on careful task/environment design, rather than specific architectural engineering alone. However, while such environments clearly play an important role in building NLU systems, they are (1) relatively narrow and fixed in terms of semantics (2) costly to create, especially multi-modal environments.

Here we propose an approach to address this limitation and extend grounded language approaches towards more general domains, by harnessing the power of language to also create and shape environments, rather than just to induce literal execution within them. In this important, yet relatively unexplored role, language helps structure semantic knowledge and serves as a proxy for expensive embodied experience (Lupyan & Bergen, 2016). To efficiently accomplish this remarkable feat, humans use the language of *affordances* (Gibson, 1979; Glenberg, 2008) to construct “mental worlds”; shaping interactions by specifying what can be done in various situations, from concrete to abstract. We propose that NLU systems should learn to understand (parse) and use such language (e.g., “This bag can hold up to 20kg before bursting”, see §2), which we suggest has a natural programmatic equivalent in the behavioral programming paradigm, such as interactive fiction languages.

In summary, we make the following more concrete contributions and proposals:

- Ecological Semantics: Outline for a theoretical and practical approach to a semantic parsing framework for creation as well as interaction with environments through language. Design considerations are informed by contemporary cognitive science, AI/NLU research and programming language theory (PLT).
- We propose methods to inject rich, *actionable* external knowledge into the framework at scale, building upon data mining and automated knowledge base construction (AKBC) research.
- We make available¹ simple interactive demonstrations as working examples showing how such methods can be applied towards open challenges such as common-sense and causal reasoning.

2 MOTIVATING CHALLENGES: INCORPORATING WORLD KNOWLEDGE

Explicitly Provided Knowledge. Consider the example in figure 1, describing an everyday situation of shopping for fruit in a market. Completely trivial for humans, current NLU methods find such “what-if” questions highly challenging, even though the relevant affordances are made explicit in the text. A textual entailment model judges it very likely that “The bag bursts.” for $X \in \{\text{no,one,two,three}\}$ ².

Assumed World Knowledge. In this common, yet more difficult setting, the relevant knowledge is implicitly assumed. Consider a prompt like “He put on a white t-shirt and blue jeans. Next, he wore .”. A completion produced by GPT-2 (Radford et al., 2019) is “a gray cowboy hat, black cargo pants, and white shoes. He also had a black baseball cap pulled low over his eyes”³.

Common-sense knowledge graphs are likely to be insufficient for such problems; as shown in Forbes et al. (2019), “neural language representations still only learn associations that are explicitly written down”, even after being explicitly trained on a knowledge graph of objects and affordances. As suggested by the work, mental simulations are crucial to common-sense in humans (Battaglia et al., 2013), allowing the dynamic, affordance-guided construction of relevant representations at run-time as needed, rather than wasting valuable space in memorizing large, ever-incomplete relation graphs.

Importantly, the first problem should be simpler than the second: the required background knowledge is made available in the text. It would be highly desirable to be able to act upon such information. Recent work has begun to explore such capabilities (Zhong et al., 2020), but current methods are largely limited in this respect (Luketina et al., 2019). In the following section, we propose a general problem formulation for incorporating affordances, building upon cognitive linguistics theory.

¹<https://eco-sem.github.io/>

²<https://demo.allennlp.org/textual-entailment/>

³<https://talktotransformer.com/>

(1) You're shopping for fruit in the market. Your plastic bag can hold 20 kilograms before bursting. One watermelon weighs 10 kilograms. What would happen if you put 3 watermelons in the bag?				
(2)	A weight is a kind of value. 1kg specifies a weight. Everything has a weight. A thing usually has weight 1kg. The verb to weigh means the weight property. A container has a weight called breaking strain. The breaking strain of a container is usually 5kg. Definition: A container is bursting if the total weight of things in it is greater than its breaking strain. Every turn when a container (called the sack) held by someone visible (called the chump) is bursting:say "[The sack] bursts under the weight! [if the player is the chump]You discard[otherwise][The chump] discards[end if] its	remains, looking miserably down at [the list of things in the sack] on the floor.";now all of the things in the sack are in the location;now the sack is nowhere.	(4)	>[1] take watermelon Taken. // (repeat two more times) >[4] put watermelon in bag You put the watermelon into the plastic bag. // (repeat two more times) >[7] take bag Taken.The plastic bag bursts under the weight! You discard its remains, looking miserably down at the three watermelons on the floor.
		(3)	A watermelon is a kind of portable thing. A watermelon has weight 10kg. The plastic bag is a container with breaking strain 20kg. Three watermelons and a plastic bag are in the Fruit Stall.	
Legend	Compiled Knowledge	Ecological actions	Indexicalization	Executable Actions Observations

Figure 1: Inform7 ecological semantic parsing example for §2 challenge. (1) Input prompt (2) Pre-existing, compiled knowledge (3) Situation knowledge: simulation configuration and indexicalization of referent objects (4) Run simulation to answer “what-if” question.

3 ECOLOGICAL SEMANTICS

Mental simulations and affordances feature centrally in contemporary cognitive linguistics research. According to one such theory, the Indexical Hypothesis (Glenberg, 2008), language comprehension involves three key processes: (1) indexing objects, (2) deriving their affordances, and (3) meshing them together into a coherent (action-based) simulation as directed by grammatical cues. Importantly, affordances generally cannot be derived directly from words, but rather rely on context and pre-existing object representations.

3.1 COMPUTATIONAL FORMULATION

The Indexical Hypothesis (IH) can be formulated naturally within the model-based framework used in general AI mental simulation research (Hamrick, 2019). At the core of such frameworks is the partially observable Markov Decision Process (POMDP) (Kaelbling et al., 1998), which governs the relations between states (s), actions (a), observations (o) and rewards (r). Specifically, we focus on the recognition⁴ $I : O \rightarrow S$, transition $T : S \times A \rightarrow S$ and policy $\pi : S \rightarrow A$ functions.

Pre-existing knowledge regarding the environment (objects and their affordances) can be seen to be primarily represented by T , with the emulator model being the neural correlate (Grush, 2004; Glenberg, 2008). In the POMDP formulation, for a linguistic input (or observation) x , IH can be formulated as (1) compose an initial state representation $I(x) = s_0$ of objects (we assume the simple case where all objects are mentioned in x) (2) derive affordances, or the set of actions that can be taken in the current situation (3) enact mental simulation by applying T with chosen action. Typically x is composed of multiple utterances ($\bar{x}_1, \dots, \bar{x}_N$) and so the simulation may be composed of multiple actions $\mathbf{a} = (a_0, \dots, a_{L-1})$. Slightly abusing notation, we can denote the full execution $T(s_0, \mathbf{a})$ which yields a result state s_L . IH can be seen as corresponding to the standard setting in executable semantic parsing/grounded NLU works (Long et al., 2016):

Executable Semantic Parsing (Ex-SP). Given a linguistic input x and target intent (goal state) g^* , output action sequence \mathbf{a} such that $T(I(x), \mathbf{a}) = g^*$. Most grounded/executable approaches assume a fixed, programmatic, domain-specific T (navigation environments, SQL engine, etc.) and focus on learning a policy mapping from x to \mathbf{a} .

Our proposal thus focuses on “pushing the envelope” of T to allow grounded understanding of more general language. IH discusses the comprehension process in cases where the relevant object and affordance information already exists. But how do we learn such representations in the first place? Embodied experience is one way, but a costly and slow one, so here we focus on the role of language in shaping affordance knowledge, specifically modal language, like “All watermelons are portable.”

⁴Commonly denoted O^{-1} , we denote here by I for Indexicalization.

Such language can more naturally be seen as modifying⁵ the emulator T . Therefore, we propose extending the representation of T to allow it to change in time, $T^{(t)}$, modified by special *eco-actions* \hat{a} . These do not change the current state, but rather only the executor (example in fig. 1). We denote regular executed actions as \hat{a} , and a *scenario* (containing possibly both \hat{a} , \hat{a} actions) as $\tilde{\mathbf{a}}$. The full execution is then $T^{(0:L-1)}(s_0, \tilde{\mathbf{a}})$, which denotes applying $T^{(t)}$ at each timestep.

Ecological Semantic Parsing (Ec-SP). Given a linguistic input \mathbf{x} and target intent (goal state) g^* , output action sequence $\tilde{\mathbf{a}}$ such that $T^{(0:L-1)}(I(\mathbf{x}), \tilde{\mathbf{a}}) = g^*$.

Figure 1 shows how Ec-SP can be utilized towards addressing the challenge problem from §2, which is not handled by current Ex-SP methods, as the input language is out-of-domain (so a specific executor would need to be created). The implementation uses Inform7 (Nelson, 2005), an interactive fiction (IF) language (see §4). Interactive versions of the examples from §2 are available online.

We distinguish between compiled knowledge vs. situation knowledge: the former refers to existing knowledge encompassed by the emulator (analogous to code libraries that just need to be imported), the latter is new knowledge defined online using eco-acts (analogous to writing a new program). Clearly, a core issue to be managed is the scalable and incremental growth of the emulator: as in regular programming, recurring ecological information (such as watermelons being portable) should become part of the library, rather than having to be re-defined anew in every situation.

4 AFFORDABLE AFFORDANCES: TOWARDS IMPLEMENTATION

Programmatic emulation of environments requires an appropriate programming formalism with which environments can be flexibly constructed and configured⁶. Our focus here is on purely text based construction, from considerations of scale, to remain broadly applicable to general NLU; multi-modal integration is an interesting future direction. We suggest that a natural paradigm for such a purpose is Behavioral Programming (Harel et al., 2012), which can also be seen to include certain IF languages, like Inform7 (Nelson, 2005). These languages are designed to be reminiscent of natural language, and express semantics in terms of interactional affordances (indeed often using modal verbs like *can*, *mustn't*) (Harel et al., 2012). Current frameworks for creating custom IF training environments (Côté et al., 2018; Tamari et al., 2019) require extensive re-configuration for new domains, and games must be pre-compiled rather than generated dynamically from textual inputs. Most current IF works focus on solving existing games (Jain et al., 2019) or game construction for human entertainment (Ammanabrolu et al., 2020).

Learning emulators at large-scale. This task is closely related to the grand AI challenge of common-sense learning. In humans, common-sense is hard-coded through rich experience (Hasson et al., 2020); it is reasonable to expect that approximating human emulators will require extensive hard-coding as well. In rendering this task tractable, We join Kordjamshidi et al. (2018) in advocating a tighter loop between learning and programming to represent knowledge: AI should be extensively utilized in hard-coding its own common-sense. Whereas earlier approaches typically consisted of non-executable, relational knowledge graphs (KGs) (Speer et al., 2017), in our case knowledge can be represented by code, executable in interactive simulations. KGs will likely be useful for populating an initial “seed emulator”, as will AKBC methods for learning object (Elazar et al., 2019) and action (Forbes & Choi, 2017) properties at scale. In Pustejovsky & Krishnaswamy (2018), multimodal simulations are used to evaluate automatic affordance extraction. In Balint & Allbeck (2017), game designers (for human games) utilized NLU methods for learning object affordances. Finally, as symbolic knowledge is by nature incomplete, it will need to be superseded by geometric, multi-modal knowledge representations (Gärdenfors, 2014; Pezeshkpour et al., 2018).

By affording NLU systems with the ability to programmatically emulate environments in the context of both online discourse comprehension, as well as large-scale, offline common-sense knowledge mining, we hope to advance research efforts towards grounded, general NLU.

⁵This is a delicate point- we currently assume the modification is valid, and leave a more thorough discussion of the rules governing what is possible to future work.

⁶This preliminary approach is naturally biased towards literal language, which is easier to simulate than more abstract language. While a detailed analysis is out of scope, we note that literal language is seen to lay the neural foundations for abstract language understanding (Lakoff & Johnson, 1980; Davis et al., 2020)

ACKNOWLEDGMENTS

We thank the Hyadata Lab at HUJI, and Yoav Goldberg, Ido Dagan, and the audience of the BIU-NLP seminar for interesting discussion and thoughtful remarks. The last author is funded by an ERC grant on *Natural Language Programming*, grant number 677352, for which we are grateful. This work was also supported by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant no. 852686, SIAM) and NSF-BSF grant no. 2017741 (Shahaf).

REFERENCES

- Prithviraj Ammanabrolu, Wesley Cheung, Dan Tu, William Broniec, and Mark O. Riedl. Bringing stories alive: Generating interactive fiction worlds. *CoRR*, abs/2001.10161, 2020. URL <http://arxiv.org/abs/2001.10161>.
- J. Timothy Balint and Jan Allbeck. ALET: Agents Learning their Environment through Text. *Computer Animation and Virtual Worlds*, 28(3-4):1–9, 2017. ISSN 1546427X. doi: 10.1002/cav.1759.
- Lawrence W. Barsalou. Grounded Cognition. *Annual Review of Psychology*, 59(1):617–645, 2007. ISSN 0066-4308. doi: 10.1146/annurev.psych.59.103006.093639.
- Peter W. Battaglia, Jessica B. Hamrick, and Joshua B. Tenenbaum. Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences of the United States of America*, 110(45):18327–18332, 2013. ISSN 00278424. doi: 10.1073/pnas.1306572110.
- Marc-Alexandre Côté, Ákos Kádár, Xingdi Yuan, Ben Kybartas, Tavian Barnes, Emery Fine, James Moore, Ruo Yu Tao, Matthew Hausknecht, Layla El Asri, Mahmoud Adada, Wendy Tay, and Adam Trischler. Textworld: A learning environment for text-based games. *CoRR*, abs/1806.11532, 2018.
- Charles P. Davis, Gerry T. M. Altmann, and Eiling Yee. Situational systematicity: A role for schema in understanding the differences between abstract and concrete concepts. *Cognitive Neuropsychology*, pp. 1–12, jan 2020. ISSN 0264-3294. doi: 10.1080/02643294.2019.1710124. URL <https://www.tandfonline.com/doi/full/10.1080/02643294.2019.1710124>.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
- Yanai Elazar, Abhijit Mahabal, Deepak Ramachandran, Tania Bedrax-Weiss, and Dan Roth. How Large Are Lions? Inducing Distributions over Quantitative Attributes. 2019. URL <http://arxiv.org/abs/1906.01327>.
- Jerome Feldman and Srinivas Narayanan. Embodied meaning in a neural theory of language. *Brain and Language*, 89(2):385–392, 2004. ISSN 0093934X. doi: 10.1016/S0093-934X(03)00355-9.
- John R Firth. A synopsis of linguistic theory, 1930-1955. *Studies in linguistic analysis*, 1957.
- Jerry A Fodor. *The language of thought*, volume 5. Harvard university press, 1975.
- Jerry A Fodor and Zenon W Pylyshyn. Connectionism and cognitive architecture: A critical analysis. *Cognition*, 1988.
- Maxwell Forbes and Yejin Choi. VERB PHYSICS: Relative physical knowledge of actions and objects. *ACL 2017 - 55th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, 1:266–276, 2017. doi: 10.18653/v1/P17-1025.
- Maxwell Forbes, Ari Holtzman, and Yejin Choi. Do neural language representations learn physical commonsense? *Proceedings of the 41st Annual Conference of the Cognitive Science Society*, 2019.
- P. Gärdenfors. *The Geometry of Meaning: Semantics Based on Conceptual Spaces*. The MIT Press. MIT Press, 2014. ISBN 9780262026789. URL <https://books.google.co.il/books?id=QDOKAgAAQBAJ>.

- Matt Gardner, Jonathan Berant, Hannaneh Hajishirzi, Alon Talmor, and Sewon Min. On making reading comprehension more comprehensive. In *Proceedings of the 2nd Workshop on Machine Reading for Question Answering*, pp. 105–112, Hong Kong, China, November 2019. Association for Computational Linguistics. doi: 10.18653/v1/D19-5815. URL <https://www.aclweb.org/anthology/D19-5815>.
- James Jerome Gibson. The ecological approach to visual perception. 1979.
- Arthur M. Glenberg. Toward the Integration of Bodily States, Language, and Action. In Gun R. Semin and Eliot R. Smith (eds.), *Embodied Grounding*, pp. 43–70. Cambridge University Press, Cambridge, 2008. ISBN 9780511805837. doi: 10.1017/CBO9780511805837.003. URL https://www.cambridge.org/core/product/identifier/CBO9780511805837A010/type/book_{_}part.
- Rick Grush. The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences*, 27(3):377–396, 2004. ISSN 0140525X. doi: 10.1017/S0140525X04000093.
- Jessica B Hamrick. Analogues of mental simulation and imagination in deep learning. *Current Opinion in Behavioral Sciences*, 2019. ISSN 23521546. doi: 10.1016/j.cobeha.2018.12.011.
- David Harel, Assaf Marron, and Gera Weiss. Behavioral programming. *Communications of the ACM*, 55(7):90–100, 2012. ISSN 00010782. doi: 10.1145/2209249.2209270.
- Uri Hasson, Samuel A. Nastase, and Ariel Goldstein. Direct Fit to Nature: An Evolutionary Perspective on Biological and Artificial Neural Networks. *Neuron*, 105(3):416–434, 2020. ISSN 10974199. doi: 10.1016/j.neuron.2019.12.002. URL <https://doi.org/10.1016/j.neuron.2019.12.002>.
- Felix Hill, Andrew Lampinen, Rosalia Schneider, Stephen Clark, Matthew Botvinick, James L. McClelland, and Adam Santoro. Environmental drivers of systematicity and generalization in a situated agent. In *International Conference on Learning Representations*, 2020. URL <https://openreview.net/forum?id=Sk1GryBtwr>.
- Vishal Jain, William Fedus, Hugo Larochelle, Doina Precup, and Marc G. Bellemare. Algorithmic Improvements for Deep Reinforcement Learning applied to Interactive Fiction. 2019. URL <http://arxiv.org/abs/1911.12511>.
- Leslie Pack Kaelbling, Michael L Littman, and Anthony R Cassandra. Planning and acting in partially observable stochastic domains. *Artificial intelligence*, 101(1-2):99–134, 1998.
- Parisa Kordjamshidi, Dan Roth, and Kristian Kersting. Systems AI: A declarative learning based programming perspective. *IJCAI International Joint Conference on Artificial Intelligence*, 2018-July:5464–5471, 2018. ISSN 10450823.
- Brenden Lake, Tal Linzen, and Marco Baroni. Human few-shot learning of compositional instructions. In Ashok Goel, Colleen Seifert, and Christian Freksa (eds.), *Proceedings of the 41st Annual Conference of the Cognitive Science Society*, pp. 611–616. Cognitive Science Society, Montreal, Canada, 2019.
- George Lakoff and Mark Johnson. The metaphorical structure of the human conceptual system. *Cognitive science*, 4(2):195–208, 1980.
- Percy Liang. Learning executable semantic parsers for natural language understanding. *Communications of the ACM*, 59(9):68–76, 2016. ISSN 15577317. doi: 10.1145/2866568.
- Reginald Long, Panupong Pasupat, and Percy Liang. Simpler context-dependent logical forms via model projections. *54th Annual Meeting of the Association for Computational Linguistics, ACL 2016 - Long Papers*, 3:1456–1465, 2016. doi: 10.18653/v1/p16-1138.
- Jelena Luketina, Nantas Nardelli, Gregory Farquhar, Jakob Foerster, Jacob Andreas, Edward Grefenstette, Shimon Whiteson, and Tim Rocktäschel. A Survey of Reinforcement Learning Informed by Natural Language. jun 2019. URL <http://arxiv.org/abs/1906.03926>.

- Gary Lupyan and Benjamin Bergen. How Language Programs the Mind. *Topics in Cognitive Science*, 8(2):408–424, 2016. ISSN 17568765. doi: 10.1111/tops.12155.
- Gary Lupyan and Molly Lewis. From words-as-mappings to words-as-cues: the role of language in semantic knowledge. *Language, Cognition and Neuroscience*, 34(10):1319–1337, nov 2019. ISSN 2327-3798. doi: 10.1080/23273798.2017.1404114. URL <https://www.tandfonline.com/doi/full/10.1080/23273798.2017.1404114>.
- William M Mace. James j. gibson’s strategy for perceiving: Ask not what’s inside your head, but what’s your head inside of. *Perceiving, acting, and knowing: Towards an ecological psychology*, 1977.
- Bradford Z. Mahon and Alfonso Caramazza. A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology Paris*, 102(1-3): 59–70, 2008. ISSN 09284257. doi: 10.1016/j.jphysparis.2008.03.004.
- James L. McClelland, Felix Hill, Maja Rudolph, Jason Baldridge, and Hinrich Schütze. Extending Machine Language Models toward Human-Level Language Understanding. pp. 1–8, 2019. URL <http://arxiv.org/abs/1912.05877>.
- R. Thomas McCoy, Jung-Hyun Min, and Tal Linzen. Berts of a feather do not generalize together: Large variability in generalization across models with similar test set performance. *ArXiv*, abs/1911.02969, 2019.
- Tomas Mikolov, Armand Joulin, and Marco Baroni. A roadmap towards machine intelligence. *ArXiv*, abs/1511.08130, 2015.
- Srinivas Narayanan. *Knowledge-based Action Representations for Metaphor and Aspect (KARMA)*. PhD dissertation, The University of California, 1997.
- Graham Nelson. Natural Language, Semantic Analysis and Interactive Fiction. *IF Theory Reader*, (April 2005):141–188, 2005. URL http://inform-fiction.org/manual/if_{_}theory.html.
- Pouya Pezeshkpour, Liyan Chen, and Sameer Singh. Embedding multimodal relational data for knowledge base completion. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 3208–3218, Brussels, Belgium, October-November 2018. Association for Computational Linguistics. doi: 10.18653/v1/D18-1359. URL <https://www.aclweb.org/anthology/D18-1359>.
- James Pustejovsky and Nikhil Krishnaswamy. Every Object Tells a Story. In *Proceedings of the Workshop Events and Stories in the News 2018*, pp. 1–6, Santa Fe, New Mexico, U.S.A, aug 2018. Association for Computational Linguistics. URL <https://www.aclweb.org/anthology/W18-4301>.
- Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. 2019.
- Jennifer M. Rodd. Settling Into Semantic Space: An Ambiguity-Focused Account of Word-Meaning Access. *Perspectives on Psychological Science*, pp. 174569161988586, jan 2020. ISSN 1745-6916. doi: 10.1177/1745691619885860. URL <http://journals.sagepub.com/doi/10.1177/1745691619885860>.
- Robert Speer, Joshua Chin, and Catherine Havasi. Conceptnet 5.5: An open multilingual graph of general knowledge. In *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.
- Ronen Tamari, Hiroyuki Shindo, Dafna Shahaf, and Yuji Matsumoto. 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*, pp. 62–71, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics. doi: 10.18653/v1/W19-2609. URL <https://www.aclweb.org/anthology/W19-2609>.
- Victor Zhong, Tim Rocktschel, and Edward Grefenstette. Rtfm: Generalising to new environment dynamics via reading. In *International Conference on Learning Representations*, 2020. URL <https://openreview.net/forum?id=SJgob6NKvH>.