Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 145 (2018) 724–729

Procedia Computer Science

www.elsevier.com/locate/procedia

Postproceedings of the 9th Annual International Conference on Biologically Inspired Cognitive Architectures, BICA 2018 (Ninth Annual Meeting of the BICA Society)

How Language Processing can Shape a Common Model of Cognition

Matthew A. Kelly*, David Reitter

College of Information Sciences and Technology, The Pennsylvania State University, University Park, 16802 - 6823, USA

Abstract

What role does the study of natural language play in the task of developing a unified theory and common model of cognition? Language is perhaps the most complex behaviour that humans exhibit, and, as such, is one of the most difficult problems for understanding human cognition. Linguistic theory can both inform and be informed by unified models of cognition. We discuss (1) how computational models of human cognition can provide insight into how humans produce and comprehend language and (2) how the problem of modelling language processing raises questions and creates challenges for widely used computational models of cognition. Evidence from the literature suggests that behavioural phenomena, such as recency and priming effects, and cognitive constraints, such as working memory limits, affect how language is produced by humans in ways that can be predicted by computational cognitive models. But just as computational models can provide new insights into language, language can serve as a test for these models. For example, simulating language learning requires the use of more powerful machine learning techniques, such as deep learning and vector symbolic architectures, and language comprehension requires a capacity for on-the-fly situational model construction. In sum, language plays an important role in both shaping the development of a common model of the mind, and, in turn, the theoretical understanding of language stands to benefit greatly from the development of a common model.

© 2018 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0/)
Peer-review under responsibility of the scientific committee of the 9th Annual International Conference on Biologically Inspired Cognitive Architectures.

Keywords: natural language processing; cognitive architectures; cognitive models; common model of cognition; human memory

1. Introduction

When attempting to construct a computational model of human cognition as a whole, what role does the study of natural language play in the building of that model? Language is perhaps the most complex behaviour that humans exhibit, and, as such, is one of the most difficult problems for understanding human cognition.

* Corresponding author.
E-mail address: matthew.kelly@psu.edu

1877-0509 © 2018 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0/)
Peer-review under responsibility of the scientific committee of the 9th Annual International Conference on Biologically Inspired Cognitive Architectures.
10.1016/j.procs.2018.11.047

Language thus serves as an important test for the project of producing a theory and model of all of human cognition. Language tests such a model in two ways: (1) by providing strong constraints that a model of cognition must satisfy, and (2) by providing a complex behavioural domain for evaluating predictions of the model. In short, psycholinguistic theory can both (1) *inform* and (2) *be informed by* models of cognition.

Cognitive architectures are computational implementations of theories of human cognition as a whole. By that we mean that cognitive architectures computationally specify the constraints and performance consequences placed by a theory on cognitive processes. Language and cognitive architectures impose strong constraints on each other. Language rules out cognitive architectures that are insufficiently powerful to accommodate the problem of language processing, and the domain-general processes of cognitive architectures, in turn, make strong claims about how language can be acquired and produced. To illustrate this relationship, in what follows, we provides examples of both how language theory informs and can be informed by theories and models of cognition.

2. Cognition informs language

If we understand language as a behaviour produced, at least in part, by cognitive processes that are not specific to language, theoretical insights from other domains of psychological research can inform our understanding of language.

2.1. Recency and priming

One example is the recency effect, the finding that more recent information is more easily remembered [10]. Early computational models of the recency effect postulated that the effect was caused by older information being displaced from a short-term memory system with limited information capacity [2]. However, later experimental work ruled out this account [3, 26], suggesting instead that the recency effect is a memory phenomenon that occurs irrespective of the span of time over which learning occurs. For example, when people try to remember where they parked their car every morning for the past two weeks, it is easier to recall more recent parking locations than those further in the past [8]. Given that the recency effect appears over both short and long-term learning, one would expect the recency effect to be observable in the learning of language.

Language data does, in fact, show evidence of both short and long-term recency effects. The ACT-R cognitive architecture [1] can model these effects. ACT-R posits that the recency effect is a product of the declarative memory system. Items in declarative memory have an activation that is a function of both how frequently and recently the item has been encountered. Items with higher activation are easier to retrieve from memory.

Using ACT-R as a model, Cole and Reitter [5] find that, in conversation, people are faster at speaking when using words with a higher activation in declarative memory (i.e., words that have been more frequently and recently used). Similarly, Cole, Ghafurian, and Reitter [6] conduct an analysis of social media data and find that new words (e.g., new brand names or slang) are adopted in online discussion and spread through online communities at a rate that can be understood in terms of the ACT-R activation of the word (i.e., as a function of the frequency and recency of the word).

Priming effects in grammar decisions can also be modeled by the retrieval dynamics of ACT-R's declarative memory [25]. The declarative memory can be used to store syntactic rules. These rules are more easily retrieved if used recently or frequently or when cued by associated semantic material retained in working memory. ACT-R explains a pattern of short and long-term syntactic priming effects that had previously been difficult to reconcile [25].

In this manner, cognitive architectures can inform our understanding of language processes, such as how new words are adopted into the language, or the dynamics of word choice and grammatical choice in conversation.

2.2. Forgetting and semantics

Insights from computational models of cognition have also informed our understanding of the semantics of language. While linguistics has long postulated that the meaning of words can be derived from their use in language [11], computational cognitive models provide a domain-general mechanism for this process. According to Hintzman [14], human memory operates by laying down memory traces of specific events. However, retrieval from memory is always a lossy reconstruction of these events, created by combining information from a lifetime of interrelated events. As a result, memory inherently generalizes, abstracts, and infers prototypes from across instances of an experience.

Applying such a memory system to language results in a system that naturally infers the meaning of a word from across multiple instances of the word's use.

Hintzman [14] models the process by which memories of instances become memories of prototypes using the MINERVA model, which, in turn, served as inspiration for the BEAGLE model of distributional semantics [17]. Building upon BEAGLE and evidence from research in human memory, Kelly, Reitter, and West [20] propose that human memory has the capacity to recursively generalize, allowing it to infer arbitrarily abstract relationships. Kelly, Reitter, and West find evidence that this capacity to infer abstract relationships aids in the ability to learn syntactic or part-of-speech relationships between words.

2.3. Working memory and cognitive constraints

Cognitive architectures can also impose constraints, such as working memory limits, which can help us understand how properties of language emerge from our cognitive limitations and why language is produced and parsed differently by different individuals [7].

3. Language informs cognition

Language can also play an important role in informing the development of cognitive architectures. Language processing is a hard problem, or rather, a collection of hard problems. Developing a system that can learn language from experience as a child does, analyze language into its component parts, comprehend language when spoken or written, and produce natural-sounding sentences, is, to say the least, a significant technical challenge.

3.1. Language acquisition

Any computational architecture that implements a complete theory of human cognition needs to have a memory system sufficiently powerful that it can account for how language is acquired and utilized by the mind. For example, the ACT-R architecture's declarative memory stores everything it encounters without loss of information. This presents two problems for language modelling, one practical and one theoretical:

(1) Language learning happens over the course of decades of experience as the human brain acquires a vast wealth of data in both spoken and written form. This poses a practical problem for ACT-R. ACT-R estimates the human response time for recall from declarative memory as a non-linear function of the activation of the stored data. However, the actual time taken by the ACT-R software to compute a recall scales linearly with the amount of data stored. Storing a lifetime of data in the ACT-R declarative memory is intractable given the current implementation of the ACT-R software.

(2) ACT-R lacks the capacity to infer and generalize from instances of language experience in the way that humans do when producing or comprehending novel utterances.

Thus modelling language acquisition forces us to adopt a cognitive architecture that uses a lossy memory model capable of generalization. Possible candidates include vector symbolic memory, which can replicate the behaviour of the ACT-R declarative memory but is scalable to language learning [19], and neural network models.

3.2. Language production

Some models of language production use declarative memory [25], typically instantiated as a collection of prototypes or exemplars, while other models use implicit learning, typically instantiated in a connectionist architecture [4].

According to ACT-R theory [1], newly acquired knowledge is added to declarative memory. If that knowledge is accessed sufficiently frequently, it is later added to procedural memory as a condition-action rule through the process of *routinization*.

Thus, ACT-R predicts that language representations start out stored in declarative memory as detailed, flexible, and compositional constructions (i.e., what are referred to as "analytical" constructions in diachronic linguistics). However, as these representations are used, larger constructions are acquired and stored. According to Reitter, Keller, and Moore's [25] ACT-R model of syntactic priming, the ability to process larger, more complex constructions is facilitated by syntactic rules stored in declarative memory. ACT-R allows for routinization of such rules into procedural memory for faster processing.

Assuming a single, unified architecture, Chang, Dell, and Bock's neural network model of sentence production can be understood as a realization of routinized language processing in procedural memory. The ACT-R model, by contrast, requires routinization to make plausible predictions about the timing of grammatical encoding in language production. Given ACT-R's assumptions of the timescale under which declarative memory operates, making fine-grained grammatical decisions that involve many retrievals from declarative memory would be unrealistically slow.

Note that one prediction from the ACT-R account of grammatical encoding is that syntactic priming does not occur for routinized constructions, as priming occurs only in declarative memory. This prediction is borne out empirically: it is difficult to detect syntactic priming in humans for high frequency syntactic constructions [18]. However, the exact dynamics of the ACT-R routinization process in the domain of language processing (i.e., the transfer of language knowledge from declarative to procedural memory) has not been validated against experimental data.

The choice between high-dimensional neural models of memory versus ACT-R leaves us with a trade-off between (in the case of neural models) a black box that fits large-scale data better and (in the case of ACT-R) transparent representations that can be interpreted by the scientist but account for large-scale language data less well. Models that use vector-symbolic architectures [12], such as BEAGLE [17, 20], provide a middle ground of systems that are more interpretable than conventional neural networks and more scalable the traditional symbolic models like ACT-R [19].

3.3. Bilingualism

Bilingual speakers pose a challenge for cognitive modelling. Bilinguals can be observed to *code-switch* at opportune times within a sentence, from one language into another. Such code-switching plausibly interacts with the way language is stored and processed, providing a behavioural window into the compressed, shared structural and lexical representations of several languages [24]. A plausible approach to modelling code-switching is deep neural networks, which can represent linguistic information in high-dimensional spaces and can be trained on large-scale language data.

3.4. Prediction and surprisal

Generally, the success and widespread use of neural language models as an engineering solution to predict, for example, words within a sequence of words [22, 23], is a strong argument to examine them as possible plausible cognitive models. It is widely recognized that *expectations* guide language processing [13, 21], which is observable, for example, in eye-tracking [9]. A language user's *surprisal* can be observed in response to novel information contained in a sentence, and sensitivity to information density in language is evident from behaviour [15, 27].

3.5. Language comprehension

The capacity to produce and comprehend language also tells us about working memory. Working memory must be sufficiently powerful to accommodate on-the-fly sentence construction and comprehension, while having capacity limits that reflect human cognitive constraints on sentence complexity [7]. Comprehension implies an ability to track a working model of whats being described in language by constructing an imaginary environment [16].

3.6. Discretization

Language also seems to imply an analytical process, capable of detecting units of statistical invariance at multiple levels of analysis: from letters and phonemes to morphemes, words, sentences or phrases, discourse, meaning, et cetera. These units and levels are discovered by human minds with little guidance. A cognitive architecture capable of language processing (and more) needs to be equally capable of discovering patterns in the environment at varying levels of analysis and abstraction.

4. Conclusion

The project of developing a common, computational model of cognition will play an important role in furthering the theoretical understanding of how language is acquired, comprehended, and produced by the human mind. Computational models, developed to explain behaviours such as the recency effect, priming, attentional bottlenecks, learning, recall, forgetting, and the inference of prototypes from exemplars, have already provided important insights into language acquisition, semantics and syntax, and the dynamics of word choice and grammatical structure in language production. A unified, computational, cognitive model will allow us to model how the systems of the mind interact to produce human language, giving us new theoretical insight into one of the most complex domains of human behaviour.

But just as psycholinguistic theory stands to benefit from the development of a common model, so too will language processing play an important role in testing and shaping the common model. The task of modelling language requires large-scale and long-term learning, as well as the capacity to abstract and generalize from exemplars. These demands suggest that traditional symbolic models of memory (such as proposed by the ACT-R cognitive architecture [1]) are insufficient. Neural language models have met with considerable success [23], suggesting that they should be taken seriously as cognitive models. Likewise, vector symbolic approaches are capable of handling large-scale language data [20].

The task of modelling language requires answering fundamental questions about the cognitive architecture. How does the brain allow for language to be processed (both produced and comprehended) so rapidly? Does this speed entail a central role for a procedural memory system? If so, what is the relative contribution of the declarative and procedural memories to language processing?

Prediction and surprise have been found to play a central role in language processing. Predictive neural language models have been highly successful at modelling language production. Is this indicative of a central role for prediction and surprise in the common model and cognition as a whole?

Language is discretized into phonemes, morphemes, words, sentences, and so on. What are the statistical mechanisms that underlie this process and are these the same mechanisms that underlie the parsing of the environment into objects?

A unified theory of cognition needs to address these questions, as well as other challenges posed by language processing, such as how bilinguals are able to develop cognitive representations that allow for a mapping between different languages, or how situational models are constructed in real time in working memory during language comprehension. The task of addressing these questions and challenges will help constrain the space of possible architectures and guide development towards a truly human-like common model of the mind.

5. Acknowledgments

The authors gratefully acknowledge funding from NSF grants SES-1528409 and BCS-1734304.

References

[1] Anderson, J.R., Lebiere, C., 1998. The atomic components of thought. Lawrence Erlbaum Associates, Mahwah, NJ.

[2] Atkinson, R.C., Shiffrin, R.M., 1968. Human memory: A proposed system and its control processes. Psychology of learning and motivation 2, 89–195. doi:10.1016/S0079-7421(08)60422-3.

[3] Bjork, R.A., Whitten, W.B., 1974. Recency-sensitive retrieval processes in long-term free recall. Cognitive Psychology 6, 173 – 189. doi:10.1016/0010-0285(74)90009-7.

[4] Chang, F., Dell, G.S., Bock, J.K., 2006. Becoming syntactic. Psychological Review 113, 234–272. doi:10.1037/0033-295X.113.2.234.

[5] Cole, J., Reitter, D., 2018. The timing of lexical memory retrievals in language production, in: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), pp. 2017–2027. URL: http://www.aclweb.org/anthology/N18-1183.

[6] Cole, J.R., Ghafurian, M., Reitter, D., 2017. Linking memory activation and word adoption in social language use via rational analysis, in: van Vugt, M.K., Banks, A.P., Kennedy, W.G. (Eds.), Proceedings of the 15th International Conference on Cognitive Modeling. University of Warwick, Coventry, United Kingdom, pp. 205–210. URL: https://core.ac.uk/download/pdf/148333135.pdf#page=214.

[7] Cole, J.R., Reitter, D., 2017. Examining working memory during sentence construction with an act-r model of grammatical encoding, in: van Vugt, M.K., Banks, A.P., Kennedy, W.G. (Eds.), Proceedings of the 15th International Conference on Cognitive Modeling. University of Warwick, Coventry, United Kingdom, pp. 211–216. URL: https://core.ac.uk/download/pdf/148333135.pdf#page=220.

[8] da Costa Pinto, A., Baddeley, A.D., 1991. Where did you park your car? analysis of a naturalistic long-term recency effect. European Journal of Cognitive Psychology 3, 297–313. doi:10.1080/09541449108406231.

[9] Demberg, V., Keller, F., 2008. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. Cognition 109, 193 – 210. doi:https://doi.org/10.1016/j.cognition.2008.07.008.

[10] Ebbinghaus, H.E., 1885. Das Gedchtnis: Untersuchungen zur experimentellen Psychologie. Duncker and Humboldt, Leipzig. Reprinted as Ebbinghaus H. E. (1964). Memory: A contribution to experimental psychology. New York: Dover.

[11] Firth, J.R., 1957. A synopsis of linguistic theory 1930–1955, in: Studies in linguistic analysis. Blackwell, Oxford, pp. 1–32.

[12] Gayler, R.W., 2003. Vector symbolic architectures answer jackendoff's challenges for cognitive neuroscience, in: Slezak, P. (Ed.), Proceedings of the Joint International Conference on Cognitive Science. University of New South Wales, Sydney, Australia, pp. 133–138. URL: http://cogprints.org/3983/.

[13] Hale, J., 2003. The information conveyed by words in sentences. Journal of Psycholinguistic Research 32, 101–123. doi:10.1023/A:1022492123056.

[14] Hintzman, D.L., 1984. Minerva 2: A simulation model of human memory. Behavior Research Methods, Instruments, and Computers 16, 96–101. doi:10.3758/BF03202365.

[15] Jaeger, T., Levy, R.P., 2006. Speakers optimize information density through syntactic reduction, in: Advances in Neural Information Processing Systems, pp. 849–856. URL: http://papers.nips.cc/paper/3129-speakers-optimize-information-density-through-syntactic-reduction.pdf.

[16] Johnson-Laird, P.N., 1980. Mental models in cognitive science. Cognitive science 4, 71–115. doi:10.1207/s15516709cog0401_4.

[17] Jones, M.N., Mewhort, D.J.K., 2007. Representing word meaning and order information in a composite holographic lexicon. Psychological Review 114, 1–37. doi:10.1037/0033-295X.114.1.1.

[18] Kaschak, M.P., Kutta, T.J., Jones, J.L., 2011. Structural priming as implicit learning: Cumulative priming effects and individual differences. Psychonomic Bulletin & Review 18, 1133–1139. doi:10.3758/s13423-011-0157-y.

[19] Kelly, M.A., Reitter, D., 2017. Holographic declarative memory: Using distributional semantics within act-r, in: Laird, J., Lebiere, C., Rosenbloom, P.S. (Eds.), The 2017 AAAI Fall Symposium Series: Technical Reports, The AAAI Press, Palo Alto, California. pp. 382–387. URL: https://aaai.org/ocs/index.php/FSS/FSS17/paper/view/16001.

[20] Kelly, M.A., Reitter, D., West, R.L., 2017. Degrees of separation in semantic and syntactic relationships, in: van Vugt, M.K., Banks, A.P., Kennedy, W.G. (Eds.), Proceedings of the 15th International Conference on Cognitive Modeling. University of Warwick, Coventry, United Kingdom, pp. 199–204. URL: https://core.ac.uk/download/pdf/148333135.pdf#page=208.

[21] Levy, R., 2008. Expectation-based syntactic comprehension. Cognition 106, 1126–1177. doi:10.1016/j.cognition.2007.05.006.

[22] Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J., 2013. Distributed representations of words and phrases and their compositionality, in: Burges, C.J.C., Bottou, L., Welling, M., Ghahramani, Z., Weinberger, K.Q. (Eds.), Advances in Neural Information Processing Systems 26. Curran Associates, Inc., pp. 3111–3119. URL: http://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf.

[23] Ororbia II, A.G., Mikolov, T., Reitter, D., 2017. Learning Simpler Language Models with the Differential State Framework. Neural Computation 29, 3327–3352. doi:10.1162/neco_a_01017.

[24] Putnam, M.T., Carlson, M., Reitter, D., 2018. Integrated, not isolated: Defining typological proximity in an integrated multilingual architecture. Frontiers in Psychology: Language Sciences doi:10.3389/fpsyg.2017.02212.

[25] Reitter, D., Keller, F., Moore, J.D., 2011. A computational cognitive model of syntactic priming. Cognitive Science 35, 587–637. doi:10.1111/j.1551-6709.2010.01165.x.

[26] Watkins, M.J., Neath, I., Sechler, E.S., 1989. Recency effect in recall of a word list when an immediate memory task is performed after each word presentation. The American Journal of Psychology 102, 265–270. URL: http://www.jstor.org/stable/1422957.

[27] Xu, Y., Reitter, D., 2018. Information density converges in dialogue: Towards an information-theoretic model. Cognition 170, 147–163. doi:10.1016/j.cognition.2017.09.018.