Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 145 (2018) 765–772

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)

The Common Model of Cognition and humanlike language comprehension

Peter Lindes*

University of Michigan, 2260 Hayward Street, Ann Arbor, MI 48109 USA

Abstract

The Cognitive Model of Cognition (CMC) describes a consensus among many researchers about architectural assumptions that define aspects of humanlike minds, whether natural or artificial. This consensus combines ideas from several existing cognitive architectures. Our research group has been developing a theory and an implemented system to do humanlike language comprehension within an embodied autonomous agent. This paper analyzes how this theory of humanlike language comprehension maps onto the CMC, what extensions to the CMC it might suggest, and what major challenges for the CMC come from applying it to language processing.

Peer-review under responsibility of the scientific committee of the 9th Annual International Conference on Biologically Inspired Cognitive Architectures.

Keywords: Common Model of Cognition; language comprehension; Soar; ECG

1. Introduction

Laird, Lebiere, and Rosenbloom [1] (LLR17) have defined what they call the Standard Model of the Mind (SMM). This model is intended as a more complete and refined version of a consensus on commonality between cognitive

* Corresponding author. Tel.: +1-801-755-6869; fax: +1-734-763-1260.
E-mail address: plindes@umich.edu

Peer-review under responsibility of the scientific committee of the 9th Annual International Conference on Biologically Inspired Cognitive Architectures.
10.1016/j.procs.2018.11.032

architectures reached at the 2013 AAAI Fall Symposium on Integrated Cognition. The SMM was derived specifically from a careful analysis of the commonalities between three prominent cognitive architectures: ACT-R [2], Soar [3], and Sigma [4]. There is an ongoing community working on further development of this model, and since LLR17 was submitted for publication the community has decided to change the name to the Common Model of Cognition (CMC). We will use that name in this paper, which is submitted to the 2018 AAAI Fall Symposium on A Common Model of Cognition.

LLR17 talks about different kinds of minds, but then states that the focus of the SMM/CMC will be humanlike minds, whether natural or artificial. Certainly an important aspect of human cognition is language comprehension, production, and acquisition. Therefore, studying how the CMC might be applied to language processing should help in the further development of the model.

In this paper we will examine a model of humanlike language comprehension we have developed within the Soar cognitive architecture. We will describe some of the workings of that model, and discuss how it relates to the CMC relative to three questions adapted from suggestions made by Paul Rosenbloom:

- Which consensus aspects of humanlike language processing directly map onto the common model as currently specified?
- Which aspects would suggest straightforward extensions?
- Which provide major challenges?

Finding "consensus" on any aspect of language processing is a challenge. We will explain briefly here the theory of language comprehension our group has developed, and how that theory is based on a long history of research in cognitive linguistics and related fields. We can't claim that the theory is necessarily a consensus.

In what follows we first give a brief description of the theory of humanlike language comprehension we have developed, including references to its roots in several fields of research. Our discussion will concentrate mainly on comprehension since that has been the focus of our research so far, but we will also suggest some ideas about the challenges presented by language acquisition. Next we will give a detailed analysis of how this theory of comprehension relates to the specific architectural assumptions of the CMC as presented in Table 1 of the LLR17 paper. Finally we will summarize the answers to the three above questions that emerge from this analysis and draw some conclusions.

2. A Theory of Language Comprehension

In our research group we are working on a cognitive model of human language comprehension [5, 6]. We believe this model does a reasonable job of modeling how the human comprehension process works. The model is built on theoretical principles developed by research in several fields over several decades, but here we will reference just a few key papers. Our theory has four main components:

- *Construction grammar*: Linguistic knowledge is made up of *constructions*. Each construction is a pairing of a form with a meaning. Constructions are linked in a hierarchical network of both lexical and composite constructions [7].
- *Incremental processing*: As much processing as possible is done one word at a time, on a single path through the possible search space, with local repairs being made when later words show that path to be mistaken. Words must be processed at a simulated real-time rate between about 140-160 words/minute for speech and 240-250 words/minute for reading to simulate adult human processing [8].
- *Cognitive architecture*: All processing is done within a cognitive architecture using general cognitive abilities, with knowledge of both language and processing being stored in long-term memories [9].
- *Embodiment*: Language processing is carried out within an embodied agent which must act on the results of comprehension, meaning the language must be grounded to the agent's perception and knowledge [5].

Specifically, our model is implemented using the Embodied Construction Grammar (ECG) [10] theory of how to represent syntax and semantics, and built within the Soar cognitive architecture [3].

We are working on two versions of the system, which work differently. System A translates a high-level description of the grammar in the ECG formalism, both syntax and semantics, entirely into Soar production rules. In this version all the linguistic knowledge is stored in procedural long-term memory, thus representing skilled performance. Additional hand-written processing rules work with the grammar-derived rules to implement the full comprehension algorithm.

System B, on the other hand, translates the same grammar description into data in long-term declarative memory. Then a different set of processing rules apply that knowledge dynamically. This approach has two major advantages. First, the grammatical information is retrieved from declarative long-term memory using spreading activation, making it possible to have ambiguities, such as multiple word senses, resolved through having situational context bias these retrievals. Second, Soar's chunking mechanism can be used to gradually convert the declarative linguistic knowledge into skilled procedural knowledge, thus simulating some aspects of language acquisition.

For the purposes of this paper, we will discuss the issues at a general level that can be easily related to the CMC. The next four sections will describe the theory and its implementation in Soar in more detail, along with an analysis organized according to the four main aspects of the CMC outlined in LLR17. The section labels starting with A through D and the paragraphs starting with A1 through D2 refer to the labels of categories and assumptions given in Table 1 of that paper.

3. Assumptions: A. Structure and Processing

3.1. How Our Theory Works

Our language comprehension model processes sentences in three nested loops. At the sentence level there are two phases: a comprehension process builds an internal model of the structure and meaning of the sentence that is grounded to the agent's perception and knowledge; then an interpretation process extracts from the grounded meanings an action message that the agent can act on to change its knowledge base, produce a linguistic response, and/or take an appropriate action in the physical world.

The comprehension process is carried out in an inner loop that processes one word at a time. Syntactic structure is built and meanings are constructed and grounded; as much as possible is done for each word before moving on to the next word. Each of these word cycles is made up of one or more construction cycles.

A construction cycle consists of three basic operations related to adding an instance of a single construction to the comprehension state at the current stage of processing: structure-building, meaning-building, and grounding. Structure-building consists of recognizing one or more constructions that match the current state, selecting among several if there are more than one, building an instance of the selected construction with its constituents, and generalizing this instance by labeling it with the labels of all its ancestors in the inheritance hierarchy of constructions. Sometimes, when no constructions match the cur-rent state, it may be necessary to perform a local repair before being able to recognize and build a new construction instance.

Each construction in the knowledge base consists of a description of the elements of linguistic form which must be recognized to instantiate this construction, along with a description of a meaning structure, called a schema, that should be evoked when that construction is instantiated. The meaning-building step, then, consists of instantiating the evoked schema and populating it according to the instructions given as part of the knowledge about the construction. Once this is done, the grounding process works to match up the linguistic data about this schema with the agent's current perception and its long-term world knowledge.

3.2. How It Relates to the CMC

A1: *The purpose of architectural processing is to support bounded rationality, not optimality.* Many computerized natural language processing systems strive for optimality rather than humanlike bounded rationality. Often they perform optimization over alternative parses for an entire sentence at once. Our incremental, single-path approach with local repair is definitely not this sort of approach, but a sort of rationality bounded by the temporal and memory limitations of the processing model.

A2: *Processing is based on a small number of task-independent modules.* The processing for this model of comprehension uses working memory, procedural memory, and in some cases the long-term declarative memory. It is driven by the cognitive cycle.

A3: *There is significant parallelism in architectural processing.* The Soar implementation uses a lot of parallelism in the production rules that fire to implement each language-processing operator.

A4: *Behavior is driven by sequential action selection via a cognitive cycle that runs at ~50 ms per cycle in human cognition.* Our system works using Soar's operator selection mechanism and decision (cognitive) cycle. Calculating its simulated processing rate based on 50 ms per cycle, it runs in the range of 130-150 words/minute, in the range of human speech rates.

A5: *Complex behavior arises from a sequence of independent cognitive cycles that operate in their local context, without a separate architectural module for global optimization (or planning).* Language processing operators are selected in each cognitive cycle based on the local context. This dynamic process, including local repairs when needed, achieves effective language comprehension within the domain we are working in.

In answer to our basic questions, our theory of humanlike language comprehension maps well directly onto aspects A1 to A5 of the CMC.

4. Assumptions: B. Memory and Content

4.1. How Our Theory Works

Fundamental to any incremental language comprehension model is a dynamic representation of the state of the comprehensional analysis of a sentence. The theory we have developed uses a graph structure in working memory of this comprehension state. This model is built with two kinds of nodes: *c-nodes*, each of which represents an instance of a linguistic construction, and *m-nodes*, each of which represents an instance of a meaning schema.

Each construction cycle creates a c-node, and most create an m-node as well. These nodes are linked together into a tree by parent-child compositional links, links from each c-node to its evoked m-node, a link from each c-node to the c-node which was the root of the tree when it was created, and grounding links from m-nodes to the concepts they are grounded to. The entire comprehension state at any time, even in an intermediate state, is a fully connected graph, with the most recently created c-node being a root from which all the other nodes can be reached.

4.2. How It Relates to the CMC

B1: *Declarative and procedural long-term memories contain symbol structures and associated quantitative metadata.* Our System A stores all linguistic and processing knowledge in procedural memory; only world knowledge is stored in long-term declarative memory. System B, on the other hand, stores all its grammatical knowledge in long-term declarative memory, and depends heavily on the use of the quantitative metadata for spreading activation. There is no notion of spreading activation in the CMC as specified in LLR17, although both ACT-R and Soar have such a feature. This seems like a good candidate for a "straightforward extension" to the common model.

B2: *Global communication is provided by a short-term working memory across all cognitive, perceptual, and motor modules.* Working memory is central to language processing as well as to cognition in general. Symbols and relations in the graph structure used as the comprehension state are an essential part of comprehension. For comprehension to work, the principle stated in LLR17 that "All of working memory is available for inspection and modification by procedural memory" is also essential, since the rules that do the structure and meaning building sometimes need to look at nodes several steps away.

The CMC has metadata for working memory, but how it is used is not well defined. Our implementation does not currently use this metadata, but it is important to simulate some more detailed aspects of humanlike processing, where human working memory has limits on its capacity and retention time. Christiansen and Chater [11] call this the "now-or-never bottleneck," saying that unless words and other units are composed into larger "chunks" quickly, they will be lost. To simulate this property of human processing, it is necessary to postulate some additional constraints on working memory that involve the metadata.

The main constraint is that unactivated nodes in working memory will decay over time and be lost. As linguistic units are composed into larger chunks, their activation may be maintained by spread from other nodes. Since the comprehension graph always has a root node which presumably has the main focus of attention, this attention can cause activation to spread through the rest of the graph, maintaining the nodes activated to at least a few levels deep.

B3: *Global control is provided by procedural long-term memory. a. Composed of rule-like conditions and actions. b. Exerts control by altering contents of working memory.* This model of control by procedural memory fits exactly our theory of comprehension processing.

B4: *Factual knowledge is provided by declarative long-term memory.* Declarative long-term memory serves our model well for both grounding to world knowledge and storing grammatical data. Full functionality, however, requires an extension to use spreading activation.

In this area, then, our language comprehension model maps well onto CMC assumption B3. However, some functionality under B1, B2, and B4 requires an extension to include spreading activation in both working memory and long-term declarative memory. Perhaps something modeled after what already exists in ACT-R or Soar would work well. For B2, a forgetting mechanism along with spreading activation seems to be needed in working memory.

5. Assumptions: C. Learning

5.1. How Our Theory Works

Our theory of language comprehension does not yet include a theory for how grammatical knowledge is learned. However, we speculate that a process beginning with understanding specific experiences, then storing generalizations over these data as declarative facts in long-term declarative memory, and finally compiling this knowledge somehow into procedural memory might be a good model.

5.2. How It Relates to the CMC

C1: *All forms of long-term memory content, whether symbol structures or quantitative metadata, are learnable.* Any learning theory consistent with our model will certainly require learning both symbol structures and quantitative metadata.

C2: *Learning occurs online and incrementally, as a side effect of performance and is often based on an inversion of the flow of information from performance.* This fits well with our current understanding of language acquisition.

C3: *Procedural learning involves at least reinforcement learning and procedural composition*. It is clear that some form of procedural composition, and perhaps reinforcement learning as well, will be required for a model of language processing where comprehension based on learned grammar can be performed in simulated real time.

There is a significant problem with this approach. For B1 above we said that spreading activation in declarative long-term memory is necessary for our System B to perform some of the kinds of ambiguity resolution that humans do easily. However, learning language processing as a skill [12] requires procedural composition to achieve real-time processing. The procedural long-term memory in the CMC, as well as in ACT-R and Soar, does not pro-vide any way for situational context to bias rule selection. Not only that, there is no way to modify already composed rules when new experience requires a change to some grammatical generalization. Solving this problem seems like a major challenge for the common model.

C4: *Declarative learning involves the acquisition of facts and tuning of their metadata.* Although declarative learning may play a role in specific increments of grammar learning, humans do not seem to have access to their grammatical knowledge, since it is stored as a skill rather than a set of facts. Thus this point is not very relevant to language.

C5: *More complex forms of learning involve combinations of the fixed set of simpler forms of learning*. This seems very general and reasonable. Any implications it might have for language are not presently apparent.

What we understand so far about language acquisition maps well onto principles C1 and C2 of the CMC. C4 and C5 seem reasonable, but do not affect very much our current theory. For C3, language processing presents a major challenge to the common model. Language processing is a learned skill in humans. Current architectural methods for accessing declarative long-term memory are too slow for real-time processing, so procedural composition seems necessary. However, com-posed procedural knowledge does not have the flexibility needed for either context-aware comprehension or incremental learning of the sort assumed by C2.

6. Assumptions: D. Perception and Motor

6.1. How Our Theory Works

Our comprehension model receives input words as fairly simple structures on Soar's input link, a form of working memory buffer. We do not yet have a humanlike model of language production, but such a model would certainly use a working memory buffer of some sort to drive language output.

6.2. How It Relates to the CMC

D1: *Perception yields symbol structures with associated metadata in specific working memory buffers.* Our model maps directly onto this assumption.

D2: *Motor control converts symbolic relational structures in its buffers into external actions.* Our theory does not yet include language production, but when it does it certainly seems that symbolic structures would drive motor control for speech articulation or writing.

7. Summary

To summarize our analysis, Table1 describes the mapping from what our theory says about humanlike language comprehension and the architectural assumptions about the Common Model of Cognition as presented by Laird, Lebiere, and Rosenbloom [1] (LLR17) in their Table 1.

Table 1: Mapping language comprehension to the CMC

A1	✓	B1	❄	C1	✓	D1	✓
A2	✓	B2	❄	C2	✓	D2	✓
A3	✓	B3	✓	C3	☒		
A4	✓	B4	❄	C4	✓		
A5	✓			C5	✓		

✓ Maps Well ❄ Needs Extension ☒ Major Challenge

Overall, the mapping is very good. This supports the validity of the common model itself, as well as the hypothesis that the domain-general abilities of a cognitive architecture are sufficient to form the basis of humanlike language processing. Nevertheless, we have found that there is an important area of extension needed, and a major challenge to how the com-mon model is now constructed.

The area of needed extension is that quantitative metadata needs to be structured and used for spreading activation in two ways. Within working memory the use of metadata can provide a model of how attention spreads within the graph of the dynamic comprehension state, allowing recent chunks to be accessible for recognizing constructions, while at the same time allowing less recent chunks to fade away so as to simulate limited working memory capacity. Within long-term declarative memory, spreading activation is needed to allow for retrievals to be biased by situational context.

The major challenge we see is the apparent conflict between the necessity to compose procedural knowledge to achieve skilled, real-time comprehension while at the same time maintaining the flexibility to allow for both contextual bias and the modification of procedural knowledge through incremental learning. Perhaps metadata could help, but LLR17 has no specifics on what this data is or how it is used.

8. Conclusions

The theory we have implemented only begins to approximate an accurate model of human language comprehension, but it suggests certain principles that a more accurate model should incorporate. Two kinds of memory are needed: a short-term memory with limited capacity to hold the current comprehension state as incremental processing proceeds, and a long-term store of grammatical knowledge. Processing mechanisms interact dynamically with these two memories to achieve comprehension. This is done rapidly in real time, and the long-term grammatical knowledge is a skill not accessible to conscious awareness.

At the core of processing is the construction cycle, where one or more constructions in the long-term store match the cur-rent comprehension state, and one is selected and instantiated to augment that state. Our System A implements this core operation of recognizing, selecting, and instantiating a construction using Soar production rules. This provides reasonable real-time performance and makes the long-term grammatical knowledge inaccessible, but it does not provide a way to use situational context to bias the selection, nor does it give a clear route to learning. Our System B stores the grammatical knowledge in Soar's semantic memory and uses spreading activation for recognition and selection. This allows for contextual bias and provides a more plausible route to learning, but it is slow and the grammatical knowledge is easily accessible.

What is needed is a set of memories and processing mechanisms that can provide recognition, selection, and retrieval of long-term knowledge driven by the short-term state. Both our implementations accomplish this much. However, the entire mechanism must allow for contextual bias, and provide a means for the long-term knowledge to be an acquired skill that keeps the long-term grammatical knowledge inaccessible. The definition of the CMC presented in by LLR17 does not give sufficient detail about how metadata works to be able to determine how it could help address this major challenge.

Although the solution to this challenge is not at all clear yet, we would like to speculate about a possible approach that, though radical, might be fruitful. Suppose all the grammatical knowledge were stored in a new form of declarative memory. Input word symbols, whether from recognition of speech or vision, would be sources of spreading activation to this memory. That spread would activate constructions, which in turn would evoke meanings as well as activate other constructions. This pattern of activation would represent the dynamic comprehension state with attention focused at its root. The results of this process would appear as perceptual input into the main working memory. Over time the patterns of activation would in-crease the rate of spread through given links so as to produce skilled performance. Thus the whole comprehension process would appear to the main model of cognition as a skilled form of perception whose details are not directly accessible.

It has been asked whether there are issues with interference between language processing and other task operators when a task is being performed at the same time. This is an important question, but it is beyond the scope of this paper since it applies not just to language processing but to any kind of multi-tasking situation. We also do not address here the implications for localist versus distributed representations, although our implementations certainly use a localist approach.

Acknowledgements

The work described here was supported by the AFOSR under Grant Number FA9550-18-1-0168. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressly or implied, of the AFOSR or the U.S. Government.

References

[1] Laird, John E., Christian Labiere, and Paul S. Rosenbloom (2017). A Standard Model of the Mind: Toward a Common Computational Framework across Artificial Intelligence, Cognitive Science, Neuroscience, and Robotics. *AI Magazine* 38(4).

[2] Anderson, John R. (2007). *How Can the Human Mind Occur in the Physical Universe?* Oxford University Press, Oxford and New York.

[3] Laird, John E. (2012). *The Soar Cognitive Architecture*. Cambridge, MA: The MIT Press.

[4] Rosenbloom, Paul S., Abram Demski, and Volkan Ustun (2016). The Sigma Cognitive Architecture and System: Towards Functionally Elegant Grand Unification. *Journal of Artificial General Intellilligence*, 7(1) 1-103. DOI: 10.1515/jagi-2016-0001

[5] Lindes, Peter, Aaron Mininger, James R. Kirk, and John E. Laird (2017). Grounding Language for Interactive Task Learning. *Proceedings of the 1st Workshop on Language Grounding for Robotics at ACL*.

[6] Lindes, Peter and John E. Laird (2017). Cognitive Modeling Approaches to Language Comprehension Using Construction Grammar. *Proceedings of the AAAI Spring Symposium 2017 track on Computational Construction Grammar and Natural Language Understanding.*

[7] Goldberg, Adele E. (2013). Constructionist Approaches. In Thomas Hoffman and Graeme Trousdale, eds, *The Oxford Handbook of Construction Grammar*. Oxford University Press, New York, pp. 15-31.

[8] Lewis, Richard Lawrence (1993). *An Architecturally-based Theory of Human Sentence Comprehension*. PhD dissertation in Computer Science, Carnegie Mellon University.

[9] Newell, Allen (1990). *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.

[10] Bergen, Benjamin and Nancy Chang (2013). Embodied Construction Grammar. In Thomas Hoffman and Graeme Trousdale, eds, *The Oxford Handbook of Construction Grammar*. Oxford University Press, New York, pp. 168-190.

[11] Christiansen, Morten H. and Nick Chater (2016). The Now-or-Never bottleneck: A fundamental constraint of language. *Behavioral and Brain Sciences (2016)*, 1-72.

[12] Chater Nick and Morten H. Christiansen (2018). Language acquisition as skill learning, *Current Opinion in Behavioral Science* **21**.