A Cognitive Substrate for Natural Language Understanding

Nicholas L. Cassimatis, Arthi Murugesan and Magdalena D. Bugajska
Human-level Intelligence Laboratory, Rensselaer Polytechnic Institute, Troy

Abstract. Our goal is to understand human language use and create systems that can use human language fluently. We argue that to a achieve this goal, we must formulate all of the problems for language use from morphology up to pragmatics using the same *cognitive substrate* of reasoning and representation abilities. We propose such a substrate and described systems based on it. Our arguments, results with real-world systems and ongoing work suggest that the cognitive substrate enables a significant advance in the power of cognitive models and intelligent systems to use human language.

Keywords. Natural language, cognitive substrate.

Introduction

Enabling computers to use and understand natural language would greatly increase our ability to model human language use and the power of the human-computer interface. In this paper we will refer to the goal of creating computers that can use and understand natural language as human-level language use (HLU). Achieving HLU will require serious advances in the state of the art in artificial intelligence and cognitive modeling. One consequence of these advances will be inferential and representational computational abilities that can be applied to other interface modalities.

The following is a (simplified) enumeration of computational tasks involved in HLU:

1. Infer phonemes, morphemes and words being uttered from an acoustic signal.
2. Infer the syntactic structure the words were intended to form.
3. Given knowledge about the world, people and the specific participants of a conversation, infer the goal of the speaker which generated his utterance.
4. Given a goal, produce an utterance that achieves it.
5. Articulate the utterance.

In this paper, we will focus on the goal of inferring syntactic structure and speaker intention, partly because we suspect the insights required to solve these problems will enable fairly straightforward solutions to the other problems.

1. Current AI is not adequate for HLU

Let us know examine the state of the art in artificial intelligence and evaluate whether it is adequate to the task of the solving these problems. Many of the individual observations we will make are obvious to many. They are mentioned here to consider their ramifications for achieving HLU.

1.1. Parsing and intention recognition are closely related

The goal of recognizing the syntax of a sentence and recognizing the speakers' intent are closely related. Consider the sentence:

"Show me the weather for the Red Sox game."

Does "for the Red Sox game" modify "the weather" or "show"? That is a syntactic question; but to answer it you have to infer that the speaker intends you to show him the weather forecast involving the Red Sox game ("for …" modifies "the weather") and does not intend you to show him the weather in order to somehow facilitate the Red Sox game ("for …" modifies "show", as in "show me the weather for the second time").

1.2. Modern algorithms for inference and parsing are incompatible

The algorithms used in AI for parsing are incompatible with those used in inference. For now, we consider "modern" mainstream AI algorithms and later discuss algorithms based on more structured formalisms such as scripts and frames. Probabilistic and statistical approaches are popular in AI today, with Markov chain Monte Carlo (MCMC) methods being perhaps the most popular. However, in parsing research, MCMC is shunned for parsers of probabilistic context-free grammars. This is because MCMC (and other algorithms for propagating beliefs in graphical models) require that all of the relevant state variables in a problem be identified before inference proceeds. However, in parsing we do not have all the state variables in advance. We are simply given a string of words and it is our job to infer is which phrases those words were intended to utter. For MCMC to be used to find the most likely parse of a sentence, it must be given the possible parses of a sentence to begin with. For example, to decide whether "the dog" form a noun phrase or whether, as in "the dog food", they do not, MCMC would need to have both possibilities specified in advance. But that is precisely the job of a parser. Even if you had all (or most) of the possible parses, the size of the corresponding graphical model MCMC would be operating over would probably be prohibitively large. (More on this later). Thus, devotees of the statistical approach in the parsing community must use a different inference algorithm than their kin in the rest of AI.

The upshot of all this is that it is very difficult to guide parsing using perceptual information or world knowledge because the "general" inference algorithms for those are different than the algorithms used by the parsing community.

1.3. Dynamism and scalability

The root cause of the tension between PCFG parsers and MCMC is that the latter are not dynamic[1] (i.e., they do not enable unforeseen objects) to be inferred and do not scale well as the number of state variable increase. These problems are not specific simply to probabilistic inference algorithms but also apply, e.g., to SAT-based search

---

[1] Dynamic Bayesian Networks are not dynamic in the present sense because they assume that all the state variables are fixed in advance and make inference about how these unfold through time.

algorithms. Recent approaches come closer to addressing these problems but do not address them fully. For example, [1] only deals with the case were new objects are perceived, not when they are actually inferred. [2] does not enable new objects to be introduced at all, but only delays the representation of grounded relations involving those objects until they are needed.

1.4. World knowledge and scalability

The amount of world knowledge and the consequent size of the inference problem are enormous. Consider again our example sentence: "Show me the weather for the Red Sox game." To retrieve a weather forecast, one requires a time and a location. Neither of these is explicitly specified in the sentence. The Red Sox could be playing away from Boston, the speaker might not even care about the weather for the game itself, but for the reception of his television satellite, say, in St. Louis, as he watches the game. Which game? The next game, the last game, or the game for which the speaker has tickets? Finally, almost anything can be relevant to answering this question. If marketing for a new kind of egg white product is being debuted during a Red Sox game in two weeks and the speaker is a commodity trader, he might be referring to that game. Thus, quite literally, the price of eggs in China could be relevant to even the most prosaic utterance.

This example illustrates that for any given utterance a wide variety of knowledge, involving the speaker's history and goals, typical intentions of speakers, the behavior of objects and event can all be relevant to using language.

What this means is that the state/search/constraint space for the HLU problem is enormous. To illustrate, let us assume very conservatively, that people know on the order of one million facts (as does CYC), that each fact is expressed with about 4 open variables that can each range over 100 objects and that there are 1 billion seconds in a lifetime. This amounts to 100 billion trillion (10^23) concrete constraints. That is well beyond the scope of all general inference algorithms available today (and, we suspect, ever).

The problem is even made worse to the extent that understanding a sentence requires an understanding of multiple people's mental states. For example, if the listener knows that the speaker thinks a glass of water is a glass of gin, then he will understand who "the man with the gin" refers to. Each person's beliefs thus become a factor in the understanding of an utterance, which adds at least an integral multiple to the calculations in the last paragraph. It is worse, however. The listener's beliefs about the speaker's beliefs about the listener's beliefs about … can potentially be relevant. The regress is infinite and, formally at least, so is the state/constraint/search space.

HLU will thus require inference algorithms that are extremely scalable.

Integration

One solution is to abandon general inference algorithms for approaches based on larger structures such as scripts or frames. However, these approaches, *by themselves,* have proven to be too rigid and inflexible. When a situation changes slightly, a structure that worked might, *by itself,* not work.

What is needed is a combination of both approaches: the speed of structured approaches along with the flexibility of more general approaches.

Most researches who have taken integration seriously have taken a "modular" approach. They build systems that have, for example, a syntax module that outputs possible parses to other (e.g., semantic and pragmatic) modules that settle on the most likely parse and attach a meaning to it. The problem with this approach is that information from each module is needed to constrain information in the others. The flow of information needs to be deeper and more frequent than the modular approach makes easy to achieve.

So far, we have enumerated many grounds for pessimism. However, we at least have come to a deeper understanding of what the obstacles. We need better inference (more dynamic, scalable, structured and flexible) and a way to tightly integrate syntactic, semantic and pragmatic inference. The next section outlines an approach.

In the preceding discussion, we argued that two key steps to achieving HLU are developing common sense inference algorithms that are scalable, flexible and dynamic and closely integrating inference required for all aspects of language use. In our work, we are addressing these two tasks using the Polyscheme cognitive architecture [3] and the notion of a cognitive substrate [4].

A cognitive substrate for integrating reasoning about language with common sense reasoning

Every step of understanding a sentence can be influenced by a wide variety of factors. Deciding which sense of a word to use in parsing a sentence, for example, can be influenced by the perceived state of the world, statistical regularities in word use, the beliefs and goals of the people involved in the conversations and knowledge about how people and the rest of the world work.

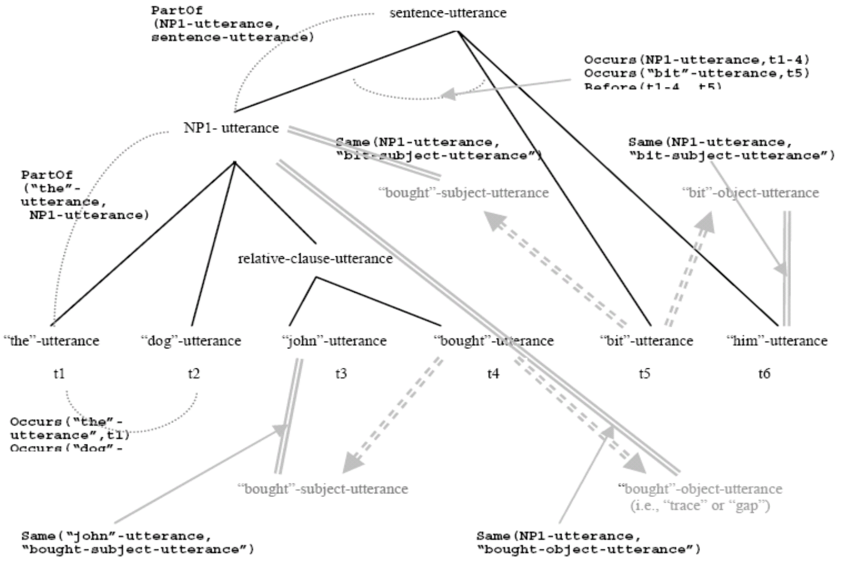

Figure 1: Syntactic structure expressed using substrate representations.

Our goal is to formulate each of these kinds of constraints on language interpretation using the same "language" or "framework". By so doing, language understanding becomes not several inference problems (e.g., parsing, intention recognition, etc.) with algorithms for solving each encapsulated within separate modules, but a single inference problem in which every aspect of the problem can constrain every other.

To exploit this fact computationally, we needed to find a way to represent the superficially very different realms of cognition, e.g., syntax and people's beliefs, using the same basic conceptual elements. We have achieved this by showing how to reduce these conceptual fields into the same basic cognitive substrate [4]. This substrate includes reasoning and representation abilities for time, space, causality, identity, parthood and desire. For example, [5] has shown how the problem of syntactic parsing can be reformulated straightforwardly using the same concepts used to formulate constraints on physical actions by agents. Figure 2 shows how to reduce a grammatical rule in HPSG to this substrate. Figure 1 shows an example of a sentence being described in terms of substrate concepts.

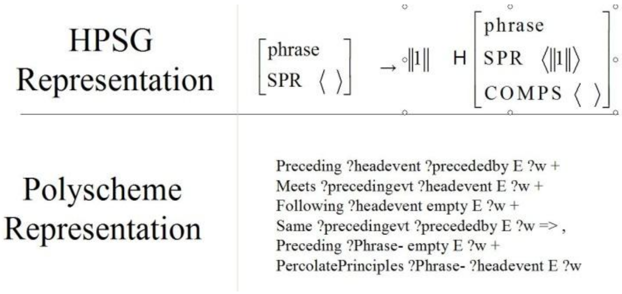

Figure 2: An HPSG rule formulated using substrate concepts.

Table 1 illustrates Cassimatis' [6] mapping between elements of syntactic structure and substrate concepts.

We have also shown [7] how to reduce representations of beliefs to this same basic substrate of concepts. For example, we can represent belief using a combination of identity and counterfactuality. To say John believes Fred is Honest, we say that in the counterfactual world where John is like Fred (Same(John,Fred,world)), that Fred is honest (Honest(Fred,w)).

These kinds of reductions to substrate representations allow all aspects of language understanding to be represented within the same language, making it easier to let each influence the other.

Table 1: Mapping of syntactic structures onto syntactic structures.

Grammatical Structure	**Cognitive structure**
Word, phrase, sentence	Event
Constituency	Meronomy
Phrase structure constraints	Constraints among (parts of) events
Word/phrase category	Event category
Word/phrase order	Temporal order
Phrase attachment	Event identity
Coreference/binding	Object identity
Traces	Object permanence
Short- and long-distance dependencies	Apparent motion and long paths.

2. Implementing a substrate in Polyscheme

A substrate that enables all the factors involved in using language to mutually constrain each other must conform to the algorithmic requirements described above. It must integrate the best characteristics of several classes of algorithms by being fast, flexible, scalable and dynamic. In this section we briefly outline how this integration is achieved in the Polyscheme cognitive architecture and direct readers who want more detail to other sources [3, 4, 8].

The fundamental idea behind Polyscheme is that many different kinds of algorithms can be implemented as sequences of attention fixations. These attention fixations in Polyscheme each concentrate multiple computational resources on a single aspect of the world. When algorithms are executed as sequences of attention fixations, integrating these algorithms is as simple as interleaving these sequences.

This enables hybrid algorithms that combine the benefits of multiple algorithms. For example, consider the tension between a general search algorithm and a script-based reasoner. The search algorithm is general and can adapt to many changes in circumstances but is very slow because meaningful state spaces are prohibitively large. The script-based reasoner will work quickly and effectively when the appropriate script exists for a particular situation. However, slight changes in a situation can invalidate a script and we cannot expect each of these changes to have been anticipated in advance. Our goal is to be able to achieve the generality and flexibility of general search without sacrificing the speed of more structured reasoning algorithms.

In Polyscheme, we can achieve this by implementing each algorithm as a sequence of attention fixations. We can look at applying a script as a special case of search in which at each step the operator is chosen from the script. If, when applying a script a problem occurs at a particular step, we can begin to choose operators that a more general search algorithm would choose. The result will be that the elements of a script that are not appropriate for a particular situation will be repaired by the integrated search algorithm. Figure 3 illustrates this process.

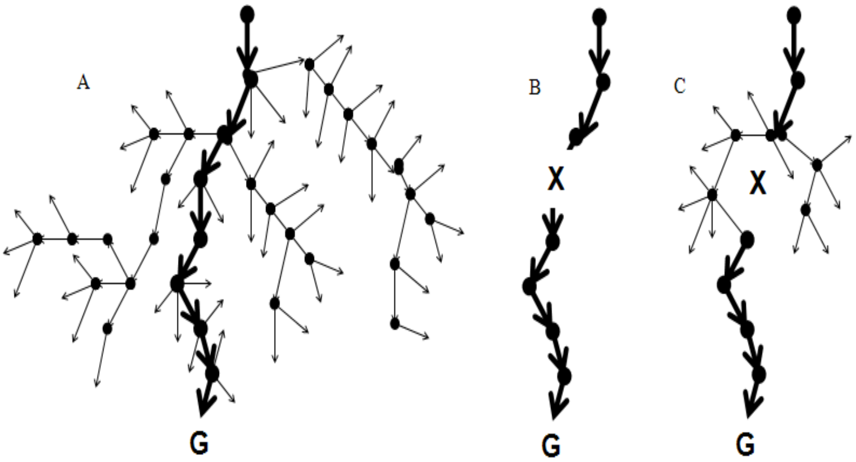

Figure 3: A hybrid (C) of systematic, general, flexible but slow search (A) with fast but rigid case-based reasoning (B).

In ongoing work, we have used this approach to address the tension alluded to above between PCFG parsers and more general probabilistic inference. The fundamental issues that prevent general inference algorithms from parsing PCFGs is that they cannot in the middle of inference add new objects to their reasoning. PCFG parsers constantly do this (e.g., when they infer that a determiner and a noun imply the existence of a noun phrase). Our approach has been to treat PCFG parsing as a maximization problem. We want to find the parse of a sentence with the highest probability of being the parse intended by the speaker. This has led us to implement WalkSAT-inspired algorithm in Polyscheme that maximizes the probability of parses. By integrating this algorithm with a rule matcher we have been able to represent the parsing problem using relational first-order representations instead of propositionally. This also enables objects and relations to be instantiated in memory only when they are needed. Because nonlinguistic relationships can be captured using the rule system, the algorithm automatically takes into account nonlinguistic information when making linguistic inferences. From its point of view, there is no difference between the two types of constraints. Finally, because each step of our algorithm is implemented by a focus of attention that includes information from memory or perceived from the world, this approach enables an embodied approach to sentence processing all the benefits of modern PCFG and probabilistic inference research.

This kind of integration in Polyscheme enables suitable cognitive substrate to be implemented. As alluded to before, the substrate implementation we have so far has enabled a unified and integrated account of physical, epistemic and syntactic inference.

3. Implemented systems

In our preliminary work combining parsing and general inference, we have been able to formulate grammatical rules and world knowledge in the same substrate and develop a system that used the world knowledge to disambiguate the meaning of a sentence. For example, with knowledge about mechanical devices and the relevant syntax and semantics, our parser could find the right sense of “bug” in “the bug needs a new battery”. Confirming the power of the substrate approach, nothing special needed to be done to the grammar rules or the world knowledge to make them interact. It was a zero-effort natural consequence of formulating each within the substrate.

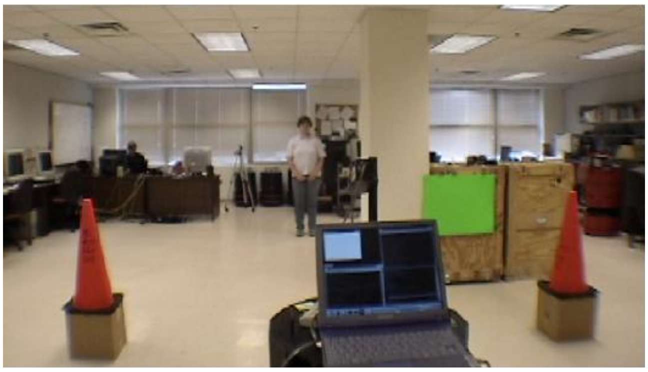

Figure 4: The person cannot see everything the robot can from its perspective.

The last section described how basing a system on a substrate implemented in Polyscheme enables language understanding to be facilitated by rich, embodied inference. This approach was used to enable human-interaction with a mobile robot. As is typically done in Polyscheme, a focus of attention was used to implement inference algorithms for semantic understanding and inference about the world. This focus of attention concentrated several computational resources, including a perceptual input from a video processing system. The result was that visual and spatial information was used to infer a person's perspective and use all that information was used to disambiguate natural language references. For example, in one situation (illustrated in Figure 4), a robot could see two cones (A and B) while it could see that the person it was interacting with only ever saw one cone (B). When the person said "the cone", the robot was able to use perceptual information and spatial inference to infer that the robot was looking at cone B. This required no special encoding of semantic information. The inference was simply a byproduct of Polyscheme's ability to implement a cognitive substrate in which several kinds of constraints (linguistic, epistemic and spatial) could be given a unified formulation.

4. Conclusions

The arguments we have presented, the results with real-world systems we have just described and ongoing work demonstrate that a cognitive substrate and a suitable inference mechanism enable natural language capabilities that will greatly improve the interface between humans and computers.

References

[1] B. Milch, B. Marthi, D. Sontag, S. Russell, D. L. Ong, and A. Kolobov, "BLOG: Probabilistic Models with Unknown Objects," presented at IJCAI-05, Edinburgh, Scotland, 2005.

[2] P. Domingos and M. Richardson, "Markov Logic Networks," *Machine Learning*, vol. 62, pp. 107-136, 2006.

[3] N. L. Cassimatis, "Integrating Cognitive Models Based on Different Computational Methods," presented at Twenty-Seventh Annual Conference of the Cognitive Science Society, 2005.

[4] N. L. Cassimatis, "A Cognitive Substrate for Human-Level Intelligence," *Artificial Intelligence Magazine*, vol. 27, 2006.

[5] A. Murugesan and N. L. Cassimatis, "A Model of Syntactic Parsing Based on Domain-General Cognitive Mechanisms," presented at 8th Annual Conference of the Cognitive Science Society, Vancouver, Canada, 2006.

[6] N. L. Cassimatis, "Grammatical Processing Using the Mechanisms of Physical Inferences," presented at Twentieth-Sixth Annual Conference of the Cognitive Science Society, 2004.

[7] P. Bello and N. L. Cassimatis, "Developmental Accounts of Theory-of-Mind Acquisition: Achieving Clarity via Computational Cognitive Modeling," presented at 28th Annual Conference of the Cognitive Science Society, Vancouver, Canada, 2006.

[8] N. L. Cassimatis, J. G. Trafton, M. Bugajska, and A. C. Schultz, "Integrating Cognition, Perception, and Action through Mental Simulation in Robots," *Robotics and Autonomous Systems*, vol. 49, pp. 13-23, 2004.