A Formal Model of the Mechanism of Semantic Analysis in the Brain

Yuuji Ichisugi and Naoto Takahashi

National Institute of Advanced Industrial Science and Technology (AIST), Central 1,Tsukuba,Ibaraki 305-8560, Japan
{y-ichisugi,naoto.takahashi}@aist.go.jp

Abstract. We propose a formal model of the mechanism of semantic analysis in the language areas of the cerebral cortex. The framework of Combinatory Categorial Grammar, a framework of grammar description in theoretical linguistics, is modified so that it does not use lambda calculus to represent semantic rules. This model uses a novel form of semantic representation named hierarchical address representation, and uses only fixed-length data structures. The knowledge of syntax and the knowledge of semantics are clearly separated in this model. Therefore, it is possible to reproduce disorders specific to syntax (utterance similar to Broca's aphasia) and disorders specific to semantics (utterance similar to Wernicke's aphasia) by disabling different modules in the model. We estimate that the model can be implemented using the Bayesian network model of the cerebral cortex that we have proposed earlier. We believe that this research will connect computational neuroscience and theoretical linguistics, and greatly evolve both of them.

1 Introduction

The language areas, which are considered to be centers of human language activities, are parts of the cerebral cortex. There is a hypothesis[3] that claims "the cerebral cortex is a kind of *Bayesian network*." If so, we must be able to build a system that reproduces the behavior of the human language areas using a Bayesian network. Thus we aim at constructing a system that processes *Combinatorial Categorial Grammar(CCG)*[2], a framework of grammar description, using a Bayesian network[9]. As the first step towards the aim, we propose a formal model of the mechanism of semantic analysis in the brain based on CCG.

Theoretical linguistics is one field of linguistics that analyzes the characteristics of natural languages by mathematical methods. The relation between linguistics and theoretical linguistics resembles the relation between neuroscience and computational neuroscience. One purpose of theoretical linguistics is to find out some characteristics shared by all existing natural languages. Such characteristics can be considered as the characteristics of the information processing of the language areas in the brain.

Lambda calculus is usually used as a tool to describe semantic rules of CCG; however, it is difficult to handle variable-length data structures like lambda terms

This is a pre-print of a contribution published in Samsonovich A. (eds) Biologically Inspired Cognitive Architectures 2018 (BICA 2018), published by Springer.
The final authenticated version is available online at: https://doi.org/10.1007/978-3-319-99316-4_17

by Bayesian networks. The proposed model uses a novel form of semantic representation named *hierarchical address representation*, which does not use lambda terms. The model uses only fixed-length data structures. We estimate that the model can be implemented within the framework that we have proposed[5][9], which is based on Bayesian networks.

Although it is unknown whether the proposed method is applicable to the syntactic rules and the semantic rules of all the existing natural languages, we believe that this research will connect computational neuroscience and theoretical linguistics and greatly evolve both of them.

Sec.2 briefly explains CCG. Sec.3 describes the proposed model and examples of analysis of some sentences. Sec.4 describes the correspondence between the modules of the model and some areas of the cerebral cortex, then we demonstrate reproduction of utterance of aphasia by disabling several modules.

2 Combinatory Categorial Grammar (CCG)

CCG is one of the most successful frameworks of grammar description. Its expressive power of grammar description is "mildly context-sensitive", which locates in between context-sensitive and context-free in the Chomsky hierarchy. Although the framework is very simple, grammars defined in CCG successfully explain many language phenomena (even though it is not complete). Therefore, we consider that CCG is the theory of information processing of the language areas in the brain.

In theoretical linguistics, some frameworks use *unification* as the core operation of analysis, and they are called *unification grammars*. CCG is one of unification grammars.

In CCG, general syntactic categories have structures that consist of ground categories (e.g., S for sentence, NP for noun phrase) combined by the operators "/" and "\". Theoretically, the length of a syntactic category is not restricted.

A ground category may have *syntactic features*. A syntactic feature may be a discrete variable whose value will be determined by the unification operations during the process of syntactic analysis. A ground category G with a syntactic feature F is denoted as G_F.

Production rules are defined by the form of *inference rules* in CCG. Syntactic analysis (i.e. parsing) is formalized as proof search showing a given word sequence being a sentence. A parse tree obtained as a result of syntactic analysis corresponds to a proof diagram.

The inference rules in CCG are accompanied by semantic rules that compose meaning of phrases. For example, the function which is the meaning of the word "black" : $\lambda x.black(x)$ is applied to the term that means "cats" : $cats$ to get the meaning of the phrase "black cats" : $black(cats)$. The semantic representation of the whole sentence is obtained by performing function applications (i.e. beta reductions) or function compositions sequentially from the leaves to the root, along with the parse tree.

In theoretical linguistics, including CCG, lambda calculus has been used as a tool for describing semantics; however, it is hard to imagine that the actual neural networks in the language areas have realized the complicated lambda calculus. In the following section, we propose a model of semantic analysis that does not use lambda calculus.

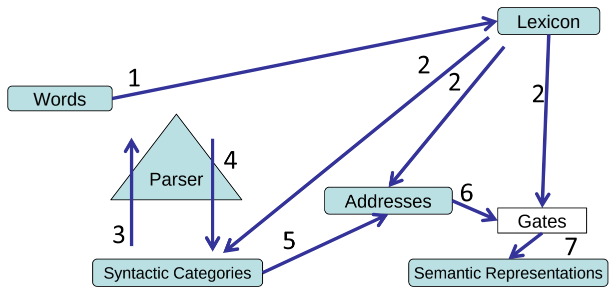

Fig. 1. The architecture of the model and typical information flow among the modules during the parsing process.

3 The proposed model

3.1 Scope of the model

We aim at constructing a model of the mechanism for unconscious and instant interpretation of the superficial and literal meaning of comparatively simple sentences. Neither conscious interpretation of sentences with complicated structures nor presumption of unexpressed intention is a target of this model.

The length of word sequences, the length of syntactic categories, and the length of generated semantic representations are all limited. Although human can interpret a long sentence incrementally from the head, this model assumes that all words are given at once.

We adopt a simplified English grammar for a straightforward explanation of the behavior of the proposed model. Although the model accepts complex sentences consisting of two clauses connected with a subordinating conjunction ("if", "after", etc.) and sentences containing relative pronouns ("which", etc.), it is assumed that a subordinating conjunction or a relative pronoun appears at most once in a sentence.

3.2 The outline of the process of semantic analysis

In this model, syntactic analysis and semantic analysis are conducted simultaneously. Fig.1 shows a typical information flow among the modules during parsing. (1) First, a word sequence is given. (2) By referring to the lexicon, three data

Address	SR
$(sconj, -, -)$	if
$(c1, agent, size)$	big
$(c1, agent, color)$	*
$(c1, agent, entity)$	dogs
$(c1, modality, -)$	*
$(c1, action, -)$	chase
$(c1, patient, size)$	small
$(c1, patient, color)$	*
$(c1, patient, entity)$	mice
$(c2, agent, size)$	*
$(c2, agent, color)$	black
$(c2, agent, entity)$	cats
$(c2, modality, -)$	may
$(c2, action, -)$	eat
$(c2, patient, size)$	*
$(c2, patient, color)$	*
$(c2, patient, entity)$	mice

Fig. 2. All addresses in the prototype system and semantic representations obtained as a result of the semantic analysis of the sentence "if small mice areChasedBy big dogs black cats may eat mice". Undetermined values are denoted by "$*$".

$$> \frac{\mathbf{X/Y} \quad \mathbf{Y}}{\mathbf{X}} \qquad < \frac{\mathbf{Y} \quad \mathbf{X\backslash Y}}{\mathbf{X}}$$

$$> B\ \frac{\mathbf{X/Y} \quad \mathbf{Y/Z}}{\mathbf{X/Z}} \qquad < B\ \frac{\mathbf{Y\backslash Z} \quad \mathbf{X\backslash Y}}{\mathbf{X\backslash Z}}$$

$$> T\ \frac{\mathbf{X}}{\mathbf{T/(T\backslash X)}} \qquad < T\ \frac{\mathbf{X}}{\mathbf{T\backslash (T/X)}}$$

Fig. 3. The inference rules in the proposed model. These are identical to the inference rules of the usual CCG except that there are no semantic rules using lambda calculus.

structures, a syntactic category, an *address* and a *semantic representation*, are obtained for each word. Variables may be unbound in syntactic features and in addresses at the time. (3) The parser merges syntactic categories and finds a parse tree in which the whole word sequence forms a sentence. During this process, unbound syntactic features receive a value through unification operations. (4) If a complete parse tree is generated, all values of syntactic features are determined. (5) All values of variables contained in the addresses are also determined. (6) A set of pairs of addresses and semantic representations that represents the meaning of the whole sentence is obtained. (7) Each semantic representation is written into the "memory" at the corresponding address.

Generally speaking, the information flow is not limited in the direction above. For example, when a prior knowledge for the meaning is given, the information will flow backwards from the semantic representation module to the parser module and the parser can use it to resolve lexical or syntactic ambiguities. Moreover, it is also possible to infer appropriate word sequences when a semantic representation is given. We show such an example in Sec. 4.

3.3 Hierarchical address representation, inference rules and lexicon

In this model, the meaning of a sentence is represented as a set of pairs of addresses and semantic representations. We call this form "hierarchical address representation".

Each address is a tuple of the three values $(\mathbf{C}, \mathbf{R}, \mathbf{F})$. Each semantic representation is written into an address. All possible addresses in the prototype system are listed in Fig.2.

The topmost hierarchy of an address, $\mathbf{C}$, represents the index of a clause in a complex sentence, which uses either a subordinate conjunction (e.g. "if") or a relative pronoun (e.g. "which"). The value of $\mathbf{C}$ is either $c1$ (for the first clause) or $c2$ (for the second). The type of the subordinate conjunction is written into the special address: $(sconj, -, -)$. For example, when the sentence starts with an "if", the semantic representation of "if" is written into the said special address.

The second hierarchy of an address, $\mathbf{R}$, primarily indicates the *semantic role* of a word in the sentence. In the current prototype system, its value is limited to either *agent* or *patient*, but it can be extended for other semantic roles (e.g. instrument, location, time) easily. Also, $\mathbf{R}$ can take the value *action* (for the semantic representation of a verb) or *modality* (for the semantic representation of an auxiliary verb).

When the value of $\mathbf{R}$ is *agent* or *patient*, i.e. when the word is a part of a noun phrase, the third hierarchy, $\mathbf{F}$, indicates a *feature* of the word. In the prototype system, the value of $\mathbf{F}$ is limited to *color*, *size*, and *entity*. $\mathbf{F}$ being *entity* means that the word is the head noun of a noun phrase.

The hierarchical address representation is inspired by a neuroscientific finding about the encoding of sentence meaning in the brain[8], which suggests that each semantic role (agent, patient) has its own representing place in the brain, regardless of the superficial voice (active or passive) of the sentence. We suppose that fixing positions for all elements of meaning facilitates learning and processing of language for neural networks and Bayesian networks in the brain.

Fig.3 shows the inference rules in the proposed model. There are no semantic rules using lambda calculus. The semantic analysis is performed by merely unification operations of syntactic features as explained in the next subsection.

The *lexicon* is a set of *lexical items*. A lexical item in this model is represented as a tuple of four data structures (a word, a syntactic category, an address, and a semantic representation), as shown in Fig.4.

3.4 Examples of analysis

Fig.5 shows the parse tree of the sentence "black cats eat mice".

Let us explain the process of analyzing the phrase "black cats" in detail. First, by referring to the lexicon for these words, the corresponding syntactic categories $NP_{\mathbf{C}_1,\mathbf{R}_1}/NP_{\mathbf{C}_1,\mathbf{R}_1}$ and $NP_{\mathbf{C}_2,\mathbf{R}_2}$ are obtained. Next, the inference rule ">" is chosen because this rule has the premises that unify with the obtained syntactic categories. Then, the inference rule is applied to the syntactic categories and the merged syntactic category $NP_{\mathbf{C}_1,\mathbf{R}_1}$ is derived as a result. This derived syntactic

Word	Syntactic Category	Address	SR
'if'	$(S_{c1}/S_{c2})/S_{c1}$	$(sconj, -, -)$	if
'black'	$NP_{\mathbf{C},\mathbf{R}}/NP_{\mathbf{C},\mathbf{R}}$	$(\mathbf{C}, \mathbf{R}, color)$	black
'big'	$NP_{\mathbf{C},\mathbf{R}}/NP_{\mathbf{C},\mathbf{R}}$	$(\mathbf{C}, \mathbf{R}, size)$	big
'cats'	$NP_{\mathbf{C},\mathbf{R}}$	$(\mathbf{C}, \mathbf{R}, entity)$	cats
'eat'	$(S_{\mathbf{C}}\backslash NP_{\mathbf{C},agent})/NP_{\mathbf{C},patient}$	$(\mathbf{C}, action, -)$	eat
'areEatenBy'	$(S_{\mathbf{C}}\backslash NP_{\mathbf{C},patient})/NP_{\mathbf{C},agent}$	$(\mathbf{C}, action, -)$	eat
'may'	$(S_{\mathbf{C}}\backslash NP_{\mathbf{C},\mathbf{R}})/(S_{\mathbf{C}}\backslash NP_{\mathbf{C},\mathbf{R}})$	$(\mathbf{C}, modality, -)$	may
'which'	$(NP_{c1,\mathbf{R}_1}\backslash NP_{c1,\mathbf{R}_1})/(S_{c2}\backslash NP_{c2,\mathbf{R}_2})$	$(c2, \mathbf{R}_2, entity)$	$(c1, \mathbf{R}_1, entity)$
'which'	$(NP_{c1,\mathbf{R}_1}\backslash NP_{c1,\mathbf{R}_1})/(S_{c2}/NP_{c2,\mathbf{R}_2})$	$(c2, \mathbf{R}_2, entity)$	$(c1, \mathbf{R}_1, entity)$

Fig. 4. Examples of lexical items contained in the lexicon. Each bold letter denotes an unbound variable. The same variables within a lexical item must have the same value at the end of the parsing. The first lexical item of "which" is used for the nominative case, and the second one is for the objective case.

category itself is merged with other syntactic categories as the analysis goes further.

The analysis progresses in such a way until the syntactic category $S_{\mathbf{C}}$ (sentence) is finally derived. At that time, the proof of the whole word sequence being a sentence is completed. By the unifications performed during the process of analysis, the semantic roles $\mathbf{R}_2$ (for "cats") and $\mathbf{R}_4$ (for "mice") are determined as *agent* and *patient*, respectively. Then, the address where the semantic representation of each word should be written is determined. The addresses and the semantic representations finally obtained are shown in Fig. 7.

Fig.6 shows a part of the parse tree of the sentence "mice which dogs chase areEatenBy cats" that uses a relative pronoun of the objective case. The analysis of this sentence requires the type raising rule "$> T$". The semantic representation of the relative pronoun "which" is the address of its antecedent (Fig.8).

$$< \dfrac{> \dfrac{\overset{\text{black}}{NP_{\mathbf{C}_1,\mathbf{R}_1}/NP_{\mathbf{C}_1,\mathbf{R}_1}} \quad \overset{\text{cats}}{NP_{\mathbf{C}_2,\mathbf{R}_2}}}{NP_{\mathbf{C}_1,\mathbf{R}_1}} \qquad > \dfrac{\overset{\text{eat}}{(S_{\mathbf{C}_3}\backslash NP_{\mathbf{C}_3,agent})/NP_{\mathbf{C}_3,patient}} \quad \overset{\text{mice}}{NP_{\mathbf{C}_4,\mathbf{R}_4}}}{S_{\mathbf{C}_3}\backslash NP_{\mathbf{C}_3,agent}}}{S_{\mathbf{C}_3}}$$

Fig. 5. A parse tree (proof diagram) of the sentence "black cats eat mice". Although the variables $\mathbf{R}_i$, semantic roles, are unbound at the beginning, their values are determined as $\mathbf{R}_1 = \mathbf{R}_2 = agent$ and $\mathbf{R}_4 = patient$ by unification operations during parsing. A parse tree of the sentence "black cats areEatenBy mice" results in the same form but the semantic roles of *agent* and *patient* are exchanged. The variables $\mathbf{C}_i$ are left unbound after parsing; however, they are restricted to have the same value, i.e. $\mathbf{C}_1 = \mathbf{C}_2 = \mathbf{C}_3 = \mathbf{C}_4$.

$$> \dfrac{\overset{\text{which}}{(NP_{c1,\mathbf{R}_{21}}\backslash NP_{c1,\mathbf{R}_{21}})/(S_{c2}/NP_{c2,\mathbf{R}_{22}})} \qquad > B\;\dfrac{> T\;\dfrac{\overset{\text{dogs}}{NP_{\mathbf{C}_3,\mathbf{R}_3}}}{\mathbf{T}_3/(\mathbf{T}_3\backslash NP_{\mathbf{C}_3,\mathbf{R}_3})} \quad \overset{\text{chase}}{(S_{\mathbf{C}_4}\backslash NP_{\mathbf{C}_4,agent})/NP_{\mathbf{C}_4,patient}}}{S_{\mathbf{C}_4}/NP_{\mathbf{C}_4,patient}}}{NP_{c1,\mathbf{R}_{21}}\backslash NP_{c1,\mathbf{R}_{21}}}$$

Fig. 6. A parse tree of the word sequence "which dogs chase". In this relative clause, whose index is $\mathbf{C}_3 = \mathbf{C}_4 = c2$, the semantic roles of "dogs" and "which" become $\mathbf{R}_3 = agent$ and $\mathbf{R}_{22} = patient$, respectively. The index of the main clause is determined as $c1$, but the semantic role of the antecedent $\mathbf{R}_{21}$, is determined only when the whole sentence has been analyzed.

Address	SR
$(c1, agent, color)$	black
$(c1, agent, entity)$	cats
$(c1, action, -)$	eat
$(c1, patient, entity)$	mice

Fig. 7. The pairs of addresses and semantic representations obtained by analyzing the sentence "black cats eat mice". Because the value of the address **C** is arbitrary, we set it as $c1$.

Address	SR
$(c1, agent, entity)$	cats
$(c1, action, -)$	eat
$(c1, patient, entity)$	mice
$(c2, agent, entity)$	dogs
$(c2, action, -)$	chase
$(c2, patient, entity)$	$(c1, patient, entity)$

Fig. 8. The pairs of addresses and semantic representations obtained by analyzing the sentence "mice which dogs chase areEatenBy cats".

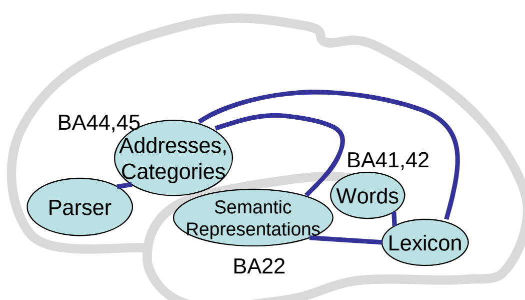

Fig. 9. A possible correspondence between the modules in the model (Fig. 1) and cortical areas.

4 Correspondence to cortical areas

A possible correspondence between the modules in the model (Fig. 1) and cortical areas is shown in Fig. 9. Broca's area (Brodmann Areas 44 and 45) participates in grammar processing and Wernicke's area (BA22; close to the primary auditory area (BA41,42) and to the angular gyrus (BA39)) participates in association between speech sounds and concepts[7]. The parser in the proposed model is the module of grammar processing, thus it is matched with Broca's area. Because it is suggested that agents and patients are represented at the left mid-superior temporal gyrus[8], we suppose that the human's module for semantic representations is located there, around BA22.

By "disabling" a part of the model, utterance that is similar to Broca's aphasia or to Wernicke's aphasia can be reproduced. Although the symptoms of aphasia[7] is complicated and largely vary from patient to patient, we give simple explanation below. Broca's aphasia arises from damage to Broca's area. Its utterance consists of scattering words that do not constitute sentences. The Wernicke's aphasia arises from damage to the Wernicke's area. Its utterance is fluent but does not make sense because of mistakenly selected words.

The proposed model has been implemented in the Prolog language. First, we show an example of a normal behavior of sentence generation. For this example, the semantic representation shown in Fig. 7, which means "black cats eat mice", is given. If the model infers all possible sentences that consist of four words, the following two sentences are obtained as solutions.

```
black cats eat mice
mice areEatenBy black cats
```

In the same condition but without the parser module, the obtained solutions include syntactically incorrect word sequences; however, only those words that are semantically suitable are chosen (Fig.10(a)). This phenomenon is essentially the same one seen in utterance of Broca's aphasia.

If the model infers all possible sentences that consist of four words without giving concrete semantic representation, all syntactically correct sentences will be obtained as solutions(Fig.10(b)). This phenomenon is essentially the same one seen in utterance of Wernicke's aphasia.

In the proposed model, the parser module processes only the *addresses* where meanings are written, but does not process the semantic representations themselves. Because the knowledge of syntax and the knowledge of semantics are clearly separated in this architecture, we can reproduce syntactic disorder and semantic disorder, like actual aphasia.

5 Related work

We aim at the model of the language processing that can be realized easily in the form of Bayesian networks or neural networks like the cerebral cortex. The essential difference from the conventional frameworks of the formal semantics is a complete exclusion of variable-length data structures. Although unification

(a)	(b)
black black black black	white dogs eat dogs
black black black cats	white dogs eat cats
black black black mice	white dogs eat mice
black black black eat	white dogs chase dogs
black black black areEatenBy	white dogs chase cats
black black cats black	white dogs chase mice
black black cats cats	white dogs areEatenBy dogs
black black cats mice	white dogs areEatenBy cats
...	...

Fig. 10. Inferred word sequences similar to (a)Broca's aphasia and (b)Wernicke's aphasia.

operations are occasionally used to express semantic rules, tree structures have been used to express semantic representations. MRS (Minimal Recursion Semantics)[4] expresses meanings not with a tree structure but with a flat structure; however, it needs to handle variable-length data structures.

There are some systems that parse natural language efficiently using loopy belief propagation. For example, the system in [6] is an efficient CCG parser; however, it does not include semantic analysis.

6 Conclusion

We proposed a model of the mechanism of the semantic analysis that does not use variable-length data structure (e.g. lambda terms) but uses a novel form of semantic representation named hierarchical address representation. The modules of the model have correspondence to cortical areas in the brain.

We can reproduce utterance of aphasia by "disabling" some modules in the model.

A chart parser for context-free grammar can be realized as a Bayesian network[9]. It should be possible to apply this method to CCG. Moreover, the mechanism of the unification that the proposed model uses is also easily realizable as Bayesian networks. If the whole model is realized as a Bayesian network, it is possible that lexical items and inference rules can be learned from pairs of word sequences and semantic representations. We believe that this research will connect computational neuroscience and theoretical linguistics, and greatly evolve both of them.

Acknowledgments. We greatly thank to Dr. Daisuke Bekki, who gave useful comments to this research.

References

1. J. Pearl , Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann, 1988.
2. M. Steedman, The Syntactic Process. The MIT Press, 2000.

3. T.S. Lee, D. Mumford, Hierarchical Bayesian inference in the visual cortex. Journal of Optical Society of America, A 20(7): pp.1434–1448, 2003.
4. Ann Copestake, Daniel Flickinger, Carl Pollard, and Ivan A. Sag. Minimal Recursion Semantics. An introduction. Research on Language and Computation, 2005.
5. Yuuji ICHISUGI, The Cerebral Cortex Model that Self-Organizes Conditional Probability Tables and Executes Belief Propagation, In Proc. of IJCNN 2007, pp.1065–1070, 2007.
6. Auli, M. and Lopez, A., A Comparison of Loopy Belief Propagation and Dual Decomposition for Integrated CCG Supertagging and Parsing, In Proc. of ACL, pp.470-480, 2011.
7. Eric R. Kandel et.al ed., Principles of Neural Science, Fifth Edition, McGraw-Hill Companies, 2012.
8. Frankland S. M., Greene J. D., An architecture for encoding sentence meaning in left mid-superior temporal cortex., Proc. Natl. Acad. Sci. U.S.A. 112, 11732-11737, 2015.
9. Naoto Takahashi and Yuuji Ichisugi, Restricted Quasi Bayesian Networks as a Prototyping Tool for Computational Models of Individual Cortical Areas, Proceedings of Machine Learning Research (AMBN 2017), Vol .73, pp.188199, 2017.