

# ベイジアンネットワークを用いた疑似日本語の係り受け解析

## Dependency analysis of pseudo-Japanese using a Bayesian network

高橋 直人

Naoto TAKAHASHI

国立研究開発法人 産業技術総合研究所 人工知能研究センター

Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology (AIST)

Recent researches revealed that certain regions of cerebral cortex serve purely syntactic processing, regardless of the meanings of the words contained in the sentence. In computational neuroscience, on the other hand, a new hypothesis claims that cerebral cortex works as Bayesian networks. If the hypothesis is correct, Bayesian networks should be able to parse grammatical structure. We describe how to perform dependency analysis for pseudo-Japanese using a Bayesian network, then discuss what should be done to evolve the network into a model of human's syntactic processing organ.

### 1. はじめに

言語障害の症例研究や、fMRI を使った実験などから、ヒトの言語能力は脳皮質の一部分（いわゆる言語野）が担っていると考えるとよい。また無意味語を使った最近の実験によると、脳皮質の中には意味の影響を受けず、純粋に文法的な統語解析に特化した部位が存在することがわかってきた [Ohta 13, Pallier 11, Goucha 15].

筆者らは、この統語解析に特化した機能に注目し、そのモデルの作成を目指している。

意味の問題を切り離して統語機能に集中する理由は、次のとおりである。現在までのところ、文あるいは文章の意味をどう表現するかに関して、研究者間で合意が取れているとはいえない。その一方で統語構造に関しては、句構造文法や依存文法など、大部分の研究者によって受け入れられている表現方法が存在する。したがって統語解析機能のモデルを作る場合は、そのモデルの妥当性に関する議論が行いやすいといえる。

本稿では意味を無視した疑似日本語を作成し、その係り受けを解析するシステムが備えているべき性質について考察する。

### 2. 脳皮質モデル BESOM と統語解析

近年の計算論的神経科学において、脳皮質はベイジアンネットワーク的計算を行っている、という仮説が登場した。実際、様々な神経科学的現象がベイジアンネットワークを用いたモデルで再現されている [George 05, Ichisugi 07, Ichisugi 11b, Hosoya 12, Dura-Bernal 12].

またベイジアンネットワーク上での推論には確率伝播アルゴリズム [Pearl 88] がよく用いられるが、このアルゴリズムを近似した場合の変数間の関係が、脳皮質における6層構造間の結合関係と対応付けられることが一杉によって報告された [一杉 08]. 両者の間には単なる偶然とは考えにくい整合性が見られる。これらをふまえて、我々は BESOM [一杉 11a] と呼ぶ階層型ベイジアンネットワーク (図 1) を、脳皮質のモデルとして研究している。

BESOM の研究を通じて得られた知見の一つに、ベイジアンネットワークの条件付き確率表に一種の制約を加えると loopy BP の計算式が単純化でき、計算量が大幅に抑えられるという事が

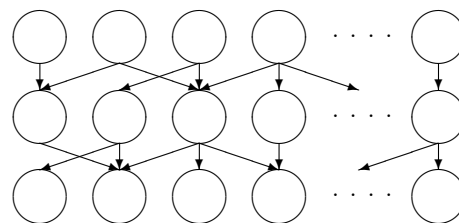


図 1: BESOM の基本構造。階層型ベイジアンネットワークとなり、ノード間を結ぶエッジは隣接した層の間のみ存在する。

ある [一杉 14]. ベイジアンネットワークをナイーブに実装した場合、ノード数に対して指数的に増大する計算量がしばしば問題となるが、上記の知見を用いればこの問題を回避できる可能性がある。

ヒトは文を聞いたときにほぼ実時間で統語解析が可能だが、これは本来ならば必要とされる計算の一部を省略しているためと考えられる。ガーデンパスセンテンスなどの解析に失敗することがあるのは、その省略が原因であろう。もし制約付き BESOM 型ベイジアンネットワークでもほぼ実時間で統語解析が実行でき、しかも類似の解析失敗現象が観察されるのならば、そのベイジアンネットワークは（仮説形成という意味で）ヒトの統語解析機能のモデルになりうる。

### 3. 疑似日本語

本節では今回の実験で採用した疑似日本語について説明する。この疑似日本語は、意味的要素を排除した上で、文節間係り受けの本質を単純化したものとなっている。

まず文節数は、係り受けが交差しうる組み合わせの中で最も小さな値である4に固定する。

次に、体言および用言語幹には、第1文節から順に「甲」「乙」「丙」「丁」を用いる。

助詞は「の」「が」「を」「に」の4種類に限定する。「が」「を」「に」は、それぞれ第1文節内、第2文節内、第3文節内のみに見えることができる。一方「の」は、第1文節から第3文節のどこにも現われることができる。

用言語尾は「する」「して」の2種類を用いる。

以上をまとめたものを表1に示す。また2文節間に係り受

第1文節	甲の, 甲が, 甲する, 甲して
第2文節	乙の, 乙を, 乙する, 乙して
第3文節	丙の, 丙に, 丙する, 丙して
第4文節	丁 (体言止め), 丁する

表 1: 各文節が取りうる形

係り元	係り先	係り受けタイプ
- の	- の/を/に	体言→体言
- が/を/に	- する/して	体言→用言
- する	- の/を/に	用言→体言
- して	- する/して	用言→用言

表 2: 係り受けの成立する文節間関係

けが成立する条件を表 2 のように定める。文全体における係り受け関係の制約としては、

1. 最後以外の各文節は、自分より後方にあるどれか 1 つの文節に係る。
2. 係り受け関係は交差しない。

という通常の 2 種類を考える。

この疑似日本語によって生成される文法的に正しい文あるいは句は、

- (1) 甲が乙を丙に丁する
- (2) 甲が乙して丙が丁する
- (3) 甲の乙の丙の丁

などとなる。

また文法的に正しくない文あるいは句には、

- (4) \*甲の乙の丙の丁する  
(第 3 文節の係り先がない)
- (5) \*甲が乙の丙する丁  
(係り受けが交差している)

などがある。

#### 4. 係り受け解析ネットワーク

本稿で説明する係り受け解析用ベイジアンネットワークの構成を図 2 に示す。BESOM 同様に階層構造を持った構成<sup>\*1</sup>となっており、文節層、係り受け層、非交差確認層の 3 層に分かれている。各々の層に含まれるノードをそれぞれ文節ノード、係り受けノード、非交差確認ノードと呼ぶ。

第  $i$  文節ノード  $B_i$  は、疑似日本語文の第  $i$  文節を値として取る。本稿では  $1 \leq i \leq 4$  である。

係り受けノードは、文節間の係り受け関係を表す。第  $i$  係り受けノード  $K_i$  ( $1 \leq i < 4$ ) は第  $i$  文節ノード  $B_i$  から第 4 文節ノード  $B_4$  までの各ノードに接続されており、第  $i$  文節が第  $j$  文節 ( $i < j \leq 4$ ) に係る時、値  $j$  を取る。第  $i$  文節の係り先が文法的に存在しない場合は、第  $i$  係り受けノード  $K_i$  は値 0 を取る。第 2 係り受けノード  $K_2$  および第 3 係り受けノード  $K_3$  の条件付き確率表を、それぞれ表 3 および表 4 に示す。

\*1 非階層型ベイジアンネットワークで統語解析を試みた例としては、[Pynadath 98] などがある。

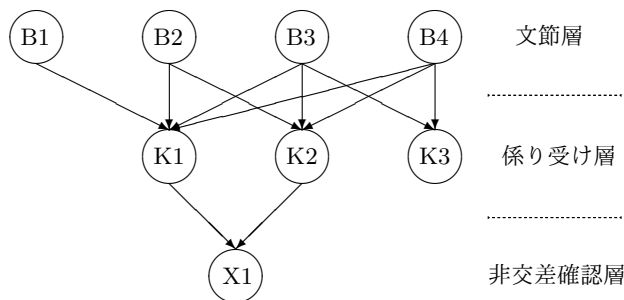


図 2: 係り受け解析を行うベイジアンネットワーク。B1 から B4 は文節ノード、K1 から K3 は係り受けノード、X1 は非交差確認ノードである。BESOM 同様に階層構造をなしており、エッジは隣接した層の間のみ存在する。同一層内のノード間を結ぶエッジは存在しない。

	B3	B4	P(K3=4)	P(K3=0)
丙の, 丙する	丁		1	0
丙の, 丙する	丁する		0	1
丙が, 丙して	丁		0	1
丙が, 丙して	丁する		1	0

表 4: 第 3 係り受けノード  $K_3$  の条件付き確率表

非交差確認ノードは、交差する可能性のある係り受けノードのペアごとに 1 個ずつ設定される。本稿では文節長 4 なので  $X_1$  の 1 ノードだけとなる。両方の係り受けノードで係り受けが成立しており、かつそれらの係り受けが交差していない場合、非交差確認ノードは値 1 を取る。片方あるいは両方の係り受けノードにおいて係り受けが成立していない場合、あるいは両方の係り受けノードで係り受けが成立しているかつそれらの係り受けが交差している場合は、値 0 を取る。

非交差確認ノードの条件付き確率表の一部を表 5 に示す。またその各行が表現する係り受け関係の例を図 3 に示す。

このベイジアンネットワークの文節ノードに値を与え、さらに係り受けが交差しないという制約を与える（非交差確認ノードを 1 に設定する）と、文法的に正しい係り受け関係を選択することができる（図 4）。また文節のみを与えて文全体の文法性を判断させたり（図 5）、係り受け関係と文節の一部のみを与えて、欠けている文節を補う（図 6）などの推論も可能である。

#### 5. 今後の課題

これまでのところは、階層型ベイジアンネットワークで係り受け解析が実行できることを、ナイーブな形で示したに過ぎない。もしヒトの統語解析機能のモデルとなるようなベイジアンネットワークを目差すのならば、そのベイジアンネットワークは以下のような性質を備えている必要がある。

- 統語解析に要する時間が、文の長さに（ほぼ）比例する。

	K1	K2	K3	P(X1=1)	P(X1=0)
a	2	3	4	1	0
b	3	3	4	1	0
c	3	4	4	0	1
d	4	0	4	0	1

表 5: 非交差確認ノードの条件付き確率表（一部）

B2	B3	B4	P(K2=3)	P(K2=4)	P(K2=0)
乙の, 乙する	丙の, 丙が	丁	0.5	0.5	0
乙の, 乙する	丙の, 丙が	丁する	1	0	0
乙の, 乙する	丙する, 丙して	丁	0	1	0
乙の, 乙する	丙する, 丙して	丁する	0	0	1
乙が, 乙して	丙の, 丙が	丁	0	0	1
乙が, 乙して	丙の, 丙が	丁する	0	1	0
乙が, 乙して	丙する, 丙して	丁	1	0	0
乙が, 乙して	丙する, 丙して	丁する	0.5	0.5	0

表 3: 第 2 係り受けノード K2 の条件付き確率表

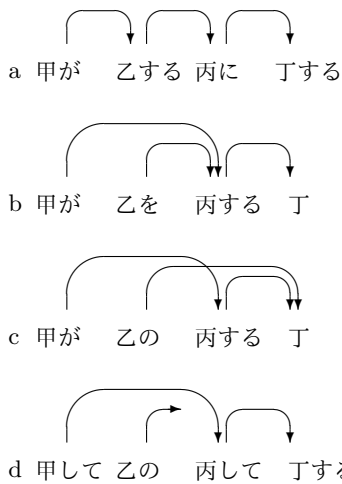


図 3: 表 5 の各行に対応する係り受け関係の例. a と b は正しい係り受け関係なので非交差確認ノードが 1 となり, c は係り受けが交差するために, また d は第 2 文節の係り先がないために非交差確認ノードが 0 となる.

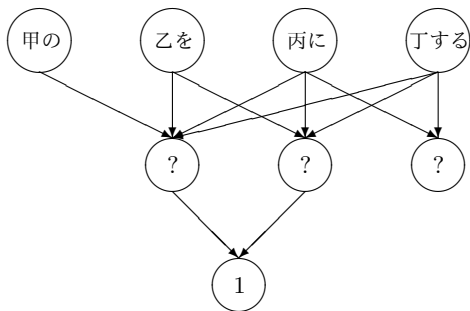


図 4: 文法的に正しい係り受け関係の選択

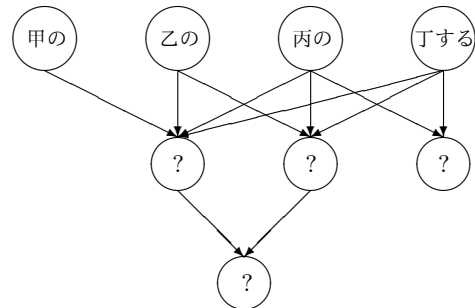


図 5: 文節列から文法性を性判断する

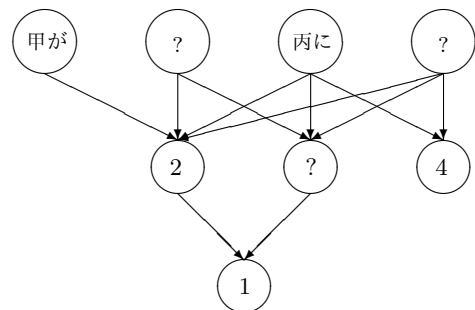


図 6: 欠けている文節を補う

これには条件付き確率表に [Ichisugi 15] で示された制約を課するなどの方法が有望であろう。

- 深さ優先方式,あるいはビーム幅の非常に狭いビームサーチに似た方式の解析を行う。ヒトは文を読んでいて解析に失敗すると,前に戻って解析をやり直す。チャートパーザのようにすべての可能性を溜め込んではいない。(少なくとも意識に上る形では溜め込んでいない。)
- 構文木の形に好みがある。特に3重以上の中央埋め込み文の解析には困難を示す。
- 漸進的な解析が可能である。今回の実験では4文節がすべて揃っている状態で係り受け解析を行っているが,ヒトは1文を読み終わってから(あるいは聞き終わってから)解析を開始するのではなく,単語が入力されるにつれて漸進的に解析を進めているように思える。特に話し言葉の場合は,言いよどみや言い直しなどにも対応できる非常に柔軟な能力を示す。

これらの性質を持ったベイジアンネットワークが,条件付き確率表に制約を与えるだけで実現できるかは不明である。

今後は,上記の性質を満たすような統語解析用ベイジアンネットワークに関して,ネットワークの構造や文法の学習も含めて研究を進める予定である。

## 謝辞

この成果は,国立研究開発法人新エネルギー・産業技術総合開発機構(NEDO)の委託業務の結果得られたものである。

## 参考文献

- [Dura-Bernal 12] Dura-Bernal, S., Wennekers, T., and Denham, S. L.: Top-Down Feedback in an HMAX-Like Cortical Model of Object Perception Based on Hierarchical Bayesian Networks and Belief Propagation, *PLoS ONE*, Vol. 7, No. 11 (2012)
- [George 05] George, D. and Hawkins, J.: A hierarchical Bayesian model of invariant pattern recognition in the visual cortex, in *2005 International Joint Conference on Neural Networks (IJCNN)* (2005)
- [Goucha 15] Goucha, T. and Friederici, A. D.: The language skeleton after dessecting meaning: A functional segregation within Broca's Area, *NeuroImage*, Vol. 114, pp. 294–302 (2015)
- [Hosoya 12] Hosoya, H.: Multinomial Bayesian Learning for Modeling Classical and Nonclassical Receptive field Properties, *Neural Computation*, Vol. 24, No. 8 (2012)
- [Ichisugi 07] Ichisugi, Y.: The cerebral cortex model that self-organizes conditional probability tables and executes belief propagation, in *2007 International Joint Conference on Neural Networks (IJCNN)* (2007)
- [一杉 08] 一杉 裕志: 脳の情報処理原理の解明状況, Technical Report AIST07-J00012, 産業技術総合研究所テクニカルレポート (2008)
- [一杉 11a] 一杉 裕志: 大脳皮質のアルゴリズム BESOM Ver.2.0, Technical Report AIST11-J00009, 産業技術総合研究所テクニカルレポート (2011)
- [Ichisugi 11b] Ichisugi, Y.: Recognition Model of Cerebral Cortex based on Approximate Belief Revision Algorithm, in *2011 International Joint Conference on Neural Networks (IJCNN)* (2011)
- [一杉 14] 一杉 裕志, 高橋 直人: 制限付きベイジアンネットワークにおける認識アルゴリズム OOBP, 第28回人工知能学会全国大会 (2014)
- [Ichisugi 15] Ichisugi, Y. and Takahashi, N.: An Efficient Recognition Algorithm for Restricted Bayesian Networks, in *2015 International Joint Conference on Neural Networks (IJCNN)* (2015)
- [Ohta 13] Ohta, S., Fukui, N., and Sakai, K. L.: Syntactic Computation in the Human Brain: The Degree of Merger as a Key Factor, *PLoS ONE*, Vol. 8, No. 2 (2013), e56230
- [Pallier 11] Pallier, C., Devauchelle, A.-D., and Dehaene, S.: Cortical representation of the constituent structure of sentences, *PNAS*, Vol. 108, No. 6, pp. 2522–2527 (2011)
- [Pearl 88] Pearl, J.: *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann (1988)
- [Pynadath 98] Pynadath, D. V. and Wellman, M. P.: Generalized Queries on Probabilistic Context-Free Grammars, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 10, No. 1, pp. 65–77 (1998)