

# 木置換文法による語彙と文法の教師なし同時学習に関する検討

相良陸成<sup>1</sup> 田口亮<sup>1</sup><sup>1</sup>名古屋工業大学大学院 工学研究科

r.sagara.628@nitech.jp taguchi.ryo@nitech.ac.jp

## 概要

人の言語学習能力をロボットにより構築する研究が行われており、単語とその意味、文法の教師なし学習モデルが提案されている。一方、木構造の文法と語彙を同時に学習するモデルは提案されていない。本研究では木置換文法により音節列から語彙と文法を教師なしで同時学習する手法について検討する。木置換文法において構文木を構成する単位である基本木として、単語と構文を同じ形式で学習する。小規模なデータセットを用いた実験により、基本木として単語と構文を抽出可能であることが示された。このような学習は、語彙と文法を連続体として捉える認知言語学との親和性があるものと考えられる。

## 1 はじめに

人間の幼児は大人とのインタラクションを通じて、単語とその意味、および文法を学習する能力を持つ。この能力の数理モデル化を目的の一つとして、ロボットによる語彙・概念の獲得に関する研究[1]–[3]や、ロボットによる文法学習の研究[4], [5]が行われている。一方、語彙と文法を統一的なモデルで学習する手法は提案されていない。人間は学習した「赤い」「赤」「緑」という単語から「緑い」という単語を作り出せるように、学習した単語から単語を生成する規則を学習することが可能である。このような学習をロボットにより実現するには、語彙と文法を統一的なモデルで学習する必要があると考える。このため本研究では、木置換文法を用いることにより、単語と構文を統一的な形式で教師なし学習する方法について検討する。

## 2 木置換文法

木置換文法 (TSG: Tree Substitution Grammar) は、文脈自由文法の生成規則の右辺を木構造へと一般化し、生成規則に確率を対応させた形式文法である。この生成規則の右辺の木は基本木 (elementary tree)

と呼ばれる。図 1 に示すように、基本木の末端にある NP などの非終端記号を基本木で置換することにより、構文木を導出する。ここでは Cohn らの提案した木置換文法[8]について述べる。この木置換文法により、品詞列から教師なしで依存構造解析ができることが示されている。

### 2.1 確率モデル

木置換文法において、非終端記号列  $y$  は以下の式により生成される。

$$p(y) = \sum_{t: \text{yield}(t)=y} p(t)$$

$$p(t) = \sum_{e: \text{tree}(e)=t} p(e)$$

$$p(e) = \prod_{c \rightarrow e \in \mathcal{L}} p(e|c, \mathcal{L})$$

ただし、 $t$  は構文木、 $e$  は基本木の列、 $\mathcal{L}$  は木置換文法のパラメータ集合、 $c \rightarrow e \in \mathcal{L}$  は基本木の列  $e$  に含まれる基本木  $e$  が非終端記号  $c$  から生成される規則を表す。また、 $\text{yield}(t)$  は構文木  $t$  の終端記号列を表し、 $\text{tree}(e)$  は基本木の列  $e$  により作られる構文木を表す。非終端記号  $c$  から生成される基本木  $e$  の分布  $p(e|c, \mathcal{L})$  は基本木の出現頻度に基づくが、一度も出現していない基本木にどのように確率を割り当てるかが問題となる。Cohn らは階層 Pitman-Yor 過程に

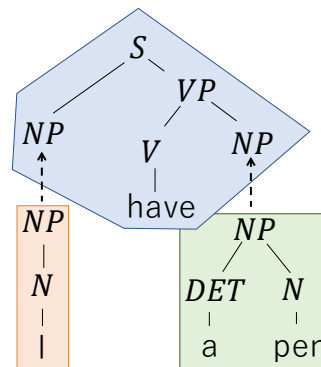


図 1 木置換文法

より、一度も出現していない基本木の確率を、文脈自由文法の生成規則の出現頻度に基づいて割り当てている。これによりモデル全体を確率的生成モデルで表現可能とした。詳細は文献[8]を参照されたい。

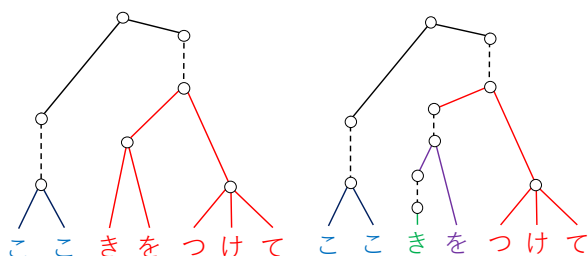
## 2.2 パラメータ推定

確率モデルのパラメータ推定に一般に用いられる Metropolis-Hastings 法を用いる。基本木からなる構文木のサンプリングには、inside アルゴリズムによる動的計画法の結果を用いる。サンプリングされた基本木からなる構文木は、Metropolis-Hastings 法のプロセスの中で、尤度に基づいて確率的に受理または棄却される。

## 3 木置換文法による語彙・文法学習

本研究では Cohn らの木置換文法を用いて、音節列から語彙と文法を学習することを試みる。図 2 に模式図を示す。音節列に対して木置換文法を適用し、得られた基本木を単語および構文と解釈する。図では実線で繋がれた部分木を基本木として表しており、点線は基本木の置換(接続)部を表している。図 2(a) は「きをつけて」を一つの単語とした場合であり、図 2(b) は「き/を/つけて」のように 3 単語とした場合である。このような分割候補に対して生成確率を取得できる。このため、ロボットのセンサ情報から得られた意味情報を加味して分割候補を選択するなどの拡張が考えられる。また、n-gram とは異なり、意味を対応付ける単位として構文を抽出可能である。

図 2 の学習を木置換文法で実現するために、図 2 の葉ノードは音節に対して一意に決まる前終端記号とし、そこから音節の終端記号が導出されるとする。学習に用いる非終端記号の性質は事前に与えられず、非終端記号の数のみが与えられる。



(a) 1 単語とする場合 (b) 単語に分ける場合

図 2 木置換文法による単語と文法の学習

## 4 実験

木置換文法により単語および構文の教師なし学習が可能であるかを検証する。

### 4.1 実験条件

データセット A, データセット B のそれぞれを用いて実験した。データセット A は 45 の音節列からなり、3 種類の色、3 種類の物体、5 種類の動作を表す単語のすべての組み合わせからなる。すべての音節列は「色・物体・動作」の順に並んでおり、例として「あかいはこのせて」のような音節列が存在する。データセット B はロボットに位置を教示する 104 の音節列からなる。4 種類の絶対位置(例: リビング)、6 種類の相対位置(例: 後ろ)、4 種類の物体(例: テレビ)の単語が含まれる。「ここわ X です」のような 19 種類の発話パターン X に、「(物体)の(相対位置)」または「(絶対位置)」をあてはめ、「ここわてれびのまえです」のような音節列を生成した。データセット A とは異なり、単語の出現頻度および 2 単語の共起頻度には偏りがある。付録に全データを記載する。

非終端記号の個数を 3, 4, 5 の 3 つの値に設定してそれぞれ実験した。構文木の根ノードには特別な記号を用いず、全ての非終端記号が等確率で出現可能とした。Metropolis-Hastings 法のサンプリング回数は 1000 回とした。最終イテレーションにおいて得られた木置換文法のパラメータを用いて、各入力音節列から事後確率最大となる基本木の列を取得し、評価に用いた。データセットと非終端記号の組み合わせごとに、初期値を変えて 10 回ずつ実験した。

### 4.2 実験結果と考察

#### 4.2.1 データセット A

非終端記号を 3 個としたときの学習例を図 3 に示す。図の PT は前終端記号を、NT は非終端記号を表す。点線部は基本木の境界を表している。点線で繋がれた 2 つのノードは構文木としては同一のノードであるが、実線で繋がれた部分木を基本木とみなせるよう、視認性のために分割して描画している。アスタリスク(\*)は、基本木の境界ノードではないことを表す。この学習例において、基本木は「あかい」「はこ」「のせて」の 3 つの単語の基本木に正しく分割されていることが分かる。また、単語は音節列だけでなく、他の単語との接続に関する情報を持つ

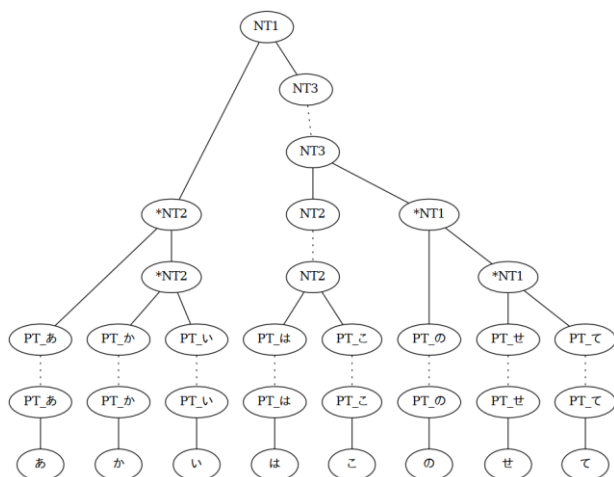


図 3 データセット A の学習例  
(非終端記号 3 個)

た、構文として学習された。この実験では 45 の音節列すべてが正しく 3 つの単語に分割された。また、この実験結果ではすべての物体の基本木の根ノードは NT2、すべての動作の基本木の根ノードは NT3 であり、構文木上の 3 単語の関係は全て同じであった。これは、各単語の品詞とそれらの関係、すなわち文法が学習されたと考えることができる。なお、構文木の構造については誤って学習されたが、これは全ての音節列「色・物体・動作」という語順で出現し、「色・物体」のみの音節列がなかったためであると考えられる。音節列のバリエーションを増加させることにより、構文木としても正しい結果が得られるようになると思われる。

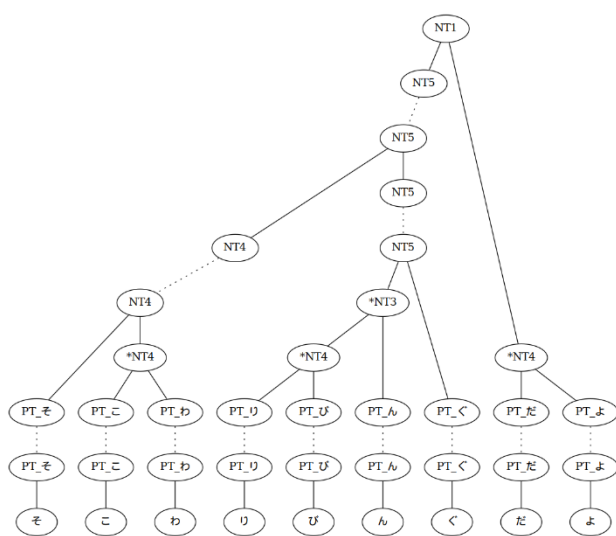
次に、定量評価について示す。データに含まれる全単語を対象として、それらが基本木として抽出された割合を算出した。非終端記号が 3, 4, 5 個の時に抽出された割合は、10 回の実験の平均でそれぞれ 99.7%, 99.3%, 96.1% となった。いずれも高い精度で音節列から単語を抽出できていることがわかる。非終端記号を増やすことで基本木の内部ノードとして活用されることを期待したが、非終端記号が少ない方が精度は高かった。基本木の置換部の非終端記号の方が重要であり、内部ノードがどの非終端記号になるかは学習に大きく影響を及ぼしていないと考えられる。

## 4.2.2 データセット B

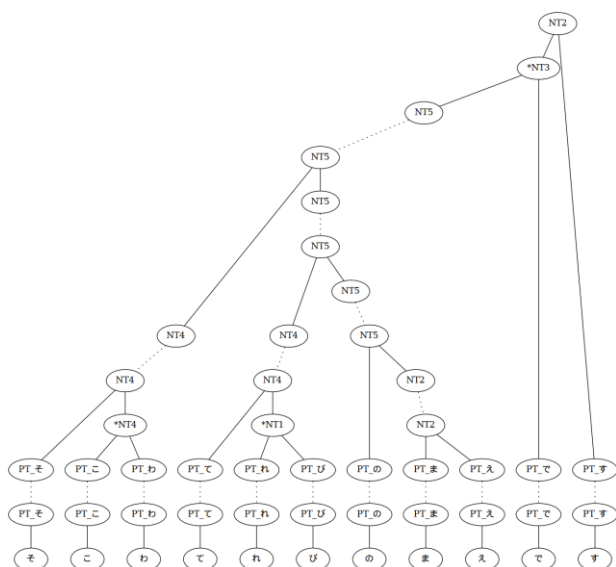
非終端記号を 5 個としたときの学習例を図 4(a)および(b)に示す。音節列が基本木としておおよそ正しく単語に分割できていることがわかる。また、位置を表す「リビング」と「てれびのまえ」は単語数が異なるが、それらを統べる部分木の根ノードはどちらも非終端記号 NT5 であった。これは、NT5 が「位置」という抽象的な単位として学習されたと考えることができる。これにより、これまでの n-gram では学習できなかった、単語数に依存しない AXB 形式の構文を学習可能となり、それに従う単語列を生成可能となることが期待される。ただし、この学習結果では NT5 は他の基本木の置換部にも用いられており、現段階では単語列の生成は正しくできないと考えられる。5 つある非終端記号がすべて置換部に用いられることを期待したが、同じ非終端記号を複数回用いる傾向が見られた。基本木が根ノードを伴って学習されるために根ノードを変更することが難しく、局所解に留まっている可能性がある。この点の改善は今後の課題である。

また、データセット A と同様に単語が抽出された割合を算出した。評価対象は、絶対位置の単語、物体の単語、「まえ」以外の相対位置の単語とした。非終端記号が 3, 4, 5 個の時、想定した単語に分割された割合は、10 回の実験の平均でそれぞれ 59.4%, 52.9%, 46.6% となった。ただしこのデータセットにおいては、この値が高いほど望ましい学習結果であるとは限らない。想定した単語に分割されなかった例として、「のひだり」や「いすのうしろ」を基本木とする場合が見られた（付録を参照）。「ひだり」の前の音節は必ず「の」であるため、「ひだり」と「のひだり」の出現回数は等しい。また、「うしろ」は「いすのうしろ」の形でしか出現せず、「いすのうしろ」の出現回数は「リビング」と同じ 8 回である。このように高頻度に出現する音節列を、「気を付ける」と同様に一つの単語と解釈し、それを意味と対応付ける単位として学習することは一つの妥当な結果であると言える。このような学習は、語彙と文法を連続体として捉える認知言語学[6]との親和性があるものと考えられる。

<sup>i</sup> 「まえ」は、「てまえ」と「ここのなまえわ」の中でも部分的に出現する音節列であり出現数では評価できないため、評価対象外とした。



(a) 非終端記号 5 個「そこわりびんぐだよ」



(b) 非終端記号 5 個「そこわてれびのまえです」

図 4 データセット B の学習例

## 5 制約

Cohn らの木置換文法を音節列からの語彙と文法の教師なし学習に利用できることを示したが、いくつかの制約が存在する。まず、Cohn らの木置換文法を教師なし学習するためには二分木とする必要があり、過度に内部ノードを必要とする点が挙げられる。これはパラメータ推定に用いる inside アルゴリズムが二分木を前提としているためである。生成規則が「ここわ・X・です」「て・れ・び」のような3つの子を持つことができれば、内部ノードを削減でき、分析が容易になると考える。また、単語学習時に基

本木の根ノードの非終端記号を定める必要がある点は、学習を難しくするだけでなく、認知言語学の観点からも不自然な仮定であると考えられる。このような技術的制約を取り払い、より認知言語学からの説明が可能なモデルに改良したい。さらに、データセットが小さいにも関わらず計算コストが大きい点が挙げられる。データセット A を用いた非終端記号 3 個での実験では、Intel Core i7-7820X CPU 3.60GHz のシングルコアでの実行で 3 時間 11 分を要した。また、非終端記号の数を増加させると実行時間は指数的に増大し、非終端記号が 8 個の場合は 54 時間 46 分を要した。このため、大規模なデータセットへの適用は困難である。RNN や Transformer を用いた統語構造の学習モデル[8], [9]を用いるなど、高速な学習を可能とする方法について検討したい。

## 6 おわりに

本研究では、木置換文法により音節列から語彙と文法を教師なし学習する手法について検討した。小規模なデータセットを用いた実験で、木置換文法により基本木という単位で単語と構文を抽出できることを確認した。

本研究では言葉の意味（概念）を扱っていない。文献[2], [3]では、語彙学習とセンサ情報からの概念学習を統合することで、語彙学習の精度が向上することが示されている。また認知言語学においては、構文は音韻と意味のペアとして抽象化されていくと考える。このため意味の抽象化についても同時に扱い、意味を伴って文法を学習可能なモデルに発展させたい。

本研究では発話音声から音節列が得られたものと仮定し、音節列を観測値として学習している。完全な教師なし学習の実現のためには、発話音声からの音素・音節の獲得との統合が必要である。谷口らは、発話音声から音素と単語を教師なし学習する手法を提案している[10]。この手法で学習される文法は 2-gram であり、AXB のような文法は学習できない。この手法の音素の獲得を取り入れることなどにより、完全な教師なし学習として発話音声からの語彙・文法学習を実現したい。

## 謝辞

本研究は、JST 次世代研究者挑戦的研究プログラム JPMJSP2112 の助成を受けたものです。

## 参考文献

- [1] T. Nakamura, T. Nagai, K. Funakoshi, S. Nagasaka, T. Taniguchi, and N. Iwahashi, “Mutual learning of an object concept and language model based on MLDA and NPYLM,” in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2014, pp. 600–607.
- [2] A. Taniguchi, T. Taniguchi, and T. Inamura, “Unsupervised spatial lexical acquisition by updating a language model with place clues,” *Rob. Auton. Syst.*, vol. 99, pp. 166–180, Jan. 2018.
- [3] R. Sagara *et al.*, “Unsupervised lexical acquisition of relative spatial concepts using spoken user utterances,” *Adv. Robot.*, vol. 36, no. 1–2, pp. 54–70, Jan. 2022.
- [4] A. Aly, T. Taniguchi, and D. Mochihashi, “A Probabilistic Approach to Unsupervised Induction of Combinatory Categorical Grammar in Situated Human-Robot Interaction,” in *2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids)*, 2018, pp. 1–9.
- [5] M. Spranger, “A basic emergent grammar for space,” *Experiments in Cultural Language Evolution*, ed. L. Steels (Amsterdam: John Benjamins), pp. 207–232, 2011.
- [6] T. Cohn, P. Blunsom, and S. Goldwater, “Inducing tree-substitution grammars,” *J. Mach. Learn. Res.*, vol. 11, pp. 3053–3096, 2010.
- [7] R. W. Langacker, *Foundations of Cognitive Grammar: Volume I: Theoretical Prerequisites*. Stanford University Press, 1987.
- [8] C. Dyer, A. Kuncoro, M. Ballesteros, and N. A. Smith, “Recurrent Neural Network Grammars,” in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, San Diego, California, 2016, pp. 199–209.
- [9] P. Qian, T. Naseem, R. Levy, and R. F. Astudillo, “Structural Guidance for Transformer Language Models,” in *Proceedings of the 59th Annual Meeting of the Association— 452 for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume1: Long Papers)*, Online, 2021, pp. 3735–3745.
- [10] T. Taniguchi, R. Nakashima, H. Liu, and S. Nagasaka, “Double articulation analyzer with deep sparse autoencoder for unsupervised word discovery from speech signals,” *Adv. Robot.*, vol. 30, no. 11–12, pp. 770–783, Jun. 2016.



