MTRNNを用いた階層的言語構造の創発

○日下航† 有江浩明‡ 谷淳‡ 尾形哲也† 高橋徹† 奥乃博†
(†京都大学大学院情報学研究科 ‡理化学研究所 脳科学総合研究センター)

Emergence of Linguistic Hierarchy in a MTRNN

*Wataru Hinoshita†, Hiroaki Arie‡, Jun Tani‡,
Tetsuya Ogata†,
Toru Takahashi†, and Hiroshi G. Okuno†
(†Graduate School of Informatics, Kyoto Univ. ‡Brain Science Institute, RIKEN.)

Abstract— We show that a Multiple Timescale Recurrent Neural Network (MTRNN) can acquire the lingistic hierarchical structure: characters → words → setences. In our experiment, we trained our model to learn language using only a sentence set without any previous knowledge about words or grammar. Our experimental results shown that the model can acquire the capabilities to recognize and deterministically generate grammatical sentences even if they were not learned. The analysis of neural activations in our model revealed that the MTRNN had self-organized mirroring the hierarchical linguistic structure taking advantage of differences in time scale among its neurons: concretely, neurons that change the fastest represented "characters," those that change more slowly represented "words," and those that change the slowest represented "sentences."

Key Words: Language acquisition, Multiple Timescale Recurrent Neural Network, Linguistic hierarchy, Self-organization

1. はじめに

人間の言語獲得機構を明らかにするため, またロボットに言語を自在に扱わせるため，計算機による言語獲得の研究が盛んに行われてきた [1, 2, 3]. 言語獲得研究では，子供がどのようにして限定的で質の悪い言語刺激のみから，複雑で多様な文を自在に生成する能力を獲得するのかが大きな争点となってきた [4, 5]. 近年, 非線形力学系が複雑で多様なパターンの生成能力を持つことが明らかになるとともに，言語を力学系として捉え，Recurrent Neural Network (RNN) などの神経力学モデルによって言語を獲得させる研究が大きな注目を集めるようになってきた [1, 6, 7].

Elman [7] らは，RNN が文集合のみから文法構造を自己組織的に獲得し，文脈に則して次の単語候補を正しく予測できるようになることを示したが, このモデルには決定論的に特定の文を生み出す生成能力を持たないという問題点がある. 文の生成能力は, 人間の言語能力の中核の一つである. 杉田ら [3] や尾形ら [8] は，RNN にパラメータノードを付加した RNN with Parametric Bias (RNNPB) [9] を用いることで，パラメータに応じた文が決定論的に生成されることを示している. これはある発話意図（パラメータ）に従って，適切な文を生成する過程のモデル化と見ることができる. しかし, RNNPB は長いシーケンスの学習に不向きであり，これらの研究では 2〜3 単語程度の極めて単純な文しか扱われていなかった.

RNN を用いた従来の言語獲得研究全般の問題として，各入出力ノードが単語に対応しているため, 既知の語彙からなる文しか学習できないという点が挙げられる. 語彙の事前知識なしに文を学習するには, 文字から単語, 単語から文といった階層的な合成能力が必要になる. このような階層的合成能力は言語表現の多様性の中核を担っており, 文の生成能力と共に言語の創造的側面を実現する上で必要不可欠な能力であると言える.

本研究の目的は, 階層的言語構造 (文字 → 単語 → 文) を, 文集合のみを用いて, 神経回路網モデルに自己組織化させ, 未知文の認識・決定論的生成を行わせることである. 我々は，時定数の異なるニューロン群から構成される Multiple Timescale RNN (MTRNN) [10] を用いて学習を行い, この階層性の自己組織化を目指す. 学習済み MTRNN を認識・生成器として用いることで, 獲得された構造を基に, 未知の文が正しく生成されるかを検証する.

また, 本研究では学習時に誤った文を与えることで, 獲得される言語能力にどのような影響が生じるのかも同時に調べる. 一般的に，子供が実際に得る言語刺激の質は必ずしも高くなく，それが言語獲得の障壁になっていると考えられている [4, 5]. それに対し我々は，学習データの誤りが逆にロバストな言語能力の獲得を促すという仮説を立て, その検証を行う.

2. 言語学習モデル

我々は言語学習に力学モデルである MTRNN[10] を用いる. MTRNN は, 現在の状態を入力とし次状態の予測値を出力することで，時系列データを扱うことができる. このモデルは，異なった時定数を持つ複数のニューロングループから構成され，時定数が大きいほどニューロンの状態は緩やかに変化する. この時定数の違いによって，情報の階層的構造化が生じる. MTRNN は一部のニューロンの初期値を変えることで，複数の

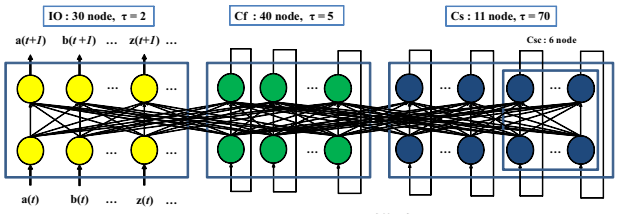

Fig.1 MTRNN 構成

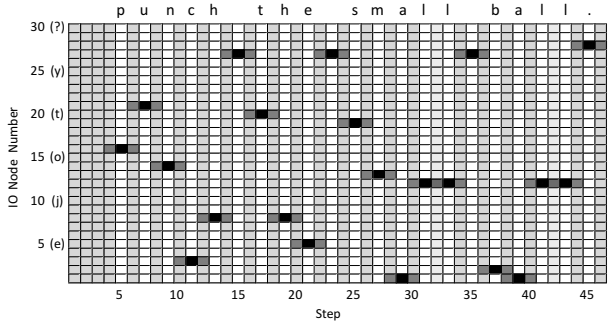

Fig.2 入力データ例: “punch the small ball.”

時系列パターンを決定論的計算によって生成できる．また，時系列パターンが与えられた時に，そのパターンを生成する初期値を計算によって求めることもできる．そのため，このモデルは時系列パターンの認識器および生成器として用いることができる．初期値の状態空間は，学習データ間の力学的相関から自己組織的に獲得される．このため，学習データの汎化によって，未知のパターンであっても認識・生成することができる．

我々が実験に用いた言語学習用 MTRNN は A から Z の各アルファベットと, スペース, カンマ, ピリオド, クエスチョンの 4 つの記号に対応する計 30 の入出力ノード (IO) を持ち, 文脈ノードには, Fast Context(Cf, 40 ノード) と Slow Context(Cs, 11 ノード) と呼ばれる 2 種類が存在する (Fig. 1)．IO, Cf, Cs の順に時定数 (τ) は大きくなり, ニューロン状態の変化が緩やかになる．また Cs のうち 6 ノードを Controlling Slow Context(Csc) とした．この Csc に与える初期値によって，モデルから生成される時系列パターンは一意に決定される．

このモデルでは, 文章は文字に対応した IO ノードが順次発火することで表現される (Fig. 2)．モデルはそれまでの入力から次状態での IO ノードの発火を予測するように訓練される．このため，学習はアノテーションデータなどのない文集合のみを用いて行われる．

時刻 t における i 番目のニューロンの発火値 ($y_{t,i}$) は以下のように求める．

$$y_{t,i} = \begin{cases} \dfrac{\exp(u_{t,i}+b_i)}{\sum_{j\in I_{IO}} \exp(u_{t,j}+b_j)} & \cdots(i\in I_{IO}) \\ \dfrac{1}{1+\exp(-(u_{t,i}+b_i))} & \cdots(i\notin I_{IO}) \end{cases} \quad (1)$$

$$u_{t,i} = \begin{cases} 0 & \cdots(t=0\wedge i\notin I_{Csc}) \\ Csc_{0,i} & \cdots(t=0\wedge i\in I_{Csc}) \\ \left(1-\dfrac{1}{\tau_i}\right)u_{t-1,i} + \dfrac{1}{\tau_i}\left[\sum_{j\in I_{all}} w_{ij}x_{t,j}\right] & \cdots(\mathrm{o/w}) \end{cases} \quad (2)$$

$$x_{t,j} = y_{t-1,j} \quad \cdots(t\geq 1) \quad (3)$$

$I_{IO}, I_{Cf}, I_{Cs}, I_{Csc}$: 各グループの要素番号集合 $(I_{Csc}\subset I_{Cs})$

I_{all} : $I_{IO}\cup I_{Cf}\cup I_{Cs}$

$u_{t,i}$: 時刻 t における i 番目のニューロンの内部状態

b_i : i 番目のニューロンのバイアス

$Csc_{0,i}$: MTRNN を制御する初期値

τ_i : i 番目のニューロンの時定数

w_{ij} : j 番目から i 番目のニューロンへの結合重み

$w_{ij}=0 \cdots(i\in I_{IO}\wedge j\in I_{Cs})\vee(i\in I_{Cs}\wedge j\in I_{IO})$

$x_{t,j}$: 時刻 t における j 番目のニューロンからの入力

MTRNN の学習は Back Propagation Through Time (BPTT) [11] によって行われる．すなわち，以下の式に従って結合重み (w_{ij}), バイアス (b_i), 初期値 ($Ccs_{0,i}$) を逐次的に更新する．

$$\begin{aligned} w_{ij}^{(n+1)} &= w_{ij}^{(n)} - \eta\frac{\partial E}{\partial w_{ij}} \\ &= w_{ij}^{(n)} - \frac{\eta}{\tau_i}\sum_t x_{t,j}\frac{\partial E}{\partial u_{t,i}} \end{aligned} \quad (4)$$

$$\begin{aligned} b_i^{(n+1)} &= b_i^{(n)} - \beta\frac{\partial E}{\partial b_i} \\ &= b_i^{(n)} - \beta\sum_t \frac{\partial E}{\partial u_{t,i}} \end{aligned} \quad (5)$$

$$\begin{aligned} Ccs_{0,i}^{(n+1)} &= Ccs_{0,i}^{(n)} - \alpha\frac{\partial E}{\partial Ccs_{0,i}} \\ &= Ccs_{0,i}^{(n)} - \alpha\frac{\partial E}{\partial u_{0,i}} \cdots(i\in I_{Csc}) \end{aligned} \quad (6)$$

$$E = \sum_t\sum_{i\in I_{IO}} y_{t,i}^* \cdot \log\left(\frac{y_{t,i}^*}{y_{t,i}}\right) \quad (7)$$

$$\frac{\partial E}{\partial u_{t,i}} = \begin{cases} y_{t,i} - y_{t,i}^* + (1-\dfrac{1}{\tau_i})\dfrac{\partial E}{\partial u_{t+1,i}} \cdots(i\in I_{IO}) \\ y_{t,i}(1-y_{t,i})\displaystyle\sum_{k\in I_{all}} \frac{w_{ki}}{\tau_k}\frac{\partial E}{\partial u_{t+1,k}} \\ \qquad +(1-\dfrac{1}{\tau_i})\dfrac{\partial E}{\partial u_{t+1,i}} \cdots(\mathrm{o/w}) \end{cases} \quad (8)$$

n : 更新回数

E : 予測誤差

$y_{t,i}^*$: 時刻 t における i 番目のニューロンの対象文に対する理想発火値

η, β, α : 学習係数

BPTT アルゴリズムを用いる際は，IO ノードの入力値 ($x_{t,j}$) には，以下の式 (9) に従って学習データからのフィードバックを得たものが用いられる．

$$x_{t,j} = (1-r)\times y_{t-1,j} + r\times y_{t-1,j}^* \cdots(t\geq 1\wedge j\in I_{IO}) \quad (9)$$

r : フィードバック係数 $(0\leq r\leq 1)$

Table 1 Lexicon

Nonterminal symbol	Words
V_I (intransitive verb)	jump, run, walk
V_T (transitive verb)	kick, punch, touch
N (Noun)	ball, box
ART (article)	a, the
ADV (adverb)	quickly, slowly
ADJ_S (adjective:size)	big, small
ADJ_C (adjective:color)	blue, red, yellow

Csc の初期値は MTRNN の振る舞いを決定付ける．この Csc の初期値を各次元の要素とするベクタを $\boldsymbol{Csc_0}$ と定義する．ネットワーク重み (結合重み w_{ij} とバイアス b_i) は学習データ全体で共有されるが，$\boldsymbol{Csc_0}$ は各学習データ毎に独立に用意される．$\boldsymbol{Csc_0}$ とネットワーク重みを同時に更新していくことで，初期値空間は学習データ間の相関に従って自己組織的に獲得される．

時系列データを認識する際は，ネットワーク重みを固定した状態で BPTT を行うことで, 対象データを生成する $\boldsymbol{Csc_0}$ を求める (式 (6))．認識フェーズにおいて，IO 入力値の計算には式 (9) が用いられるが，認識対象データの一部が未知である場合は，その部分に対して式 (3) を用いる．このため，MTRNN は対象データの情報が部分的にしか与えられていない場合においても認識を行うことができる．

時系列データの生成は，$\boldsymbol{Csc_0}$ をセットしてフォワード計算 (式 (1), (2), (3)) を再帰的に行うことで実現される．

3. 言語学習手順

言語の学習は以下の手順で行った．

1. 1 つの正規文法から 100 種類の文を生成する．
2. そのうち 80 文を MTRNN に学習させる．
3. 未学習の 20 文を含む 100 文を用いて, MTRNN の能力を評価する．評価は以下の手順で行う．
   (1) 認識： 文を入力し, 重み固定 BPTT により, $\boldsymbol{Cs_0}$ を計算する．
   (2) 生成： 得られた $\boldsymbol{Cs_0}$ から文を生成する．
   (3) 比較： 元の文と, 生成された文を比較する．

上記の方法で評価した際, もし元の文と生成された文が同じものであったならば, 学習によって獲得された力学系は, その文を表現する安定軌道を持ち, その安定軌道を生成する Cs 初期値が, $\boldsymbol{Cs_0}$ 空間内に埋め込まれていることになる．実験に用いた文は, 7 つのカテゴリに分類される 17 単語 (Table 1) と 9 個のルールからなる正規文法 (Table 2) から生成される 2〜6 単語文である．

4. 実験

本研究では, MTRNN による言語の獲得実験と, 言語の誤り訂正能力の検証実験の 2 つの実験を行った．

4·1 実験 1：言語獲得実験

3. 節に示した手順で, 言語の学習と MTRNN の評価を行った結果, 100 文中 95 文を正しく生成することが確認された．正解データの内訳は, 既学習文 75/80, 未学習文 20/20 である．比較実験として, 文法的に誤った語順を持つ文を 20 文認識させたところ, 元の通りに生成された文は 0 であった．文法的に正しければ未知の文であっても生成可能であることから, MTRNN は学習文を汎化して言語構造を自己組織的に獲得していることが分かる．また既学習にも関わらず正しい文が生成できないことがあるのは, その文を表す軌道の引き込み領域が狭く, うまく $\boldsymbol{Cs_0}$ が見つからないためだと考えられる．

Table 2 Regular grammar

S → V_I	NP → ART N
S → V_I ADV	NP → ART ADJ N
S → V_T NP	
S → V_T NP ADV	ADJ → ADJ_S
	ADJ → ADJ_C
	ADJ → ADJ_S ADJ_C

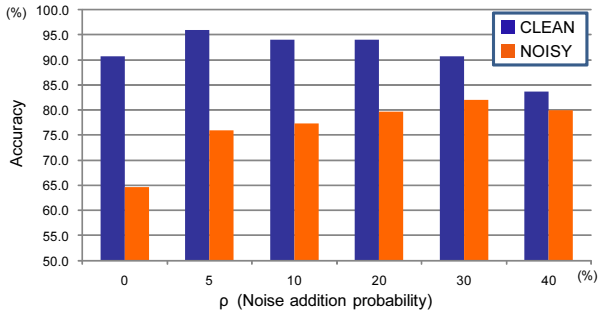

Fig.3 ノイズ生成率 ρ に対する文の認識・生成精度

4·2 実験 2：誤り訂正能力検証実験

1〜2 文字が別の文字に置き換えられた不正確な文を MTRNN に認識・生成させ, 元の正しい文に訂正する能力を検証した．また, 毎学習サイクルにおいて確率 ρ で文に誤りを付与しながら MTRNN の学習を行い, 誤り混入確率 ρ が訂正能力に及ぼす影響を調べた．実験では, $\rho = 0, 5, 10, 20, 30, 40(\%)$ の各設定に対して, 20 通りの結合重みの初期値から学習・評価を行った．

全体の中で最も優れた学習結果は $\rho = 30$ の設定で得られ, 誤り文を 86/100, 正常文を 98/100 という高い精度で正しく認識・生成した．また各 ρ に対して, 最も成績の良かった学習結果 3 パターンの平均正解数を Fig. 3 に示す．Fig. 3 より, $\rho = 30$ までは, 学習時の誤り混入率を上げるほど訂正能力が向上し, 正しい文に対する認識精度も $\rho = 0$ と同等かそれ以上を示すことが明らかになった．

以上から, データに誤りを付与して学習することで, 得られる軌道の引き込み領域が広がり, 力学系の安定性が向上することが明らかになった．

5. 解析

学習済み MTRNN が文章を生成する時の, 各ニューロン群の発火パターンを解析した結果, IO が文字を, Cf が単語を, Cs が文をそれぞれ表現していることが明らかになった．その論拠を解析結果と共に以下に示す．

- **IO**：IO ノードはそれぞれが文字に対応しているので, その発火パターンは明らかに文字を表現している．

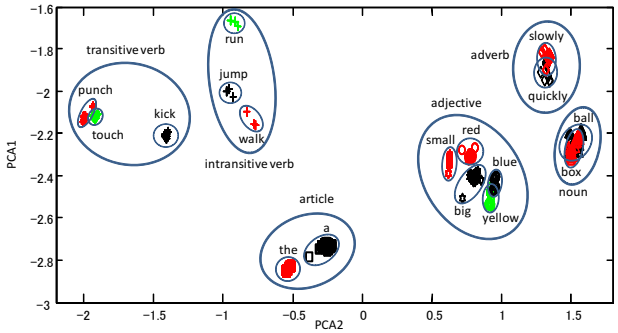

Fig.4 単語開始時の Cf 発火：品詞のクラスタが創発

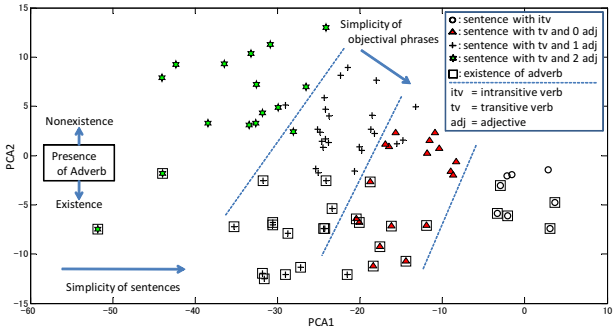

Fig.5 文開始時の Cs 発火：文構造のクラスタが創発

- **Cf**：Cf ノードの発火には以下の特徴が見られた．
  1. 同一文字に対し，異なる発火パターンを示す．
  2. 同一単語に対し，同一の発火パターンを示す．
  3. 同一品詞に対し，似た発火パターンを示す．

  これらの事実から，Cf は自己組織的に獲得した品詞情報を含む形で，単語を表現していることが分かる．参考として Fig. 4 に，単語開始時点での Cf 発火パターン (40 次元) の第一，第二主成分を示す．図において，1 文字目が同じであっても異なる発火をしている単語がある (例．“run”と “red”) ことから上記 (1) が確認できる．また，同一単語のクラスタが形成され，さらに品詞による一回り大きなクラスタが形成されていることから，(2), (3) も確認できる．
- **Cs**：Cs ノードの発火には以下の特徴が見られた．
  1. 同一単語に対し，異なる発火パターンを示す．
  2. 同一文に対してのみ同一発火パターンを示す．
  3. 類似構造文に対し，似た発火パターンを示す．

  これらの事実から，Cs は自己組織的に獲得した文構造の特徴をもとに，文全体を表現していることが分かる．参考として Fig. 5 に，文開始時点での Cs 発火パターン (11 次元) の第一，第二主成分を示す．図において，1 単語目が同じ文が多数存在するにも関わらず，全て異なる発火をしていることから上記 (1), (2) が確認できる．また，目的語の有無，目的語を修飾する形容詞の数，副詞の有無といった文構造上の特徴が状態空間に組織化されていることから，(3) が確認できる．

6. おわりに

本稿では，MTRNN による言語獲得実験を行い，その結果を報告した．実験では，単語や文法などの事前知識を与えることなく，文字の列として表現された文の集合のみを用いて MTRNN を学習させた．その結果，訓練後の MTRNN は未学習の文であっても正しく認識・生成可能であることが確認された．また，訓練後の MTRNN の発火を解析した結果，最も発火速度の速いニューロン群には “文字” が，中間的な発火速度を持つニューロン群には “単語” が，最も遅いニューロン群には “文構造” がそれぞれ自己組織化されていることが明らかとなった．この結果から，言語表現の多様性を担う仕組みである階層的な組み合わせ能力が，ニューラルネットによって獲得可能であることが示された．

また学習時に誤った文を与えることが逆にロバストな構造を生み，文の認識・訂正能力が向上することも確かめられた．一般に言語獲得においては与えられる言語刺激の質が低いことが問題視されるが，この結果はむしろ質の低い言語刺激がロバストな言語認知能力の発達を促す可能性を示唆している．

将来研究として，実世界と接地した言語認知の獲得を扱う予定である．具体的には，言語 MTRNN とロボットのセンソリ-モーター系を扱う MTRNN を，少数のニューロンを介して結合させる．これにより 2 つの MTRNN が力学的に相互作用し，ロボットの感覚運動系に接地した形で言語が獲得されることが期待される．

謝辞 本研究は，JST さきがけ「情報環境と人」，科研費学術創成研究（課題番号：19GS0208），科研費基盤研究 (B)（課題番号：21300076）の支援を受けた．

参考文献

[1] K. D. Bot, W. Lowie, M. Verspoor: “A dynamic systems theory approach to second language acquisition,” Bilingualism: Language and Cognition, vol.10, pp. 7–21, 2007.
[2] S. Lawrence and C. L. Giles: “Natural Language Grammatical Inference with Recurrent Neural Networks,” IEEE Trans. on Knowledge and Data Engineering, vol.12, no.1, pp. 126–140, 2000.
[3] Y. Sugita, J. Tani: “Learning semantic combinatoriality from the interaction between linguistic and behavioral processes,” Adaptive Behavior, vol.13, no.1, 2005.
[4] N. Chomsky: “Barrier,” MIT Press, 1986.
[5] N. Chomsky: “Rules and Representations,” Columbia University Press, 2005.
[6] J. B. Pollack: “The induction of dynamical recognizers,” Machine Learning, vol.7, no.2–3, pp. 227–252, 1991.
[7] J. L. Elman: “Language as a dynamical system,” Mind as Motion: Explorations in the Dynamics Cognition, MIT Press, pp. 195–223, 1995.
[8] T. Ogata et al.: “Two-way Translation of Compound Sentences and Arm Motions by Recurrent Neural Networks,” IROS, pp. 1858–1863, 2007.
[9] J. Tani and M. Ito: “Self-Organization of Behavioral Primitives as Multiple Attractor Dynamics: A Robot Experiment,” IEEE Trans. on Systems, Man, and Cybernetics Part A: Systems and Humans, vo.33, no.4, pp. 481-488, 2003.
[10] Y. Yamashita and J. Tani: “Emergence of Functional Hierarchy in a Multiple Timescale Neural Network Model: a Humanoid Robot Experiment,” PLoS Comput. Biol., vol.4, no.11, 2008.
[11] D. E. Rumelhart, G. E. Hinton, and R. J. Williams: “Learning internal representations by error propagation,” MIT Press., Ch. 8, pp. 318–362, 1986.