

科学研究費助成事業 研究成果報告書

平成 29 年 5 月 13 日現在

機関番号：11501

研究種目：挑戦的萌芽研究

研究期間：2014～2016

課題番号：26580105

研究課題名(和文) コネクショニズムを応用した第二言語習得研究：複数のシミュレータを用いた検証

研究課題名(英文) A Connectionist Study of Second Language Acquisition: Verified by Two Different Simulators

研究代表者

石崎 貴士 (ISHIZAKI, Takashi)

山形大学・大学院教育実践研究科・准教授

研究者番号：20323181

交付決定額(研究期間全体)：(直接経費) 2,200,000円

研究成果の概要(和文)：これまで言語学に対しても示唆に富んだ研究が多くなされているコネクショニズムという研究手法は、第二言語習得研究への応用も大いに期待できる。しかし、概説書等で紹介されている実践例は第二言語習得研究への応用を想定していないため、それらをそのまま適用することはできない。そこで本研究では、学習の対象となる同一の事象や項目を異なる二言語間で同様に記号変換できる共通の変換フォーマットを提案し、第二言語習得を模擬する学習シミュレーションが可能であることを示した後、複数のシミュレータを用いて検証し考察を行った。その結果、これまで単独のシミュレータを用いていた際にはわからなかった新たな側面が明らかになった。

研究成果の概要(英文)：This study is aimed at demonstrating a way of connectionist simulation for SLA studies. For the SLA simulation, two languages like Japanese and English need to be converted into binary signals based on the same common format. Therefore, as one of the main purposes of this study, a common format was proposed. Another main purpose of this study is utilizing two different computer simulators to conduct the SLA simulation. Verified by these simulators, some aspects which the conventional simulator had overlooked were revealed in this study.

研究分野：第二言語習得理論

キーワード：第二言語習得理論 コネクショニズム シミュレーション バイリンガル

1. 研究開始当初の背景

コネクショニズムは、人間の脳内の神経細胞がどのように働いているのか、その働きにヒントを得て考案された情報処理モデルである。従来の心理学実験のような人を対象にした実験は行わず、代わりにコンピュータによるシミュレーションを実施する。これまでコネクショニズムは、主に心理学の領域で注目されてきたが、言語学についても示唆に富んだ研究がなされている。Chomsky (1965)以降、言語学では生得論(子どもは生まれながらにしてあらゆる言語に対応可能な言語体系のもととなる規則を頭の中に備えている)による言語習得観が主流となっているが、コネクショニズムでは、従来規則を適用してなされると考えられてきた高次の情報処理も脳内の神経細胞を模して構成されたネットワークにより規則を適用することなく遂行が可能であることを、コンピュータを用いたシミュレーションによって示してきた(守1996)。このことから、コネクショニズムは第二言語習得研究においても、その応用が大いに期待できると言える。

ところが、概説書等で紹介されている実践例は、第二言語習得研究への応用を想定していないため、それらをそのまま適用することはできない。確かに、汎用性の高いOS上で動作し、プログラミング言語等の専門的な知識や技能がなくてもシミュレーションを実施できる簡便なソフトウェア(シミュレータ)も開発されている。しかし、第二言語習得研究に応用可能なシミュレーションを行うには、性質の異なる二言語間で学習の対象となる同一の項目を同様に記号変換しなければならないので、共通の変換フォーマットが必要となる。そこで本研究では、そのような具体的な共通フォーマットを提案し、コンピュータ上で第二言語習得を模擬するだけでなく、複数のシミュレータを用いてシミュレーションの結果がどの程度シミュレータに依存するものなのかも含めて検証しようという着想に至った。

2. 研究の目的

本研究では、性質の異なる二言語間で、学習の対象となる同一の項目を同様に記号変換できる共通フォーマットを提案し、第二言語習得を模擬する学習シミュレーションが可能となることを示した後、複数のシミュレータを用いて、その結果を比較・検証する。

汎用性の高いOS(Microsoft社Windows)上で動作し、プログラミング言語等の専門的な知識・技能がなくてもシミュレーションを実施できる簡便なシミュレータとしてオープンソースで公開されているものに tlearn がある。この tlearn を用いて、石崎(2012)では本研究に係る予備実験として、同一の被験者が母語での語順(統語)を学んだ後に第二言語での語順(統語)を学ぶ場合を、特定の言語に依存しない意味表象と個々の言語

に対応する単語の両方を信号化する共通のフォーマットを提案し、単純再帰ネットワークによって模擬した。しかしながら石崎(2012)では、二言語の併用を学習する際にどちらの言語で話し始めるのかについてはランダムとなるため文頭に来る語を検証の対象から外しているなど改善を要するところも見られた。そこで本研究では、新たな分析手法を導入することにより文頭に来る語についても検証の対象に加えるとともに、これらの第二言語習得を模擬する学習シミュレーションを別のシミュレータで実施し、上述の先行研究で得られた結果が、用いるシミュレータに依存するものなのかどうかを比較・検証する。

本研究によって、学習の対象となる同一の項目を性質の異なる二言語間で信号変換できる共通フォーマットが提案されれば、コネクショニズムを応用した第二言語習得研究という新たな分野の発展に寄与することができる。また、そのような共通フォーマットを用いて第二言語習得を模擬する学習シミュレーションを複数のシミュレータで実施し、結果を比較・検証することによってシミュレータの活用の仕方が認知されれば、第二言語習得研究の分野におけるコネクショニズムを応用した新たな方法論の提案にも貢献することが期待できる。

3. 研究の方法

(1) シミュレータ

本研究では、先行研究(石崎2012)で用いたシミュレータ(tlearn)とは別のシミュレータ(NeuroDimension社 NeuroSolutions Ver. 7.0.1.1)を用いる。それは、先行研究と異なるシミュレータを用いて結果を比較・検証するためであるとともに、新たな学習アルゴリズムの導入による効率化が図られ、先行研究よりも少ない試行数で学習を完了することが期待されたからである。また、製品化された有償のシミュレータを使用することは、技術的なサポートが得られるなど利点も大きい。

(2) 共通変換フォーマットの提案

まず特定の言語に依存しない形で内容(意味)を表象する信号を提示する。具体的にはTable 1に示したような7種の意味表象を(便宜的に文の形で表現しているが)特定の言語に依存しないパターンとしてとらえ、3ビットでコード化している。

Table 1
入出力信号の変換フォーマット【意味表象】

パターン	コード	日本語	英語
①	111	あなたは幸せに見える。	You look happy.
②	110	あなたたちは幸せに見える。	You look happy.
③	101	彼は幸せに見える。	He looks happy.
④	100	彼ら(男)は幸せに見える。	They look happy.
⑤	011	彼女は幸せに見える。	She looks happy.
⑥	010	彼女たちは幸せに見える。	They look happy.
⑦	001	彼ら(男女)は幸せに見える。	They look happy.
*	000	(意味表象にかかわらない)	

これらの意味表象を提示した後、それぞれに対応する文を構成する語が個々の言語で一語一語適切な語順で出力されるよう学習する。例えば、の意味を表象する信号を入力すると、日本語の場合は「あなたは」を出力し、英語の場合は“ You ”を出力する。次に「あなたは」を入力すると「幸せに」を出力し、「幸せに」を入力すると「見える」を出力、さらに「見える」を入力すると文末の句点を出力する。一方、“ You ”を入力すると“ look ”を出力し、“ look ”を入力すると“ happy ”を出力、“ happy ”を入力すると文末のピリオドを出力する。このように文を構成する全ての単語が正しい語順で予測できるようになれば当該の文を学習できたとなります。

これらの意味表象に対応する文を構成する各単語については、Elman (1991, 1993) が1語ずつ出力層のノードに割り当てていったのは異なり、複数の信号の組み合わせによって特定する形式を採用する。まず1ビットで当該の単語が日本語・英語のどちらであるかを特定し、次に2ビットでその単語が当該の文中で果たしている文法的な役割(主語・動詞・補語・文末)についてコード化する。また、単語の示す文法的な人称・性・数についても、それぞれ2ビットを用いて当該の範疇における区別があるかないか、ある場合はどちらの下位範疇に属するのか、あるいは区別は存在するが同じ形を用いて対応するのかを特定している (Table 2 を参照)。以上、意味表象に3ビット、単語の特定に9ビット、計12ビットを用いて入出力のために言語情報を変換する。意味表象を特定する場合は最初の3ビットを用いて行き、残りの9ビットには便宜的に全て“ 0 ”を割り当てて対応する。また、単語を特定する場合は、最初の3ビットには便宜的に全て“ 0 ”を割り当て、残りの9ビットを用いて対応する。

Table 2
入出力信号の変換フォーマット【単語】

	言語	役割	人称	性	数
日本語	主語(11)	二人(10)	男性(10)	単数(10)	
(0)	動詞(10)	三人(01)	女性(01)	複数(01)	
英語	補語(01)	同形(11)	同形(11)	同形(11)	
(1)	文末(00)	なし(00)	なし(00)	なし(00)	
あなたは	0	11	10	00	10
あなたたちは	0	11	10	00	01
彼は	0	11	01	10	10
彼らは	0	11	01	11	01
彼女は	0	11	01	01	10
彼女たちは	0	11	01	01	01
You	1	11	10	00	11
He	1	11	01	10	10
She	1	11	01	01	10
They	1	11	01	00	01
見える	0	10	00	00	00
look	1	10	11	00	11
looks	1	10	01	00	10
幸せに	0	01	00	00	00
happy	1	01	00	00	00
。(句点)	0	00	00	00	00
。(ピリオド)	1	00	00	00	00

このような変換フォーマットを用いて単純再帰ネットワークによる学習シミュレーションを実施した。シミュレーションでは、同

一の被験者(初期設定を同一にすることにより想定)を対象に、母語として日本語や英語を学ぶ場合、バイリンガルとして日本語と英語の両方を学ぶ場合を模擬した。

(3) シミュレーションの実施

ネットワーク：シミュレーションで構築する神経細胞のネットワークは石崎(2012)と同様の単純再帰ネットワークである。具体的なアーキテクチャは、入力層と出力層に意味表象を特定するための3ノード、単語を特定するための9ノード、計12ノードを配している。内的表象を形成する隠れ層についても同じ12ノードを配し、そのため隠れ層の状態を保持する文脈層も12ノードで構成されている。

手続き：石崎(2012)と同様、まず母語として日本語(または英語)を習得する場合を模擬する。ノード間の初期結合強度をランダムに割り振る乱数の種(ランダムシード)を特定の値に設定し、学習率と慣性項の値を統一して共通の初期状態とすることにより同一の被験者を想定する。母語として日本語(または英語)を習得する場合を想定した入力信号ファイルを用いて次に来るべき入力信号を正しく予測(出力)できるようにトレーニングを行う。母語習得を想定した入力信号ファイル(試行数を sweep と表現)として、7種類ある意味表象ごとに当該の意味表象(1 sweep)と該当する文を構成する単語3語(3 sweeps)と文末記号(1 sweep)からなるセット(計5 sweeps)を10個作り、それらをランダムに配置する。このファイルを一巡(1 epoch = 350 sweeps)すると、意味表象それぞれに対し10回のトレーニングを行ったことになるので、一巡ごとのノード間の結合強度をウェイトファイルとして保存しておく。トレーニング終了後、これらのウェイトファイルを指定してテスト機能を活用することにより、それぞれの時点での入力に対する出力値を確認することができる。その結果を踏まえて、次節に示す手法により分析を行う。

また、日本語と英語の両方を同時に習得するバイリンガルを想定した学習についても、各ノード間の結合強度の初期状態を母語習得のシミュレーションと同じにすることで同一の被験者を想定する。バイリンガルのシミュレーションでも次に来るべき入力信号を正しく出力できるようにトレーニングを行うが、7種類の意味表象それぞれに日本語文が該当するパターンと英語文が該当するパターンの両方が混在するため、作成する入力ファイルの試行数は母語習得のときの倍となる。ファイルを一巡(1 epoch = 700 sweeps)すると、意味表象それぞれに対し日本語で10回、英語で10回の試行をしたことになる。このシミュレーションでも一巡ごとのノード間の結合強度をウェイトファイル

として保存し、テスト機能を活用して分析を行う。その際、先行研究(石崎 2012)では用いていない新たな分析手法を導入することにより、バイリンガルのシミュレーションにおいても文頭に来る語を検証の対象に加えている。

分析手法: ウェイトファイルには 10 試行ごとのノード間の結合強度が保存されている。このファイルを呼び出せば当該の段階におけるネットワークの状態が復元でき、その段階での出力も確認することができる。石崎(2012)で用いたシミュレータ(tlearn)には、指定したパターンと出力値を照合して文字や記号に変換してくれる翻訳機能が備わっていたが、今回使用したシミュレータ(NeuroSolutions)には、そのような機能が備わっていないため表計算ソフト(Microsoft Excel 2010)を用いて、出力された値をIF関数で0か1かに変換(閾値は0.5に設定)し、さらにVLOOKUP関数によって指定したパターンと照合して翻訳を行う。

母語習得のシミュレーションでは、この手法によって全ての意味表象に対し日本語(あるいは英語)の該当する文が適切な語順(文末記号を含め)で出力できるようになった段階を学習完了とみなす。また、バイリンガルのシミュレーションでは日本語と英語の両方で該当する文が適切な語順で出力できるようになった段階を学習完了とみなす。但し石崎(2012)では検証の対象から除外した文頭に来る単語については、日本語・英語のいずれでも正解とし、さらにアドホック分析としてユークリッド距離の概念を導入した検証を実施して言語内・言語間のカウンターパートとの相対的な位置関係を測定する。

4. 研究成果

(1) 本実験の結果

まず、母語として日本語(あるいは英語)をネットワークに学習させたシミュレーションでは、日本語を学ぶ場合は各文 90 回の試行を経た段階で、英語を学ぶ場合については各文 140 回の試行を経た段階で、7 種類全ての意味表象に対し正しい文を出力できるようになった。先行研究(別のシミュレータを使用)である石崎(2012)でも、日本語は各文 3,200 回の試行で学習が完了したのに対し、英語は学習完了に各文 42,800 回の試行を要したので、今回のシミュレーションでも母語として学ぶ場合は英語よりも日本語の統語(語順)の方が習得しやすい可能性が示唆された。

次に、日本語と英語の両方を同時に習得するバイリンガルの想定したシミュレーションについては、異なる二言語を同時に学ぶことによる促進効果と干渉効果が見られた。バイリンガルの想定したシミュレーションでは、日本語と英語の両方を同時に学んでいるにもかかわらず学習完了に要した試行数は

各文 70 回であった。母語習得のシミュレーションで学習完了に要した試行数は、日本語が各文 90 回、英語が各文 140 回であったのでバイリンガルの場合は母語習得よりも少ない試行数で学習を完了することができたと言える。学習しなければならない課題の量は母語習得の倍になっているにもかかわらず母語習得よりも少ない試行数で学習を完了できたのは、異なる二言語を同時に学ぶことによって学習が促進されたためであると考えられる。

一方、干渉効果については、誤りが出力上のエラー(翻訳不能)としてあらわれる顕在化するタイプと出力上はエラーとしてあらわれない内在化するタイプの二つのタイプがあることがわかった。これらの干渉効果は、いずれも先行研究である石崎(2012)では検証の対象から除外した文頭に来る語で見られた。まず、顕在化した干渉効果としては次のような事例が見られた。という意味表象を入力した場合、正しい出力は日本語では「あなたたちは」、英語では“You”であるがバイリンガルの学習ではこれらのいずれでも正解としている。そこで、学習完了後である各文 270 回の試行のウェイトファイルを用いて当該の段階での出力を確認したところ、“You”と正しく出力できていた。ところが、次の各文 280 回の段階になると突然エラー(翻訳不能)を出力するようになってしまった。

Table 3
ユークリッド距離【顕在化した干渉効果】

② あなたたちは	各文 270 回 (出力: You)	各文 280 回 (出力: エラー)
あなたは	1.305	1.293
あなたたちは	0.851	0.825
彼は	2.164	2.158
彼らは	2.169	2.159
彼女は	2.164	2.158
彼女たちは	1.925	1.914
You	0.586	0.607
He	2.078	2.084
She	2.078	2.084
They	1.530	1.534

なぜ、このように突然エラーを生じるようになったのか、その原因を探るためにアドホック分析としてユークリッド距離を算出する検証を実施した(Table 3を参照)。その結果、正しい出力が見られた各文 270 回の試行の段階では正解の候補である「あなたたちは」(0.851)と“You”(0.586)がいずれもネットワークの出力値から近い位置にあり、さらに“You”の方がより近い(数値が小さい)位置にあるため“You”と出力したのだとわかった。これに対し、エラーを出力した各文 280 回の試行の段階では、正解の候補である「あなたたちは」(0.825)と“You”(0.607)がいずれもネットワークの出力値から近い位置にあることに変わりはないが、各文 270 回の試行時よりも「あなたたちは」が少し近づき、“You”が少し離れたため二つの候補の間で距離の差は縮まっている。そのため、これら二つの候補のいずれを選択するかでネ

ネットワークの判断が拮抗し、出力にエラーが生じたのだと考えられる。言い換えれば、二つのいずれにすべきかでネットワークに迷いが生じたとも言える。

一方、潜在化した干渉効果としては、次のような事例が見られた。学習完了後である各文300回の試行のウェイトファイルを用いて当該の段階での出力を確認したところ、という意味表象を入力した場合、正しい出力の候補は日本語の「彼女たちは」と英語の“ They ”のいずれかであり、実際の出力でも“ They ”と正しく出力されていた。また、という意味表象を入力した場合、正しい出力の候補は日本語の「彼らは」と英語の“ They ”のいずれかであり、実際の出力でも“ They ”と正しく出力されていた。

Table 4
ユークリッド距離【潜在化した干渉効果】

各文300回	㊶ 彼女たちは/They (出力: They)	㊷ 彼らは/They (出力: They)
あなたは	2.125	2.199
あなたたちは	1.625	1.720
彼は	1.908	1.774
彼らは	1.347	1.162
彼女は	1.640	1.711
彼女たちは	0.904	1.023
You	1.752	1.821
He	1.782	1.600
She	1.491	1.530
They	0.552	0.614

出力を見る限りは、何ら問題はないが、ここでもアドホック分析としてユークリッド距離を算出する検証を実施してみた(Table 4を参照)。すると、の意味表象を入力した場合は、正しい出力の候補である「彼女たちは」(0.904)と“ They ”(0.552)がいずれもネットワークの出力値から近い位置にあり、さらに“ They ”の方がより近い位置にあるため“ They ”と出力したのだとわかった。ところが、の意味表象を入力した場合は、正解の候補のうち“ They ”(0.614)のみがネットワークの出力値から近い位置にあり、「彼らは」は近い位置にはない。このことは、より単純な英語の方の学びを優先することによって、より複雑な日本語の学びを回避していることを示唆している。

今回の実験での学習シミュレーションは、脳における実際の言語処理と比較すれば極端に単純化されたモデルであるため、今回の結果を以てバイリンガルの学びを一般化することは早計である。さらに様々な角度からの検証が必要であることは論を俟たないが、実際の学習者の観察(あるいはスピーキングやライティングなど学習者による言語活動のテスト結果)では測定することのできない学習者の学びを客観的な数値として可視化できるという点において、このような学習シミュレーションによるアプローチは1つの可能性を示唆していると思われる。また、今回の実験では複数のシミュレータを用いて学習シミュレーションを実施したが、このことによって単独のシミュレータではわからなかった知見が得られたということも、今後、

このようなシミュレーションを実施する上で重要な指摘であると思われる。

(2) 本研究の成果

本研究では、性質の異なる二言語間で、学習の対象となる同一の項目を同様に記号変換できる共通フォーマットを提案し、第二言語習得を模擬する学習シミュレーションが可能となることを示した後、複数のシミュレータを用いて、その結果を比較・検証した。

本研究で提案された共通フォーマットを用いて第二言語習得を模擬する学習シミュレーションを実施することによって、接続主義を応用した第二言語習得研究という新たな原理の発展に寄与することができる。また、本研究では、そのような学習シミュレーションを複数のシミュレータで実施し、結果を比較・検証した。今後、このようなシミュレータの活用の仕方が認知されれば、第二言語習得研究の分野における接続主義を応用した新たな方法論の提案にも貢献することが期待できる。

参考文献(1から4に対応)

- 石崎貴士(2012)。「接続主義を応用した第二言語習得研究の試み : 第二言語による統語の学習」『山形英語研究』第13号, 11-35.
- 守一雄(1996).『やさしいPDPモデルの話 : 文系読者のためのニューラルネットワーク理論入門』東京:新曜社
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- (安井 稔(訳)(1970).『文法理論の諸相』東京: 研究社)
- Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning*, 7, 195-225.
- Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, 48, 71-99.

5. 主な発表論文等

〔雑誌論文〕(計3件)

石崎 貴士、日本語と英語のバイリンガルによる統語の学習 : コネクショニストモデルによるシミュレーション、東北英語教育学会研究紀要、査読有、第37号、2017、pp.1 - 12

石崎 貴士、第二言語学習者における言語カテゴリーの境界線 : コネクショニストモデルによるシミュレーション、東北英語教育学会研究紀要、査読有、第36号、2016、pp.1 - 12

石崎 貴士、バイリンガルにおける言語カ

テグリーの境界線：コネクショニストモデルによるシミュレーション、東北英語教育学会研究紀要、査読有、第 35 号、2015、pp.1 - 13

〔学会発表〕(計 3 件)

石崎 貴土 (代表)、日本語と英語のバイリンガルによる統語学習のシミュレーション：コネクショニストモデルを応用した第二言語習得研究、全国英語教育学会、2016 年 8 月 20 日、独協大学 (埼玉県草加市)

石崎 貴土、第二言語学習者の言語カテゴリーにおける境界線：コネクショニストモデルを応用したシミュレーション、全国英語教育学会、2015 年 8 月 22 日、熊本学園大学 (熊本県熊本市)

石崎 貴土、コネクショニスト・モデルを応用したバイリンガル研究：言語カテゴリーにおける境界線のシミュレーション、全国英語教育学会、2014 年 8 月 10 日、徳島大学常三島キャンパス (徳島県徳島市)

6. 研究組織

(1) 研究代表者

石崎 貴土 (ISHIZAKI, Takashi)
山形大学・大学院教育実践研究科・准教授
研究者番号：20323181

(2) 研究分担者

中西 達也 (NAKANISHI, Tatsuya)
山形大学・地域教育文化学部・教授
研究者番号：10217771