

単語の分節化と認知過程に関する教師なし学習モデル

牧岡 省吾

大阪府立大学 人間社会学部 人間科学科

子どもは、母語を獲得する過程において文を単語に分節化することを学習しなければならない。子どもの周囲の大人は単語を1個ずつ区切って発音するわけではない。そのため子どもは、切れ目のない音素系列として発話される文を、何らかの方法で分節化の必要がある。

単語の分節化の学習過程のモデルとして代表的なものに、Brent& Cartwrightの記号表現を用いたモデルがある¹⁾。彼らは、入力文の可能な分節化パターンに対してDR functionと呼ばれる評価関数をあてはめることにより、教師なしで分節化の学習が可能であることを示した。Bernstein-Ratnerのコーパス²⁾に対する分節化の成功率(recall率)は、DR functionのみを使用した場合に47.3%、単語の始点と終点の発音可能性に関する手がかりを外から与えた場合には69.2%だった。Recall率とは、モデルが単語の前後を正しい位置で分節化した回数(コーパス全体における延べ回数)を、コーパス中の全単語の延べ出現回数で割ったものである。Brentらのモデルは、コーパス全体を一度に入力することによって最適化を行っており、実際の幼児のように1文ずつ学習を行うわけではないという問題点がある。

一方、Yangは、音節の遷移確率に関する統計的学習と、音声的手がかりに関する生得的制約条件を組み合わせることによって分節化が可能であることを示した³⁾。彼らは、音節の遷移確率だけにに基づく分節化は局所解に落ち込みやすいことを指摘し、「単語は必ず第一強勢(primary stress)をもつ」という制約条件を用いれば

局所解への落ち込みが避けられることを示した。Yangのモデルは90.1%という極めて高いRecall率を達成したが、英語の音節と連続する子音列を音節に分解する方法に関する事前知識を与えているという問題がある。また、日本語のように第一強勢に基づく手がかりが有効でない言語が存在する。日本語の場合はアクセント変化の存在が単語境界の手がかりになる場合が多いが、いわゆる無アクセント方言ではそれも使用できない。

複数の言語を対象とした研究としては、BatchelderのBootLexモデルがある⁴⁾。このモデルは各言語における単語の平均的長さを制約に含めた評価関数を用いることにより、英語のコーパスに対して68.2%、日本語のコーパスに関しては56.0%のRecall率を示した。BootLexは、平均単語長以外の事前知識が与えられず、学習を1文ずつ行うという点で、実際の幼児のモデルとしての妥当性が高い。

本研究では、単語の分節化過程と聴覚的認知過程は不可分であると考え、両者の学習過程を統一的に説明できるモデルを構築する。両者を統一的に扱うモデルとしてはNorrisのShortlist⁵⁾があるが、このモデルは学習過程を含んでいない。これまでに著者は、入力された音素系列を比較・照合することによって単語に分節化するニューラルネットワークモデルを提案した⁶⁾⁷⁾⁸⁾。今回、英語と日本語の両方について良好な成績を示し、かつ単語の聴覚的認知過程の時間特性を再現できるようなモデルを構築したので報告する。

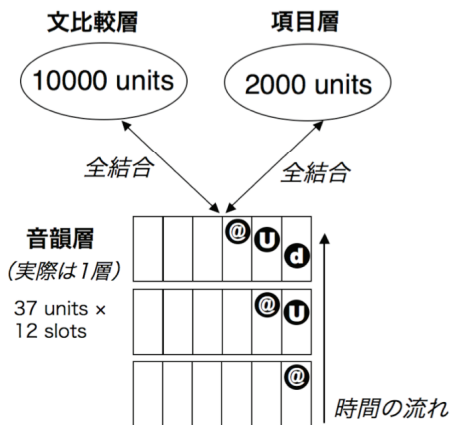


図1 モデルの全体像。入力パターンは「Oh dear」の音素表記。

モデル

提案するモデルSegReg(a model of lexical Segmentation and Recognition)の概念図を図1に示す。SegRegは、入力層である音韻層と、記憶した文と入力文を比較する

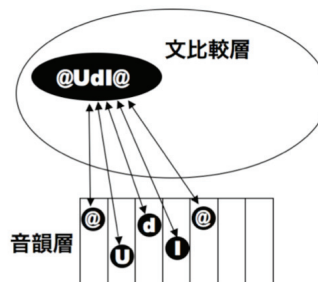


図2 文全体の学習

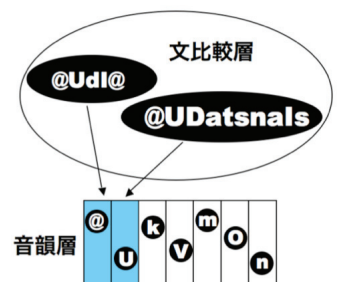


図3 比較による一致箇所の検出

文比較層，単語の表現を生成する項目層から構成される。音韻層はスロット型の構造をもち、各スロットには英語あるいは日本語の全ての音素に1対1対応するユニットが存在する。入力文はスロットの右端から呈示され、時間の経過とともに左側に移動していく（moving window型の入力表現）。

SegRegは現在の入力文を記憶済みの文と比較することによって単語を抽出する。文比較層は、入力文を順次記憶していく役割を担っている（図2）。文の最後の音素が呈示されたとき、その文を表現するユニットが文比較層に存在しなければ、新たなユニットがその文を記憶するために割り当てられる。学習はwinner-takes-all型であり、入力文に関する完全局所表現が生成される。学習規則には、牧岡⁹⁾が提案した忘却付き即時学習規則を用いる。即時学習と忘却を組み合わせることには、次のような利点がある。自然な状況下における親の発話では、近接する文において同じ単語や句が繰り返されることが多い。入力文に対する学習が即時的に行われることにより、近接する文同士の共通性を抽出し、新たな内部表現を生成することができる。一方、忘却によって、偶然生じた不適切な一致箇所に対する学習を、ある程度防ぐことができる。

文比較層は、記録された文と入力文とを常に比較し続ける。記録された3つ以上の文と現在の入力文に共通する音素系列が存在するとき、その系列の先頭部分が発音可能で、系列の中に少なくとも1つの母音が含まれていれば、項目層はその系列に対して1つのユニットを割り当てる。ここでも文比較層と同様、winner-takes-all型の忘却付き即時学習が用いられる。

複数の文同士で一致する部分は単語に該当していることが多いので、上記のような学習により、教師なしで単語の候補を獲得することができる。このような学習方法は須賀・久野が提案した「差異・共通性抽出原理」¹⁰⁾を参考にしたものである。

SegRegの特徴の一つは、項目層の表現が時間不変性をもつことである。文比較層が一致系列を抽出すると（図3）、その系列の呈示開始から終了までの各時点における音素ユニットと項目ユニットの間で結合が強化される（図4）。このような学習により、音素系列がスロット上のどこにあっても、項目層はそれを検出することが可能になる。これは、項目層がスロット位置に関する不変性、すなわち呈示開始からの経過時間に関する不変性を実現していることを意味する。さらに、このような不変性を備えることは、同時に、単語が最後まで呈示されるのを待たず、最初の音素が呈示された時点からその単語に対応する項目ユニットが活動を始めることを意味する。母語を獲得した大人は、単語を最後まで聞き終わらなくても、その単語を一意に特定できるだけの音素が入力された時点で同定できることが知られている¹¹⁾。SegRegは、項目層の表現の時間不変性を実現することにより、このような人間の反応

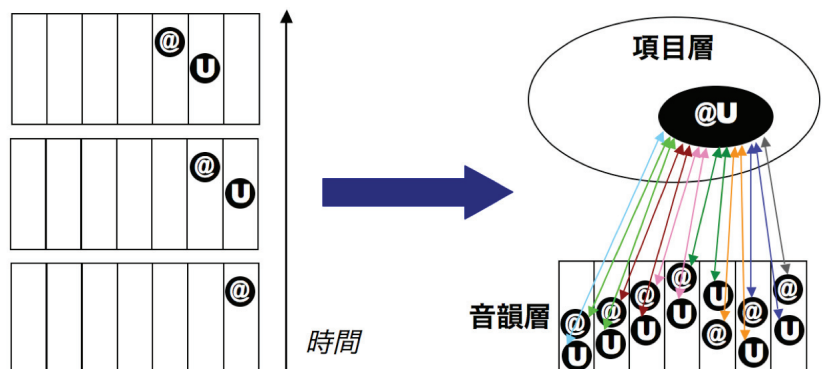


図4 項目層のユニットは、呈示開始から終了までの各時点の音韻層のユニットとすべて結合する。

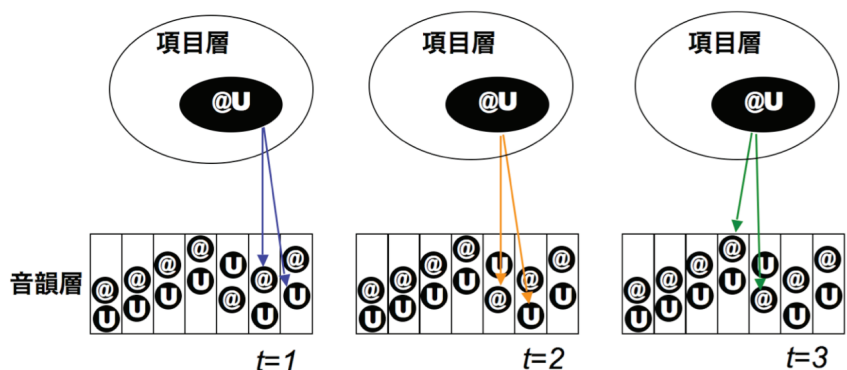


図5 項目層から音韻層へのトップダウン結合の時間依存性。項目層のユニットは「今どこにどの系列が存在するか」を知ることができる。

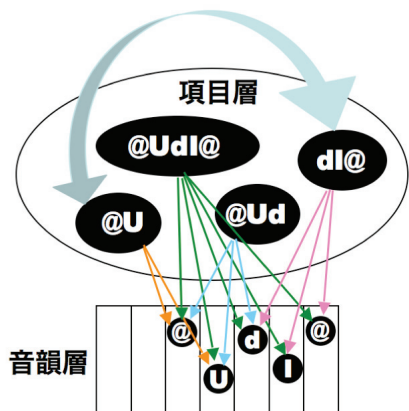


図6 項目層における競合のメカニズム。隣接する項目は、相互に活性化を高め合う。

特性を再現することを可能にしている。

このように時間不変性をもつ表現には利点もあるが、同じ音素から構成される異なる単語同士の区別が困難になるという問題点がある。これを防ぐために、項目層から音韻層に向かうフィードバック結合には、項目ユニットが活動を開始してからの経過時間に応じて動作するゲート回路が備えられている(図5)。単語は、項目層から音韻層へのトップダウン・フィードバックが入力と一致したときにだけ認知される。照合によって認知が成立する点はGrossbergのARTと同様である¹²⁾。

SegRegが学習済みの単語を認知する過程は以下の通りである。入力文に学習済みの項目が含まれていると、その最初の音素が呈示された時点から、該当する項目ユニットが活動を始める。項目の最後の音素が呈示された時点で、項目ユニットから音韻層へのトップダウン・フィードバックと入力パターンが完全に合致すると、その項目はSegRegが最終的に認知する単語の候補となる。

しかしSegRegは教師なしで学習しているため、単語に包含される短い音素系列や、単語間境界をまたぐ不適切な系列も項目として学習している。SegRegは、項目ユニット間の競合により、位置が重複する候補の中で最大活性化をもつ項目ユニットを1つ残して他を抑制する(図6)。競合のメカニズムは以下の通りである。前述のように、項目ユニットはトップダウン結合により音韻層へフィードバック信号を送る。2つの項目からのフィードバック・パターンが入力スロット上で隣接していれば、それら2つの項目には整合性があり、重複していれば整合性がない。そこで、隣接する複数の項目ユニット同士の活性化が相互に伝搬し、それらの

ユニットの活性化の総和を、項目群の活性化とするようなメカニズムを構築した。加えて、項目が表現する音素系列が長く、その音素系列の出現頻度が高いほど競合に有利となるようなバイアス項が設けた。項目群の活性化にバイアス項を加えた値が競合に用いられ、位置が重複する候補の中で最大活性化をもつ項目ユニットが1つ生き残る。この項目が、SegRegが最終的に認知する単語となる。短い音素系列の頻度をカウントすると”a”や”ta”といった高頻度の音節が競合に勝ちやすくなってしまいう傾向が見られたため、英語では3音素以上、日本語では5音素以上の系列のみ頻度バイアスの対象とした。この制約は、BatchelderのBootLexで用いられた平均単語長に関する制約と同様なものである。

シミュレーション

英語の訓練刺激としては、CHILDES¹³⁾に登録されたBernstein-Ratnerのコーパス²⁾を用いた。コーパスから養育者の発話のみを抜き出し、単語間の空白を除去した上でCMU Pronouncing Dictionary¹⁴⁾を用いて音素記号に変換した。次に個々の音素をベクトルパターンに変換した。個々の音素の表現としては、その音素に対応する1つのユニットのみが1の値をとり、その他のユニットの値が0であるような局所表現を用いた。訓練刺激には8,610文が含まれていた。

訓練時には、発話を1文ずつモデルに呈示し、教師なしで学習を行った。コーパス中では単語は分かち書きされているが、シミュレーションにおいては、1つの文に含まれる単語は切れ目なく呈示した。コーパス中の分かち書きと同じ位置で分節化された項目を認知したとき、分節化に成功したとみなした。学習の終了後、訓練刺激をもう1度呈示し、学習をオフにして分節化の成功率を計算した。日本語の訓練刺激としてはMiyataのコーパス¹⁵⁾を用い、学習の手続きは英語の場合と同様であった。訓練刺激には9,656文が含まれていた。

結果と考察

分節化の成功率を図7に示す。分節化の成功率(Recall率)は英語で68.9%、日本語で55.8%だった。日本語の方が分節化の成功率が低い理由としては、音素の種類が少ないために同一の音素系列の繰り返しが

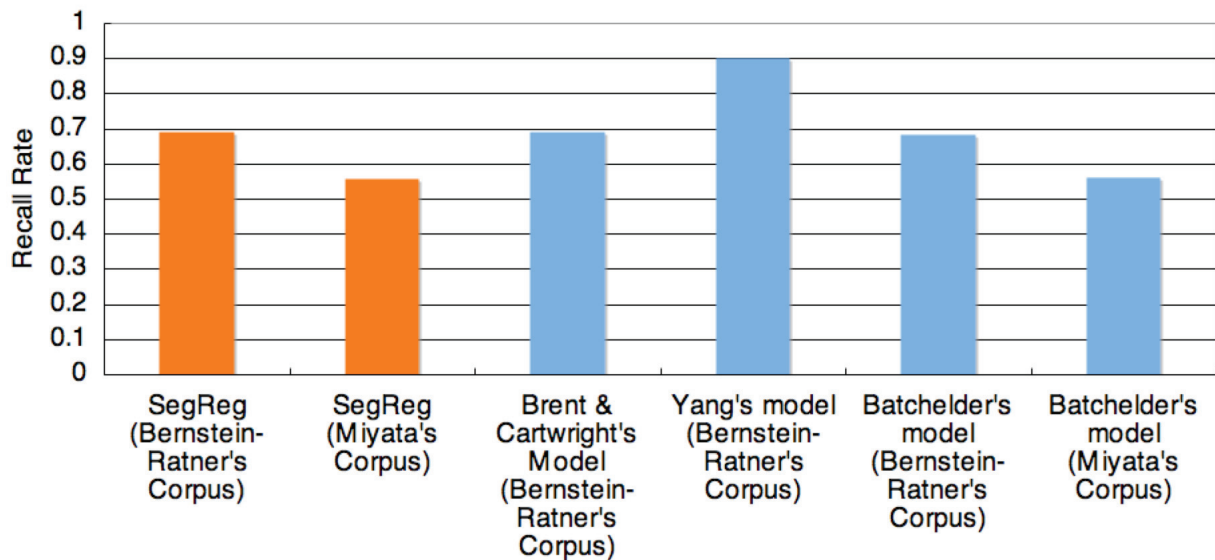


図7 本研究と先行研究のシミュレーションの結果

多くなり、項目層が単語でない系列を抽出してしまうことが多くなったためだと思われる。

SegRegの分節化成功率は、Brent¹⁾やBatchelder⁴⁾のモデルとほぼ同程度であった。Yang³⁾のモデルのような音節の発音やストレスパタンに関する事前知識をもつモデルを除外すれば、従来の記号処理モデルと同程度の分節化成績を達成することができた。一方、SegRegは、単語の聴覚的認知における競合過程を再現することができるという点で、従来のモデルより優れている。

SegRegは分節化と単語認知を包括的に扱うことができる枠組みであり、ストレスやアクセントなどの音声的情報を競合過程に取り入れることも可能である。さらに、一致箇所検出に基づく学習は動詞の過去形などの語形変化の学習に関しても適用可能であると思われる。今後、これらの拡張とともに学習アルゴリズムに関する心理学的検証を行う必要がある。

文献

- 1) Brent, M.R., Cartwright, T.A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation, *Cognition*, 61, 93-125.
- 2) Bernstein-Ratner, N. (1984). Patterns of vowel modification in motherese, *Journal of Child Language*, 11, 557-578.
- 3) Yang, C.D. (2004). Universal grammar, statistics or both?, *Trends in Cognitive Science*, 8, 451-456.
- 4) Batchelder, E. O. (2002). Bootstrapping the lexicon: A computational model of infant speech segmentation, *Cognition*, 83, 167-206.
- 5) Norris, D. (1994). Shortlist: A connectionist model of

continuous speech recognition. *Cognition*, 52, 189-234.

- 6) 牧岡省吾 (2005). 幼児は親の発話をどのように分節化するのか: 比較・照合アルゴリズムに基づく音素系列の分節化と再認のモデル, 日本神経回路学会誌, 12(3), 174-182.
- 7) 牧岡省吾 (2007). 一致箇所検出と項目間競合に基づく単語の分節化と認知のコネクショニストモデル, 日本認知科学会第24回大会発表論文集, 176-177.
- 8) Makioka, S. (2007). A Connectionist Model of Lexical Segmentation, *Studies in Language Science*, 6, 45-56.
- 9) 牧岡省吾 (2003). 音韻ループのコネクショニストモデル, 認知科学, 10, 93-103.
- 10) 須賀哲夫, 久野雅樹 (2000). ヴァーチャルインファント, 北大路書房.
- 11) Marslen-Wilson, W. D., Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech, *Cognitive Psychology*, 10, 29-63.
- 12) Grossberg, S. (1978). Adaptive pattern classification and universal recoding: Part I. Parallel development and coding of neural feature detectors. *Biological Cybernetics*, 23, 121-134.
- 13) MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk. Third Edition*, Lawrence Erlbaum Associates, Mahwah, NJ.
- 14) The CMU Pronouncing Dictionary, <http://www.speech.cs.cmu.edu/cgi-bin/cmudict>.
- 15) Miyata, S. (2000). The TAI Corpus: Longitudinal Speech Data of a Japanese Boy aged 1;5.20 - 3;1.1 Bulletin of Shukutoku Junior College 39, 77-85.