3D1-OS-37a-3

MLDA と教師なし単語分割に基づく概念と言語の相互学習 -育児語が概念形成に与える影響の解析-

Mutual Learning of Concepts and Language Based on MLDA and Unsupervised Word Segmentation -Analysis of the Effect of Infant-Directed Speech on Concept Formation-

船田 美雪 Miyuki FUnada　中村 友昭 Tomoaki Nakamura　長井 隆行 Takayuki Nagai　金子 正秀 Masahide Kaneko

電気通信大学大学院情報理工学研究科
Faculty of Infomatics and Engineering, The University of Electro-Communications

Humans acquire different languages and concepts for each culture from perceptual information and utterances given by caregivers. In addition, the utterances that the caregivers use are changed from Infant-Directed Speech (IDS) to Adult-Directed Speech (ADS) according to the growth of the infants, and humans learn both IDS and ADS in such a change. We have proposed a probabilistic model that enables robots to acquire the concept and the language by mutual learning of them. In this paper, we use IDS and ADS as teaching utterances to learn mutual learning model, and verify the influence of teaching language on concept formation.

1. はじめに

人の赤ちゃんは，外界から得られる知覚情報や養育者からの語りかけによって言語や概念を獲得する．このとき，教示される言語の特徴が母語の獲得に影響を与え，言語や文化毎に異なる概念が構築されると考えられる．また，養育者は幼児に対して，lnfant-Directed Speech（IDS）と呼ばれる特有な話し方を用いる傾向にあり，これは，成人に対する話し方である Adult-Directed Speech（ADS）と異なることが知られている [五十嵐 06]．IDS の特徴として，ADS に比べてピッチが高くなること，ピッチ幅が拡大すること，発話速度が遅いことが挙げられる．これらの IDS の特徴は多くの国や地域でも共通してみられる現象であることが知られている．特に日本においては，養育者は幼児に向けて擬音語・擬態語や音韻反復などの形態的特徴を持った育児語を用いる傾向がある．このような育児語は文化によってはほとんどみられない地域があり，日本の特徴的な傾向である．また，養育者は幼児に対して成長と共に成人語を用いる傾向があり，このような変化の中で，人は育児語と成人語の双方を学習する [小林 15].

我々は，自身が取得可能なマルチモーダルな知覚情報をクラスタリングすることで形成されるカテゴリが概念であると考え，さらに他者とのインタラクションを通して言語を学習し，概念と結びつけることで語意を獲得できると考えている．谷口らはこのような概念・言語学習を記号創発システムとして捉え，次のように説明している [谷口 14].「記号創発システムでは，自律的なシステムが『認知的な閉じ』の中で，自らの感覚器を通して得たマルチモーダル情報から，外部の教示がなくとも，概念を形成することができる．しかし，他者とのコミニュケーションをとるために，他者とのインタラクションの中で共通の記号系という大域的な秩序が形成される．そのような記号系によって，自らの概念が制約を受ける．このように，ボトムアップに創発した共通の記号系が，個々の概念形成に制約を与える．このようなループによって，記号的コミュニケーションを創発するシステムが記号創発システムである.」我々はこれまで，ロボットが取得したマルチモーダル情報から概念と言語が相互に影響しあいながら学習するアルゴリズムを提案してきた [中村 15]．このアルゴリズムにおける概念・言語獲得過程は記号創発システムだと見なすことができる．しかし，そのような学習に関わる要因と，その要因が学習モデルに及ぼす影響は十分に検証されていない．そこで本稿では，概念と言語の相互学習モデルにおける育児語と成人語の学習の過程と，それぞれの語が概念形成に及ぼす影響を検証することを目的とする．最終的には，我々が提案してきた学習モデル [中村 15] に基づいて，育児語と成人語が混在する教示発話を用いて物体概念の学習を行うことを検討しているが，本稿では育児語と成人語を区別し，それぞれの語について概念形成シミュレーションを行う．

連絡先: 船田 美雪，電気通信大学大学院情報理工学研究科，東京都調布市調布ヶ丘 1-5-1，m_funada@radish.ee.uec.ac.jp

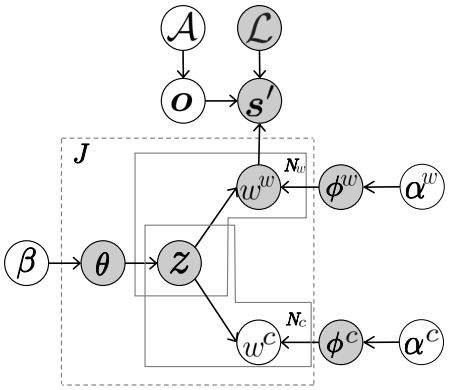

図 1: 色概念と言語のグラフィカルモデル

2. 概念と言語モデルの相互学習

文献 [中村 15] において，我々はこれまで，物体概念を扱ってきた．しかし，この物体概念と言語の相互学習モデルは非常に複雑であり，言語が概念形成に及ぼす影響を単純に解析することはできない．そこで，本稿では，文献 [船田 16] の色の概念学習を対象とする．図 1 は，言語モデルと色概念を統合したグラフィカルモデルであり，灰色で示されたノードは未観測ノードを表している．図中の o が人から教示される音声である．この音声を，$\mathcal{A}$ をパラメータとする音響モデル，$\mathcal{L}$ をパラ

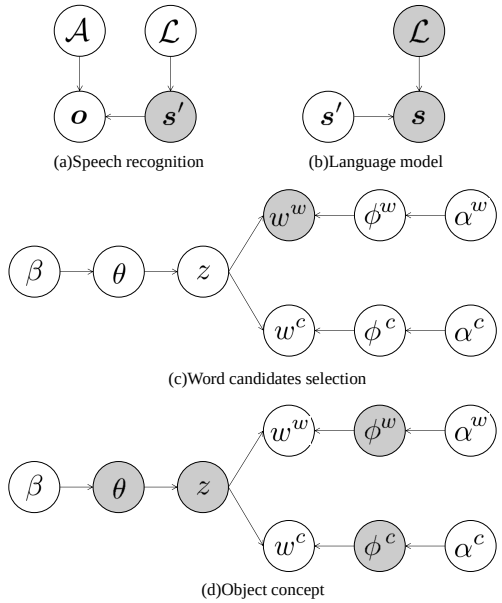

図 2: 言語と概念の近似モデル

メータとする言語モデルにより認識した結果が s である．さらに，認識結果 s を言語モデル $\mathcal{L}$ を用いて単語へ分割し，Bag of words (BoW) 表現へと変換したものが単語情報 w^w であり，さらに，w^c はそれぞれの色から得られる知覚情報を示している．

また z は色のカテゴリを表している．さらに，w^c，w^w は，それぞれ β^c，β^w をパラメータとする多項分布から発生する．これらの多項分布は，それぞれ π^c，π^w をパラメータとするディリクレ事前分布に従う．また，カテゴリ z の出現確率分布を表す多項分布のパラメータを θ とする．このパラメータは，ハイパーパラメータ α により決まるディリクレ事前分布に従う．そして，J は，教示データの数，N_c，N_w は，モダリティ c，w の情報の生起回数を表している．このモデルでは，音声認識結果 s と色のカテゴリ z が単語 w^w によって接続されているため，音声認識と色概念が相互に影響するモデルとなっており，音声認識・単語の接地・概念獲得などが統合されたモデルとなっている．このモデルにより，概念と言語モデルを相互に学習する．

3. マルチモーダル概念形成

本稿では，ロボットに対して，色を提示し，同時にその色名を音声により教示することで概念学習を行うことを想定している．しかし，図 1 のモデルは非常に複雑であり，全てを同時に学習することはできない．そこで，このモデルを図 2 のように分解し，それぞれの部分ごとに学習する．

3.1 音声認識

図 2(a) は，音声認識モデルである．ここでは音響モデルのパラメータ $\mathcal{A}$ と，言語モデルのパラメータ $\mathcal{L}$ は既知であるとし，各物体に与えられた全教示発話 $\boldsymbol{o}$ から，n-best の認識文字列 $\boldsymbol{s}'_{1:N}$ を得る．

$$\boldsymbol{s}'_{1:N} \sim P(\boldsymbol{s}'_{1:N}|\boldsymbol{o}, \mathcal{A}, \mathcal{L}) \tag{1}$$

このとき，本稿では音声認識に Julius を用い，Julius 標準の音響モデルを用いた．

3.2 単語の分割

次に，言語モデルのパラメータの推定を行う．言語モデルのパラメータ $\mathcal{L}$ は，ある物体に対して与えられた教示発話 $\boldsymbol{o}$ を生成する確率 $P(\boldsymbol{o}|\mathcal{A}, \mathcal{L})$ を最大化することで，求めることができる．

$$\mathcal{L} = \underset{\mathcal{L}}{\operatorname{argmax}} P(\boldsymbol{o}|\mathcal{A}, \mathcal{L}) \tag{2}$$

$$= \underset{\mathcal{L}}{\operatorname{argmax}} \int P(\boldsymbol{s}|\mathcal{L}) P(\boldsymbol{o}|\boldsymbol{s}, \mathcal{A}) d\boldsymbol{s} \tag{3}$$

しかし，$\boldsymbol{s}$ についての積分は，全ての文字の組み合わせの和を取ることを意味しており，直接計算することができない．そこで，ここでは $P(\boldsymbol{o}|\boldsymbol{s}, \mathcal{A})$ は一部の文字列以外の確率は非常に小さく無視できるとして，$\boldsymbol{o}$ の n-best の認識文字列 $\boldsymbol{s}'_{1:N}$ のみを用い，図 2(b) のように音声認識モデルと切り離して考える．つまり，教示発話 $\boldsymbol{o}$ を生成する確率を最大とする代わりに，教示発話 $\boldsymbol{o}$ を認識した n-best の認識文字列 $\boldsymbol{s}'_{1:N}$ から単語列 $\boldsymbol{s}_{1:N}$ を生成する確率を最大とすることで，言語モデルパラメータ $\mathcal{L}$ を近似的に計算する．

$$\mathcal{L}, \boldsymbol{s}_{1:N} = \underset{\mathcal{L}, \boldsymbol{s}_{1:N}}{\operatorname{argmax}} P(\boldsymbol{s}_{1:N}|\boldsymbol{s}'_{1:N}, \mathcal{L}) \tag{4}$$

このとき本稿では，教師なし形態素解析である Nested Pitman-Yor Language Modeling [持橋 09] を用いて，$\mathcal{L}$ と $\boldsymbol{s}_{1:N}$ を計算する．

3.3 単語の選択

次に，言語モデルと物体概念の双方を考慮した単語を生成する．

$$w^w = \underset{w^w}{\operatorname{argmax}} P(w^w|\boldsymbol{o}, \mathcal{A}, \mathcal{L}, w^c, \alpha^w, \theta, \pi, \beta) \tag{5}$$

$$= \underset{w^w}{\operatorname{argmax}} \int P(\boldsymbol{o}|\boldsymbol{s}, \mathcal{A}, \mathcal{L}) \times P(\boldsymbol{s}|w^w, \mathcal{L}) P(w^w|w^c, \alpha^w, \theta, \pi, \beta) d\boldsymbol{s} \tag{6}$$

ただし，w^c は，物体から得られる知覚情報である．しかし，この式においても $\boldsymbol{s}$ について積分することが困難であるため，直接計算することができない．そこで，ここでも $P(\boldsymbol{o}|\boldsymbol{s}, \mathcal{A}, \mathcal{L})$ は，一部の単語列以外の確率は非常に小さく無視できると考え，音声認識の n-bset から得た確率の高い単語列 $\boldsymbol{s}_{1:N}$ を利用し，図 2(c) のように音声認識部と単語の分節化部を切り離して考える．すなわち，教示発話 $\boldsymbol{o}$ と物体概念から確率が最大となる w^w を選択するのではなく，音声認識によって得られる n-best の単語列 $\boldsymbol{s}_{1:N}$ が物体概念から生成される確率が最大となる単語 $\bar{w}^w$ を選択する．

$$\bar{w}^w = \underset{w^w}{\operatorname{argmax}} P(w^w|\boldsymbol{o}, \mathcal{A}, \mathcal{L}, w^c, \alpha^w, \theta, \pi, \beta) \tag{7}$$

$$\approx \underset{w^w_n}{\operatorname{argmax}} \frac{P(\bar{w}^w|w^c, \alpha^w, \theta, \pi, \beta)}{\sum_{n=1}^{N} P(w^w_n|\boldsymbol{o}, \mathcal{A}, \mathcal{L}, w^c, \alpha^w, \theta, \pi, \beta)} \tag{8}$$

ただし，$\bar{w}^w$ は単語列 $\boldsymbol{s}_{1:N}$ の $\bar{n}$ 番目の単語集合を意味し，$\bar{n}$ は以下の式で表すことができる．

$$\bar{n} = \underset{n}{\operatorname{argmax}} P(w^w_n|w^c, \alpha^w, \theta, \pi, \beta) \tag{9}$$

3.4 単語の再分割

3.2 章で計算した言語モデルには n-best 全て使用したため，適切ではない単語も含まれている．そこで，n-best の中から

選択された認識文字列 $\bar{\boldsymbol{s}}'$ から，単語の再分割を行い，言語モデルのパラメータを再計算する．このとき言語モデルパラメータ $\mathcal{L}$ は，式 (4) と同様にして，認識文字列 $\bar{\boldsymbol{s}}'$ から単語列 $\bar{\boldsymbol{s}}$ を生成する確率を最大とすることで，近似的に計算する．

$$\mathcal{L}, \bar{\boldsymbol{s}} = \underset{\mathcal{L}, \bar{\boldsymbol{s}}}{\operatorname{argmax}} P(\bar{\boldsymbol{s}} | \bar{\boldsymbol{s}}', \mathcal{L}) \tag{10}$$

3.5 概念形成

ロボットが取得した知覚情報と，3.2 節の処理によって得られる単語をクラスタリングすることで，概念の獲得が可能となる．クラスタリングには統計モデルの一種であるマルチモーダル LDA[中村 10] を用いる．図 2(d) にマルチモーダル LDA のグラフィカルモデルを示す．図中の z がカテゴリを表しており，θ をパラメータとする多項分布から生成される．w^* はロボットが取得したセンサ情報であり，w^c が色情報である．本稿では，色情報として RGB 画像を L*a*b*表色系に変換し，a*と b*の値のヒストグラムを用いた．a*成分が 8 個，b*成分が 8 個となるように量子化を行い，合計 $8 \times 8 = 64$ 次元のヒストグラムとした．また，w^w が言語情報であり，単語の発生頻度ヒストグラムを用いた．これらの情報 w^* は，ϕ^* をパラメータとする多項分布より生成される．α^* と β は，それぞれ ϕ^* と θ の事前分布であるディリクレ分布のパラメータである．ロボットが取得した複数の情報 w^* からモデルのパラメータをギブスサンプリングにより推定することにより，ロボットはセンサ情報を範疇化した概念 z を教師なしで獲得することができる．

4. 概念と言語モデルの相互学習のシミュレーション実験

本稿では，3.2 章の色概念の学習において，特に「ぐるぐる」や「とんとん」といった音韻反復型の育児語の影響を解析する．

4.1 実験設定

まず，ランダムな色の単色画像とその画像の特徴を表す教示発話の組を 160 組生成した．次に，このうち 100 組を学習データとし，1 個から 100 個まで変化させながら学習を行い，各データ数毎に学習されたモデルを用いて 60 組のデータの認識を行った．このとき成人語での教示の場合には，「あか」や「あお」などの色語を用い，「これはあかだよ」といった教示発話を与えた．一方，育児語では，必ずしも全ての色語と対応した音韻反復型の育児語が存在するわけではない．そのため，各色に対して擬似的に音韻反復語型の単語を割り当てて，育児語を用いた学習をシミュレーションした．例えば，赤色に対しては「ぐるぐる」といった単語を割り当てて，「これはぐるぐるだよ」といった教示発話を与えた．すなわち，育児語での教示の場合は，成人語での「あか」や「あお」といった色語の代わりに，それらの色語と対応させた音韻反復語を用いている点のみが異なり，他の条件は全て同じである．これにより，本実験では音韻反復型の育児語が，学習にどういった影響をあたえるのかを検証した．

4.2 音声認識一致率

学習データ数を増加させたとき，教示発話全体から音声認識された文字列と正解文字列の一致率は図 3 のようになった．このとき，認識文字列 S_r と正解文字列 S_c の一致率 C は式 (11) によって表すことができる．すなわち，C は比較する文字列の編集距離 $\mathrm{LD}(S_r, S_c)$ を文字列長で割り，それを 1 から引いた値として定義される．一致率の値が大きいほど正解に近い音声認識ができていることを意味する．

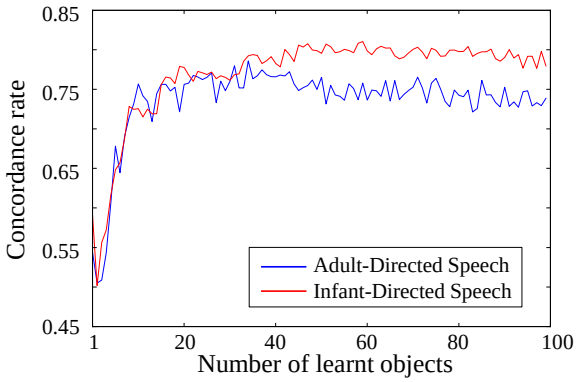

図 3: 学習データ数に対する育児語と成人語を含む教示発話全体の音声認識一致率の推移

$$C = 1 - \frac{\mathrm{LD}(S_r, S_c)}{\max(S_r, S_c)} \tag{11}$$

図 3 において，青線が成人語，赤線が育児語を用いて学習した結果を示す．図 3 から，学習前期には成人語の方が育児語よりも正解に近い音声認識をすることができ，学習後期には育児語の方が成人語よりも正解に近い音声認識をすることができていることが分かる．

次に，学習データ数 n が 1 個，17 個，83 個の場合の音声認識された文字列と正解文字列の一致率のヒストグラムを図 4 に示す．青のバーが成人語，赤のバーが育児語の一致率を表している．図 4(a) では，音声認識された文字列と正解文字列の一致率は成人語も育児語も広く分布している．図 4(b) では，学習が進み，成人語も育児語も図 4(a) に比べて一致率が高い方向へ分布が移動している．このことから，学習が進むにつれて音声を正しく認識できるようになっていることが分かる．図 4(c) では，全体としては図 4(b) と同様に一致率は高い範囲に分布しているように見える．しかし，一致率の値が 0.4 以下の範囲に関しては，図 4(b) では成人語も育児語もわずかではあるが分布しているのに対して，図 4(c) では成人語のみ分布している．これは，学習後期における育児語での教示の場合では，殆どの教示音声を正しく認識できるようになったのに対して，成人語での教示の場合では音声認識を正しく行うことが難しい教示音声が存在するためだと考えられる．つまり，育児語は学習が進むにつれて正解に近い言語モデルを学習する傾向があるのに対して，成人語は正解とは異なる言語モデルを学習する要因を孕んでいると考えられる．このような言語の学習傾向の違いは，概念と言語の相互学習において無視できるものではなく，概念の学習に影響を及ぼすと考えられる．

5. おわりに

本稿では，ロボットの概念と言語の相互学習において教示発話が学習に及ぼす影響の解析を目的として，成人語と育児語を用いた概念形成シミュレーション実験を行った．学習前期には成人語の方が育児語よりも正解に近い音声認識をすることができ，学習後期には育児語の方が成人語よりも正解に近い音声認識をすることができるていることが分かり，明らかな差があることを示した．今後，それぞれの語が概念学習にどのような影響を及ぼすかを検証する．

謝辞

本研究は JST CREST，JSPS 科研費 JP16H02835 の助成を受け実施したものである．

参考文献

[五十嵐 06] 五十嵐 陽介，馬場 れい子，“母親特有の話し方（マザリーズ）は大人の日本語とどう違うか,” 電子情報通信学会技術研究報告，pp. 31-35，2006.

[小林 15] 小林 哲生，奥村 優子，服部 正嗣，“幼児における育児語と成人語の学習しやすさの違いを探る,” NTT 技術ジャーナル，2015.

[村瀬 07] 村瀬 俊樹，“1 歳半の子どもへの育児語の使用傾向と養育者の信念,” 日心第 71 回大会，2007.

[谷口 14] 谷口 忠大，“記号創発ロボティクス 知能のメカニズム入門,” 講談社，2014.

[中村 15] 中村 友昭，長井 隆行，船越 孝太郎，谷口 忠大，岩橋 直人，金子 正秀，“マルチモーダル LDA と NPYLM を用いたロボットによる物体概念と言語モデルの相互学習,” 人工知能学会誌，Vol. 30，No. 3，pp. 498-509，2015.

[船田 16] 船田 美雪，中村 友昭，長井 隆行，金子 正秀，“MLDA と教示なし単語分割に基づく概念と言語モデルの学習過程の解析,” 計測自動制御学会 システム・情報部門学術講演会，GS13-5，2016.

[中村 10] 中村 友昭，長井 隆行，岩橋 直人，“Gibbs Sampling による物体のマルチモーダルカテゴリゼーション,” 第 28 回日本ロボット学会学術講演会，2I1-3，2010.

[持橋 09] Mochihashi, D., Yamada, T., and Ueda, N., “Bayesian unsupervised word segmentation with nested Pitman-Yor language modeling,” Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP，Vol. 1，pp. 100-108，2009.

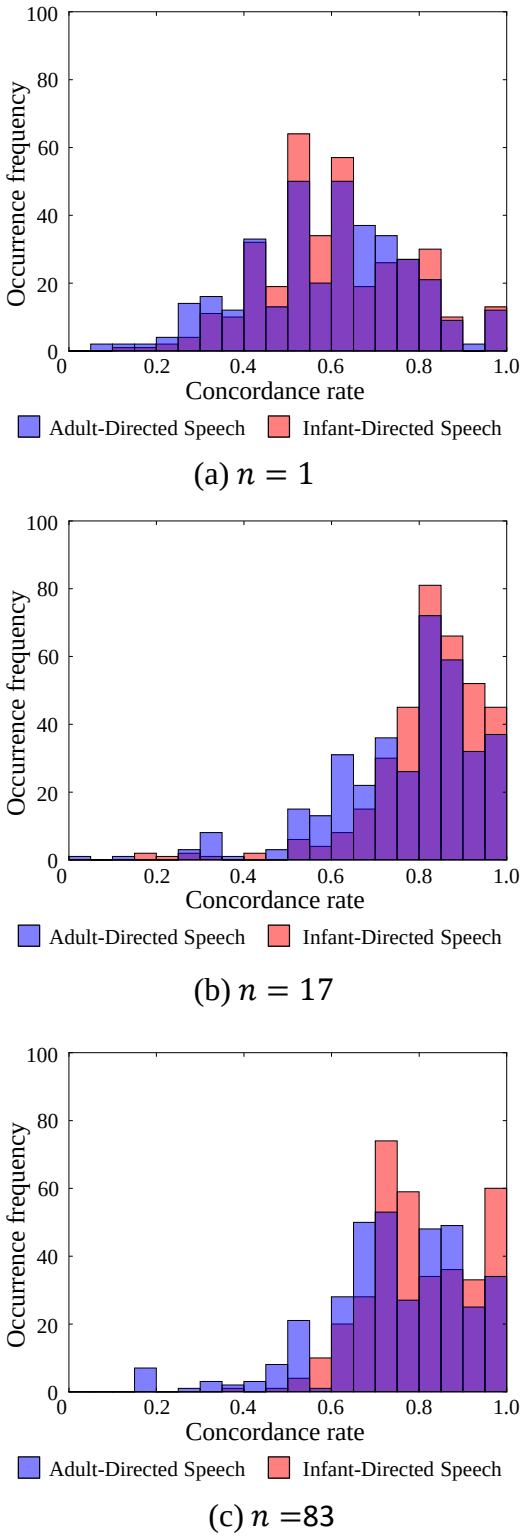

図 4: 育児語と成人語の音声認識一致率のヒストグラム (n は学習データ数)