1O2-5

マルチモーダル概念形成における概念と言語の相互作用の解析

Analysis of Interaction Between Concepts and Language in Multimodal Concept Learning

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

It is considered that concepts are not only formed based on perceptual information in a bottom-up fashion but also affected by language and culture in a top-down fashion. Different concepts depending on each culture can be formed by such interaction between concepts and language. We previously proposed the method by which the robot obtains the concept while concepts and language mutually affect each other. However, we did not analyze the interaction between concepts and language in detail. In this paper, we target a concept formation of colors based on multimodal LDA, which we have proposed, and analyze the interaction between concepts and language. Furthermore, we verify a validity of the proposed method by comparison with a human concept formation.

1. はじめに

人は経験を通して概念を形成し，その概念に単語を接地することで，単語の意味を理解することができる．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．この際，知覚情報によりボトムアップに形成された概念に対して，言語の情報がトップダウンに作用することによって概念が変化する．このように概念と言語は相互に作用し合いながら学習され，最終的に言語や文化によって異なる概念が形成されると考えている．

我々は，これまで概念と言語モデルの相互学習モデルを提案してきた．文献 [中村 15] において，概念と言語が相互に影響し合いながら学習することで，より人間の感覚に近い概念が学習でき，さらに概念の影響を受け言語が同時に学習されることで音声の認識精度が向上することを示した．しかしながら，言語と概念の相互作用に関しては，十分に検証できてはいなかった．

このような相互作用に関しては，人の色概念を対象とした研究が行われている．文献 [今井 15] では，隣接する色同士の関係や境界と色名などの言語が概念形成に密接に関わっていることが指摘されている．人間の子供は 3 歳の時点で既に色の基礎語のほとんどを知っているが，それぞれの色名の境界を理解しておらず，隣接する色に対して色名の過剰汎用をしていることが示されている．さらに，3 歳から 5 歳にかけて色名の範囲の学習は緩やかに進んでいくと同時に，色名の過剰汎用は知覚的な類似に制約され，色の境界が明確になっていく．さらに学習が進むと，言語による制約を受け，色の境界が曖昧になる部分が生じる．このように，概念と言語は相互に影響しあいながら学習される．

また，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．

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

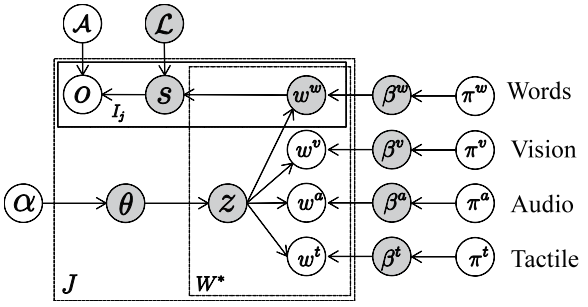

図 1: 物体概念と言語のグラフィカルモデル

本研究では，文献 [中村 15] において提案した物体概念モデルにおいて，どのような相互作用が起きているかを解析することが最終的な目的であるが，モデルの構造が複雑であり，その解析が容易ではない．そこで本稿では，色概念の学習をシミュレーションすることで，我々の提案アルゴリズムでどのような相互作用が起きているかを解析する．色概念は，様々な色と，その色を表現する教示文から学習する．その際，学習データ数と言語の認識精度をシミュレーションにより変化させることで，記号創発システムとして概念の学習と言語の学習がどのように影響しあうかを検証する．さらに，人の概念形成過程と比較することで，我々が提案してきた概念形成モデルの妥当性を検証する．基礎的な解析・検証は文献 [船田 16] で行ったが，本稿ではさらに詳細に解析した結果について報告する．

2. 概念と言語モデルの相互学習

文献 [中村 15] において，我々は言語とが概念が相互の影響しあいながら学習するモデルを提案してきた．図 1 は，言語モデルと物体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，$\mathcal{L}$ をパラメータとする言語モデルにより認識し

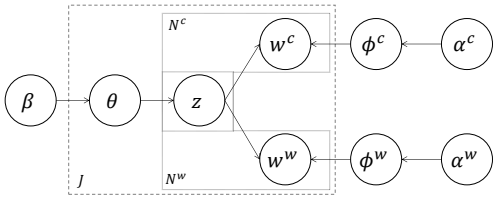

図 2: マルチモーダル LDA のグラフィカルモデル

表 1: 学習データ例

色情報	言語情報
	これは水色です
	これは黄色です
	これは赤です

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^v，w^a，w^t はそれぞれ物体から得られる視覚情報，聴覚情報，触覚情報を示している．

また z は物体のカテゴリを表している．さらに，w^v，w^a，w^t，w^w は，それぞれ β^v，β^a，β^t，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^* 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．J, W^*, I_j は，それぞれ物体数，モダリティ*の情報の生起回数，j 番目の物体への教示発話数を表している．このモデルでは，音声認識結果 s と物体カテゴリ z が単語 w^w によって接続されているため，音声認識と物体概念形成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ことで，音声認識と物体概念が影響しあい，それぞれの精度が向上する学習が可能となった．

最終的には，このモデルで起きている物体概念と言語の相互作用を解析することが目的であるが，このモデルは非常に複雑であり，単純に解析することができない．そこで，本稿では色概念を対象として，シミュレーションによりその相互作用の影響を解析する．

3. マルチモーダル LDA に基づく色概念の形成

本稿では，解析をより容易にするために，文献 [中村 15] に比べてシンプルなモデルであるマルチモーダル LDA(Mulitimodal Latent Dirichlet Allocation)[中村 10] を用いて概念学習を行う．ロボットに対して，色を提示し，同時にその色名を教示することで概念学習を行うことを想定している．ただし，本稿では概念と言語の相互作用を分かりやすくするため，色情報はシミュレーションによって生成したものを使用し，教示はテキストにより入力した．

3.1 マルチモーダル LDA

ロボットが取得した情報と，人から与えられる言語情報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 を用いる．図 2 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，w^w が単語情報となる．これらの情報は，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．

言語と概念の相互作用の解析のため，認識データが各カテゴリに分類される確率と認識データに対する対数尤度を用いる．認識データが各カテゴリに分類される確率とは，学習されたモデルのパラメータを Θ とすると，認識用の色情報 $\boldsymbol{w}^c$ と言語情報 $\boldsymbol{w}^l$ からカテゴリ z が発生する確率 $P(z|\boldsymbol{w}^c, \boldsymbol{w}^l)$ である．対数尤度とは，認識用の全データの色情報 $\boldsymbol{W}^c$ と言語情報 $\boldsymbol{W}^w$ が，Θ をパラメータとするモデルから生成される確率の対数 $\log P(\boldsymbol{W}^c, \boldsymbol{W}^w|\Theta)$ である．マルチモーダル LDA では，この対数尤度は次のように，色情報に対する対数尤度と，言語情報に対する対数尤度に分解することができる．

$$\log P(\boldsymbol{W}^c, \boldsymbol{W}^w|\Theta) = \log P(\boldsymbol{W}^c|\Theta) + \log P(\boldsymbol{W}^w|\Theta) \quad (1)$$

対数尤度は，パラメータを Θ とするモデルがデータをどれだけ表現できているか，を表す指標である．すなわち，$\log P(\boldsymbol{W}^c|\Theta)$ と $\log P(\boldsymbol{W}^w|\Theta)$ を比較することで，モデルにおけるそれぞれの情報の重要度を見ることができる．

3.2 色情報

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色情報として用いた．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．画像は，単色の画像に対してガウスノイズを付与して生成した．

3.3 言語情報

教示された言語に対して形態素解析を行い，単語に分割し，単語の発生頻度ヒストグラムを言語情報として用いる．ただし，形態素解析には，文献 [kudo 04] のアルゴリズムに基づいたオープンソース形態素解析エンジン Mecab[Mecab] を用いた．表 1 が教示文の例である．さらに，音声の誤認識をシミュレーションするため，音声の誤認識率 p を定義し，確率 p に従い別の単語へと置換する．例えば，$p = 0.5$ で “赤　だよ” といった教示文が与えられた場合，50%の確率で単語の置換が発生するため，“緑　だよ” といった文が学習データとして使用される．ただし，この方法は正確には音声の誤認識のシミュレーションにはなっていないため，将来的には実際の音声を用いて実験する予定である．

4. 実験

4.1 データセット

実験には，画素数 320×240 画素のランダムな色の単色画像を用いた．ただし，無彩色は含まない．まず，ランダムに 160 枚の画像を生成し，その画像の特徴を表す言語情報を与え，画像と言語の組を 160 組生成した．実験では，160 組のうち 100 組を学習に，60 組を認識に用いた．

4.2 実験設定

音声の認識率と学習物体数を変化させ，概念と言語が徐々に獲得されていく過程をシミュレーションした．この実験では，学習データ数が増えると共に音声の誤認識率を減少させること

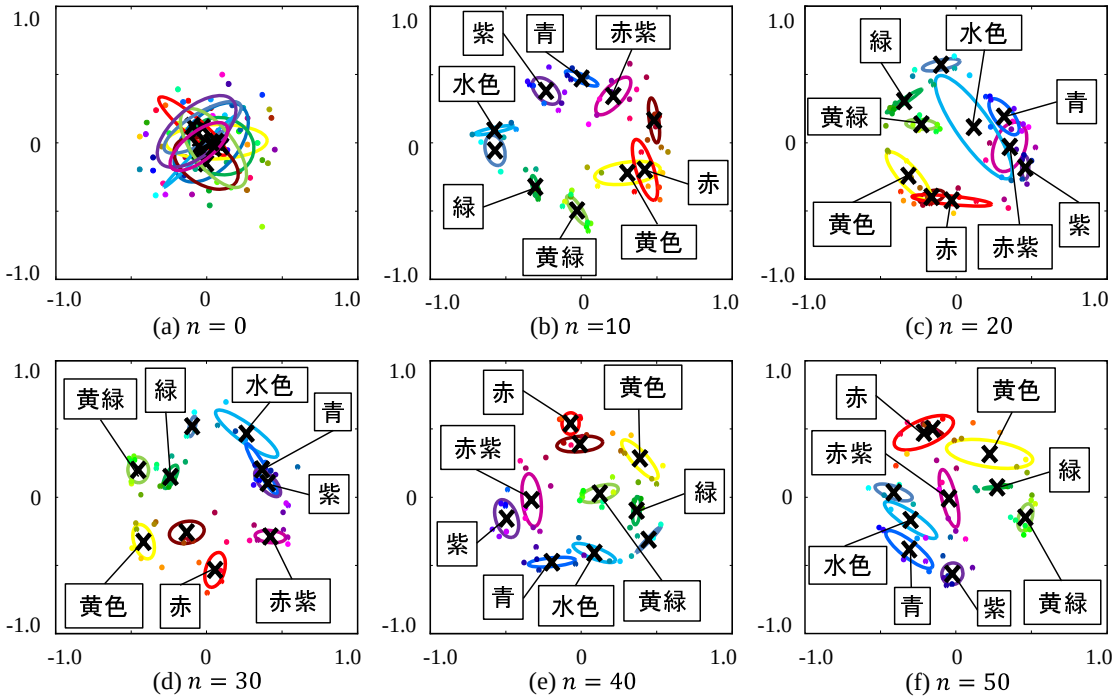

図 5: MDS による色の分類結果の可視化 (n は学習データ数)

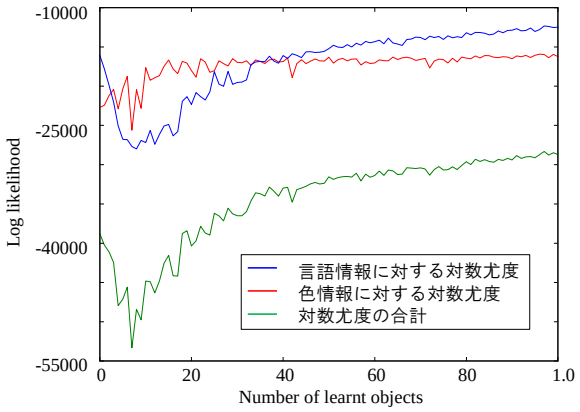

図 3: 学習データ数と音声認識率の向上に伴う対数尤度の変化

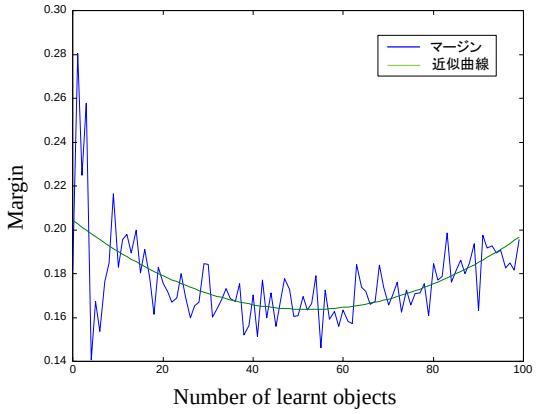

図 4: マージンの変化

で，我々の提案したモデル [中村 15] に近い概念形成過程を再現した．学習物体数 i と誤認識率 p の関係には次式を用いた．

$$p = 1 - \frac{i}{300} \tag{2}$$

学習データ数を 0 個から 100 個まで変化させて学習を行い，各データ数で学習されたモデルを用いて認識データの認識を行った．これを学習データのみを変えて 10 回行った．

4.3 対数尤度の変化

各学習データ数における 10 回の対数尤度の平均を図 3 に示す．この図より，学習初期に言語情報に対する対数尤度が急激に減少している．これは音声の誤認識が多いため，色情報優位な概念が形成されたためだといえる．その後，徐々に言語情報の尤度が上昇し，物体数が 40 付近において，色情報に対する対数尤度が言語情報の尤度を上回っている．この結果は，文献 [今井 15] によって示された発達とともに，色情報優位な概念から，言語情報優位な概念に変化するといった内容と一致している．

次に，認識用データが各カテゴリに分類される確率のマージンを計算した．マージンは，j 番目の認識用の色情報 $\boldsymbol{w}_j^c$ と言語情報 $\boldsymbol{w}_j^w$ が与えられた際の，確率 $P(z|\boldsymbol{w}_j^c, \boldsymbol{w}_j^w)$ の最大値と二番目に大きい値を引いた差である．

$$margin_j = P(z_{\max}|\boldsymbol{w}_j^c, \boldsymbol{w}_j^w) - P(z_{2\mathrm{nd}}|\boldsymbol{w}_j^c, \boldsymbol{w}_j^w) \tag{3}$$

$$z_{\max} = \operatorname{argmax} P(z|\boldsymbol{w}_j^c, \boldsymbol{w}_j^w) \tag{4}$$

$$z_{2\mathrm{nd}} = \operatorname*{argmax}_{z \neq z_{\max}} P(z|\boldsymbol{w}_j^c, \boldsymbol{w}_j^w) \tag{5}$$

すなわち，このマージンは確率が最大となるカテゴリ $z_{\max}$ へ，認識用データを分類する確信度と見なすことができる．各学習ステップ毎の全物体のマージンの平均が図 4 である．学習初期では過学習のためマージは大きくなっている．その後学習が進むにつれて，マージンは下がり 40 を過ぎたあたりでもっとも小さな値となっている．学習前半では色情報を優先して学習しており，モデルが言語情報を正しく表現出来ていないために，言語が聞き取れるようになるにつれて分類の曖昧さが増加しているためだと考えられる．その後，40 を過ぎると言語を重視する分類となり，言語情報を正しく表現できるモデルへと変化するために，分類の曖昧さが減少すると考えられる．

4.4 学習結果の可視化

次に，学習データ数 n が 0 個，10 個，20 個，30 個，40 個の場合の分類を可視化した結果を図 5 に示す．可視化には多次元尺度構成法 (MDS，Multi Dimensional Scaling)[Torgerson 58] を用いて，式 (6) によって表される非類似度の関係が可能な限り維持できるよう，各認識データを二次元空間上に表示した．

$$D_n(i,j) = 1 - \sum_{k}^{K} \min(P(k|\boldsymbol{w}_i^c, \boldsymbol{w}_i^w, \Theta_n), P(k|\boldsymbol{w}_j^c, \boldsymbol{w}_j^w, \Theta_n)) \tag{6}$$

ただし，$D_n(i,j)$ は n 個のデータで学習したモデル Θ_n において，i 番目と j 番目の認識データを認識した際の主観的な色の距離を表しており，距離が近いほど小さな値となる．また，右辺の第 2 項はインターセクション [Swain 91] であり，K はカテゴリ数，$\boldsymbol{w}_i^c$ と $\boldsymbol{w}_j^c$ は i 番目と j 番目の認識データの色情報，$\boldsymbol{w}_i^w$ と $\boldsymbol{w}_j^w$ は i 番目と j 番目の認識データの言語情報を表している．

図中の各プロットの色は， 認識データの色である．また，X は正解カテゴリの平均位置，色付きの楕円は正解カテゴリの分散を表している．この 2 次元空間では，縦軸と横軸には意味はなく，プロット間の距離が学習されたモデルによって認識した際の主観的な非類似度を表している．すなわち，この空間においてプロット間の距離が大きく広がっている場合には明確に分類できていることを意味し，距離が小さく中心付近に集まっている場合には明確な分類ができていないことを意味している．図 5(a) では色は正しく分離できておらず，中心付近に集中していることが分かる．図 5(b) では，プロットは広がり円状に配置されている．この空間上でプロットが広がることは，各カテゴリ間の非類似度が増加し，より明確な分類ができていることを意味している．しかし，それぞれの正解カテゴリの平均に注目すると，赤と黄色に相当するカテゴリの平均は近く，それぞれのカテゴリの分散を示す楕円同士が重なっていることが分かる．また，図 5(c) の水色と青と紫と赤紫，図 5(d) の青と紫についても同様の現象が起きている．つまりこれは，人間の子供が色名の過剰汎用をする現象に相当すると考えられる．さらに，図 5(e) で図 5(f) では，再びプロット間の類似度が大きくなり中心付近に集まっている．すなわち，カテゴリ間の類似性が大きくなり，明確な分類ができなくなっていることを意味している．しかし，それぞれの正解カテゴリの平均に注目すると，図 5(b)，図 5(c)，図 5(d) においてそれぞれ近くに位置していたカテゴリ間の距離が離れ，正解に近い分類がなされたことが分かる．また，カテゴリの分散くを示す楕円同士の重なりが少なくなったことからも言語情報により色の境界が変化したことが分かる．

以上より，提案モデルにおいて言語の学習とともに，言語情報が概念形成のための情報として重視されることが分かる．さらに，言語情報優位な学習へと変わることで，色の境界が変化し，色概念は形成されていく．そして，最終的には，色情報優位な概念とは異なる，複雑な概念が形成されると考えられる．この結果も，文献 [今井 15] で示された結果と一致しており，人間の学習における概念と言語の相互作用が我々のモデルにおいても再現できていることを示唆している．さらにこれは，文献 [谷口 14] における記号創発システムとして捉えることができる．学習初期には自らの知覚情報に基づいた概念が形成される．その後，他者とのインタラクションを通じて形成される共通の記号系である言語が，個の概念に制約を与え概念が変容する．本実験においては，学習データ数が増加する毎に，音声を聞き取る能力が上がることをシミュレーションすることで，そのような現象を再現することができた．

5. おわりに

本稿では，マルチモーダル LDA を用いて，色情報と言語による教示から色の概念形成実験を行った．学習物体数と言語情報の誤認識率を変化させることで，色情報優位な学習から言語情報優位な学習への推移を示した．実験結果より，音声が正しく認識できない学習初期では，色概念は色情報優位な学習によって形成されている．つまり，文献 [今井 15] における 5 歳の色情報優位な学習過程に相当していると考えられ，色語によって指し示される色の境界が曖昧なため，色語の過剰汎用を伴うと考えられる．そして，言語の発達により言語情報が概念形成に作用することで，それまでとは異なる境界を持つ概念が形成される．このように，言語と概念を同時に学習しながら概念形成することで，人間がもつ複雑な概念が形成されると考えられる．

今後の課題として，まず実際の画像と音声を使って実験することが挙げられる．さらに，色概念だけではなく，我々がこれまで提案してきたマルチモーダル情報を用いた物体概念形成においても，概念と言語間の相互作用の解析を行う．また，概念の学習には，単に音声認識精度だけが関わっているのではなく，様々な要因が影響していると考えられる．今後，そのような要因を調査し，我々のモデルにおける影響について検証する予定である．

謝辞

本研究は JST CREST の助成を受け実施したものである．

参考文献

[今井 15] 今井 むつみ，佐治 伸郎，浅野 倫子，大石 みどり，岡田 浩之，“記号の意味はシステムの中で生まれる，” 2015 年度人工知能学会全国大会, 2D3-OS-12b-2，2015.

[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，“マルチモーダル LDA に基づく概念学習における概念と言語の相互作用の解析”，メディア工学研究会 学生研究発表会，pp, 13-16，2016.

[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.

[kudo 04] Taku Kudo，Kaoru Yamamoto and Yuji Matsumoto，“Appliying Conditional Random Fields to Japanese Morphological Analysis，” EMNLP，Vol.4, pp.230–237，2004.

[Mecab] “MeCab: Yet Another Part-of-Speech and Morphological Analyzer，” http://taku910.github.io/mecab/

[Torgerson 58] Torgerson and Warren S，“Theory and methods of scaling，” New York，John Wiley and Sons, Inc., p.460，1958.

[Swain 91] M. J，Swain and D. H. Ballard，“Color indexing，” International Journal of Computer Vision，Vol.7，pp.11–32，1991.