

状況を分解する学習モデルによる言語の獲得と分類

Word acquisition and classification with situational decomposition learning models

永野秀明*¹ 山川宏*² 荒川直哉*² 岡田浩之*¹
 Hideaki Nagano Hiroshi Yamakawa Naoya Arakawa Hiroyuki Okada

*¹ 玉川大学
 Tamagawa University, Graduate school of Brain Science

*² ドワンゴ 人工知能研究所
 Dwango Artificial Intelligence Laboratory

The symbol grounding problem would require conceptual representation to be ready for the grounding of words. As word classification depends on the native language and culture, the analysis of features underlying word classification could lead to the understanding of the cultural dependency. In this study, we chose to analyze action verbs meaning “carry” and “hold” in three languages. Verbs with static features abstracted in advance were analyzed with conceptual clustering. The results indicate kinetic and temporal features need be analyzed by means of methods such as situational decomposition to simulate verb acquisition by human beings.

1. はじめに

言語獲得において、名詞は事物と対応付けられる単語レベルであるのに対し、動詞は行為を示すことから、動詞を獲得するためには対象となる行為を前後で連続している一連の動作のなかから抽出し、対応付ける必要がある。杉浦は、ユーザがロボットに動作を教示する際の軌道を汎化する手法を開発し、その有効性を示した [杉浦 12]。また谷口らは、単位となり得る動作を抽出するという分節化処理とその前後の非分節な動作時系列のモデリングとの二つの学習プロセスを考え、両者を統合することの必要性を指摘している [谷口 12]。

一方、幼児の言語獲得に関して佐治らは、動詞の獲得において中国語児と日本語児は異なる概念空間を形成していると指摘している [佐治 08]。例えば中国語の場合は「体のどこでモノを持っているか」に対する重み付けが年齢とともに増加するのに対し、日本語の場合は、低年齢ほど「する」「やる」で表される軽動詞の使用が多く、年齢が進むほど「持つ」や更に具体的な動詞へと移っていく。このように、実際の間人は母語や文化によって異なる特徴量に着目することで異なる分類をするが、特に動詞の場合はその動きの様態にも依存して異なる分類をする場合があることから、時間的な特徴量に着目することが適切な概念空間の形成のためには重要であると考えられる。そして、それら動詞の分類に対する分析を通じて、動詞の背景にある多元的な概念空間の性質を知りうると期待できる。本稿では、母語によって異なる「運ぶ」系の動詞に対応付けられる動作の特徴量を事前に抽象化し、概念クラスタリングを用いて分類することで、動詞が接地する概念空間についての分析を行う。

“lin” のように使い分けしているが、その使い分けもまた韓国語、日本語とは異なる。今回は、佐治らが用いた 13 の中国語動詞を分析の対象とする。その動作様態と、韓国語、日本語の分類を図 1 に示す。

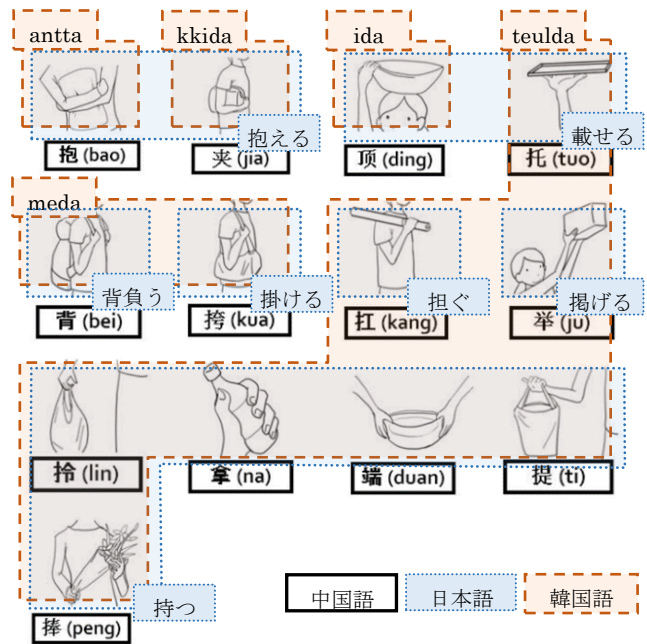


図 1 「運ぶ」系動詞の動作様態
 ([佐治 11]より一部改変)

2. 動詞の分類

2.1 文化による「運ぶ」系動作動詞の違い

英語では「モノを運ぶ」動作は「移動しているかどうか」によってのみ区別され、移動している場合は“carry”であり、移動せずにただ持っているだけの場合は“hold”である。一方、中国語や韓国語、日本語では移動の是非は区別せず、持ち方の様態で使い分けしている。例えば中国語において頭の上に対象物を載せる場合を“ding”，手からぶら下げるようにして持つ場合を

表 1 特徴量

対象物の性質	対象物の大きさ
	対象物の細長さ
	対象物のかたさ
	対象物の際立ち*
	対象物の種類 (容器, かばん, どちらでもない)*
体との関係	接触部位*
	体との高さ関係
	支持の仕方 (下から支える / 上からぶらさげる)

(* 印のあるものは佐治による)

連絡先: 永野秀明, 玉川大学大学院脳科学研究科, 東京都町田市玉川学園 6-1-1, Tel: 042-739-8326, E-mail: ngnh4re@engs.tamagawa.ac.jp

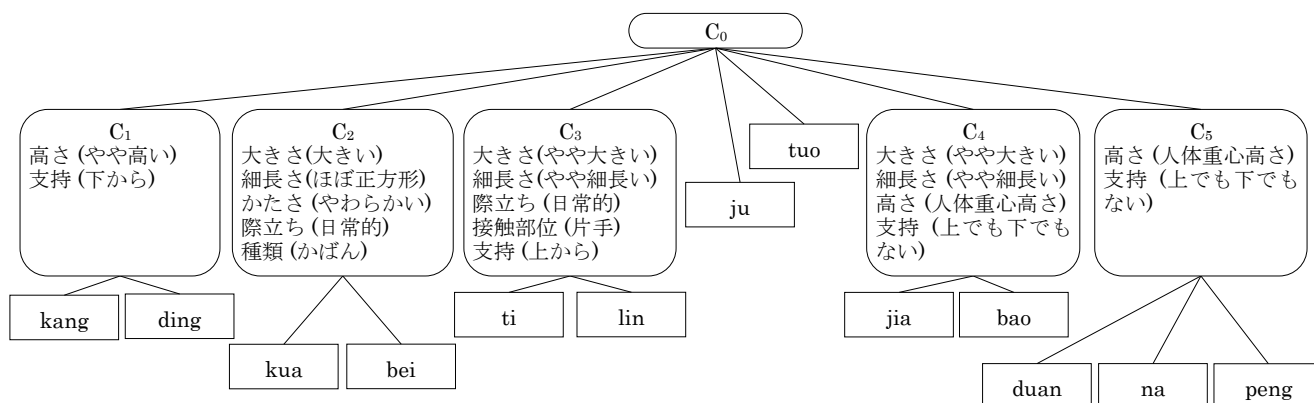


図 2 形成された概念構造

2.2 特徴量の抽出と抽象化

本稿では、13 の中国語の様態から分類の判断基準となると考えられる 8 の特徴量を事前に抽出し、抽象化したデータセットを入力データとして用いた。抽象化したデータセットを表 1 に示す。

3. シミュレーションと結果

3.1 概念クラスタリングによる分析

本稿では解析ツールである Weka [Weka 15] に実装されている概念クラスタリング手法である COBWEB を用いた。分析の結果生成された概念の木構造を図 2 に示す。クラスタ C_4 には、両手で抱くように持つ “bao” と脇で挟んで持つ “jia” が属しており、このクラスタは日本語における「抱える」に対応しているといえる。日本語における「抱える」は、「腕によって対象物を囲むようにして持つこと」を指すが、そのような様態を端的に表す特徴量は今回抽象化したデータセットにはない。一方、体と対象物との高さ関係や、対象物の支持の仕方といった特徴量が “bao” と “jia” において等しいことから、間接的に「抱える」に相当するクラスタが形成されたと考えられる。一方、手で握るように持つ “na” および両手で大事そうに持つ “peng” もそれらの特徴量において “bao” および “jia” と等しいが、“na” においては対象物のかたさが大きく異なり、“peng” においては体と対象物の接触部位の違いが差別化に寄与したと考えられる。実際、“peng” における接触部位を “bao” と同じ「胸」として分析したところ、“bao” や “jia” と同一のクラスタに分類された。また、肩から下げて持つ “kua” と背中に背負う “bei” は、韓国語においてはともに “meda” と表せられ、これはクラスタ C_2 に対応している。

一方、手のひらの上で水平に持つ “tuo” と頭の上に載せて持つ “ding” はともに日本語では「載せる」と表されるが、この 2 つは同一クラスタに分類されなかった。

3.2 着目する特徴量と状況分解手法の有効性

以上のように、言語や動作動詞の一部では実際の人間と同様のクラスタが形成されたものの、十分に人間の動詞分類を再現できていない。そこで、8 の特徴量のうち、接触部位と支持の仕方だけに限定して分析を行ったところ、“din” と “tuo” の 2 つがひとつのクラスタに分類され、日本語の「載せる」に対応付けられる結果となった。このように、着目する特徴量によって形成される概念構造が異なるのは当然であるが、このことは実際の人間も文化や母語によって着目する特徴量が異なることを示している。

今回の分析では、動作を表す図(図 1)から抽出した静的な特徴量のみに着目し、同じく事前に想定した動詞との対応付けに着目することで動詞の獲得とその概念構造について検証した。しかし実際の人間の動作は動的なものであり、したがってその特徴量も時系列に沿って変化する。また、ひとつの文化・母語内においても、その状況によって異なる特徴量に着目する場合がある。例えば中国語において “tuo” はお盆のような平たいものを手のひらの上で水平に持つことを表すため、対象物の形状と角度は無視できないが、手で握るように持つ “na” においては対象物の角度は分類を左右しない。また、今回は日本語で「掲げる」とした動作も、その前後の流れによっては高いところから低いところへ対象物を「降ろす」であるかもしれない、動作の持つそのような時間的な側面を考慮に入れ、都度適切な特徴量を選択し、分析する手法こそが動詞の獲得のモデル化には必要だと考えられる[山川 04]。

4. おわりに

本稿では中国語・韓国語・日本語における「運ぶ」系動詞の分類と構造を、概念クラスタリングを用いて分析した。今後は、動的な特徴量を考慮しその選択が可能な手法である状況分解手法を用いて、よりプリミティブなデータからの動詞の獲得を再現することを目指す。

参考文献

- [杉浦 12] 杉浦孔明: ロボット対話: 実世界情報を用いたコミュニケーションの学習, 人工知能学会誌 27 (6), 人工知能学会, 2012 年.
- [谷口 12] 谷口忠大: 二重分節構造に基づく教師なし学習と記号創発: 時系列情報の分節化による記号創発ロボティクスへのアプローチ, 人工知能学会誌 27 (6), 人工知能学会, 2012 年.
- [佐治 08] 佐治伸郎, 今井むつみ: 語彙獲得における動詞の使い分けに関する研究: 中国語の「持つ」系動詞を事例として, 日本認知言語学会論文集第 8 巻, 2008 年.
- [佐治 11] 佐治伸郎: 母語及び第二言語の習得における語意の再編成過程に関する研究, 慶應義塾大学博士論文, 2011 年.
- [Weka 15] Weka, <http://www.cs.waikato.ac.nz/ml/weka/>, 2015 年
- [山川 04] 山川宏, 馬場孝之, 岡田浩之: ETMIC 基準を用いた状況分解によるカード分類課題での概念獲得と予測過程, 認知科学 11(2), 日本認知科学会, 2004 年.