

言語変化と共通語彙の形成を両立する能力： 一般化学習の優先と同音異義語生成の抑止

Compatibility of Language Changes and Shared Vocabulary: The role of Priority of Generalization and Preventing Production of Homonyms

金野 武司*¹

Takeshi Konno

橋本 敬*¹

Takashi Hashimoto

*¹北陸先端科学技術大学院大学 知識科学研究科

School of Knowledge Science, Japan Advanced Institute of Science and Technology

Humans communicate with others using common vocabularies. And the vocabularies should change dynamically in the cultural evolutionary process. If the language evolved in response to the need to communicate using the common vocabularies, the acquisition of the different expressions impedes advantages of the communication. How the language changes and the common vocabularies can be compatible? We study the mechanisms of compatibility using a computational model for the evolution of compositionality by Kirby (2002) with slight extension. As a result, we suggest that at least two abilities have an important role for sharing vocabulary in learning compositional language. One is to give priority to generalization over the persistence of parent's expressions, and the other is to prevent the production of homonyms.

1. はじめに

言語は意思疎通の道具として機能しながらも変化し続けている [Steels 07]. この特徴は一見すると矛盾している。例えば、意思疎通のためには共通の語彙が必要だと考えられるが、その語彙が変化してしまうことを意味しているからである。人工知能の分野では意思疎通の道具となる言語の実装が盛んに取り組まれているが、これを進めるためには、変化し続ける言語という特徴のメカニズム的な理解が重要だと考えられる [橋本 99].

言語はなぜ意思疎通の道具として機能しながらも変化し続けるのだろうか。本論では文化進化的なスケールで変化する語彙に集中してこの問題を考える。今簡単に、言語話者は状況から何らかの内部表現を形成し、その内部表現に対して発話を割り当てるシステムを持つ者と捉えてみる。このとき、内部表現の形成を認識と呼び、内部表現に発話を割り当てる規則の集合を文法と呼ぶ (この文法の中に語彙表現の規則が含まれる)。すると語彙の変化は、認識の変化を起因とするメカニズムと、文法そのものが変化するメカニズムの2つで考えることができるようになる。

本論では、後者のメカニズムに関して計算モデルを用いて検討する*¹。このとき、文法の変化過程として本論が扱うのは合成的な言語の形成過程である。ここで合成的な言語とは、単語の意味の組み合わせによって文の意味が決まるような形式を持った言語のことであり、ヒトの言語は広くこの特徴を備えている [Smith 06]。以下では、言語話者が同じ認識を持った状態でも、合成的な言語の形成過程では語彙が変化することを示しつつ、語彙の共通化が言語話者のどのような能力によって実現されるのかを議論する。

2. 計算モデル

本論の計算モデルでは、認識の変化を起因とする語彙変化のメカニズムを分離するため、状況の認識はいつでも一致す

連絡先: 〒 923-1292 石川県能美市旭台 1-1,

{t-konno,hash}@jaist.ac.jp

*¹ 認識の変化を起因とする語彙変化のメカニズムも扱うべき重要な問題だが、まずは2つのメカニズムを切り分けて後者のメカニズムを検討する。

るものと仮定する。このため、言語の伝達は一方的に行なわれるものとする。このような一方の言語伝達過程を、親から子への言語継承としてモデル化するものに ILM (Iterated Learning Model) がある。本論はこの ILM において、言語伝達が世代を超えて文化進化的に行なわれた結果として合成的な言語を形成することのできる学習メカニズムを提示した先行研究のモデル [Kirby 02] を用いる。

Kirby の計算モデルでは、親子は同じ状況を見て同じ内部表現を構成する。親は自分の文法を参照して、その内部表現に対する発話を生成する。子は状況に対する内部表現と、親の発話を一つの規則 (これを文ルールと呼ぶ) にして自分の文法に加える。このとき、子は文ルールが加えられるたびに次のような2つのルール操作 (学習) を行なう。i) 2つの文ルールの類似性を使って単語を切り出す。例えば、「赤いリンゴ」と「赤い車」を見聞きした際のルールを比較して、文ルールを「赤い x 」 (x は一般化要素で、リンゴや車を当てはめることができる) と一般化し、さらにリンゴと車という単語を切り出す (これを単語ルールと呼ぶ)。このとき、2つの単語ルールは同じカテゴリになる。ii) 切り出した単語を別の文ルールに適用する。例えば「丸いリンゴ」を見聞きすると、「リンゴ」に相当する単語ルールを使って「丸い x 」と一般化する。この操作によって、リンゴの単語ルールと同じカテゴリにある車の単語ルールを「丸い x 」に適用して「丸い車」という発話を作ることができるようになる。つまり、子は「丸い車」という状況を見聞きしていないのにも関わらずそれを表現するための文法を獲得する。

3. 合成的な言語の形成と語彙の変化

親から子への言語伝達を一代とし、そこで親子が体験する状況の数 (T) を設定して言語伝達を繰り返す。環境には全部で100個の状況を用意する。すると、世代を経るにつれてそれぞれのエージェントが獲得する文法は合成性を高めることになる。合成性が高まると、用意された全ての状況を少数のルールで発話することができるようになる。これは Kirby [Kirby 02] の結果と同様である。

では、この合成的な言語の形成過程で語彙はどの程度変化しているだろうか。これを調べたのが図1である。これは、状

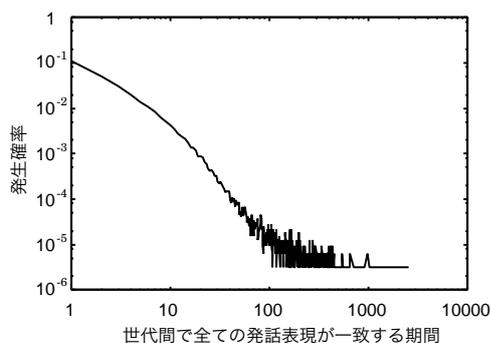


図 1: 語彙不変期間の確率分布

況に対する発話の集合が全く同じに維持される世代数を横軸にとったとき、その世代数がどの程度の頻度で発生するのかを表したグラフである。このグラフを見ると、例えば 1000 世代といった長い世代に渡って語彙の変化が起こらない場合がありつつも、多くは数世代でどんどん変化することが分かる。

我々は、このような語彙の変化を減らす可能性があると考えられる 2 つの仕組みをテストした。2 つの仕組みは共に子の学習方法に関するものである。1 つは親から聞いた発話表現を維持する仕組み (聞いた親の表現を自分の子にそのまま伝える仕組み) であり、もう 1 つは単語の切り出しにおいて発生する同音異義の単語ルールを生成しないようにする仕組みである。この 2 つの仕組みをそれぞれに実験した結果が図 2 である。これは、一世代で親子が体験する状況の数 (T) を横軸にとったときに、それぞれの状況に対して親子が同じ発話表現を用いる割合を全ての世代に渡って平均した数値を示したものである。

この図から読み取るのは次の 3 つの結果である。1. 親から聞いた発話表現を維持すると、予想に反して発話表現の一致度が低下する*2。2. 同音異義の単語ルールを生成しないようにすると発話表現の一致度が向上する。3. 全ての状況に対する発話を聞いてしまうと発話表現の一致度が低下する。

この結果は、エージェントが文法を一般化する観点から理解できる。親から聞いた発話表現を維持する仕組みは、一見すれば語彙の変化が減るのではないかと思える。ところが、実際は一般化しにくい表現が残りに残ってしまうため、子の学習において親と異なる表現を学習する機会が減らなくなり、結果的に発話表現の不一致が起こり続けることになる。これと同じ理由で、全ての状況に対する発話を親から聞いてしまうと、一般化を進めることのできない表現が残りに残って、発話表現が変化し続けることになる。また、同音異義の単語ルールを作らないようにすると、発話表現の一般化方向を限定することになるために発話表現の一致度が高くなる。

合成的な言語を形成する過程においては、一般化を進めやすい表現へと積極的に文法を変化させる方が、結果的に語彙の変化を減らすことができる。この結果は、一般化が進めば同じ学習規則を持つ者は誰でも同じ語彙表現を持った文法を獲得するようになることを意味している。これは、文法の一般化 (特に構文的規則の汎化) が、語彙の変化を引き起こしつつも集団内で共通の語彙を持った言語を形成する役割を担っていることを示唆するものである。

*2 $T=100$ で発話表現の一致率が 100% 近くになっているのは、親の発話表現を維持しながら全ての状況に対する発話を聞いているからである。このため、世代間の発話表現の不一致が起こらなくなるのは当然である。紙面の関係で提示はしないが、この状態で学習を行なったときには一般化はほとんど進まずにルール数はまったく減らなくなる。

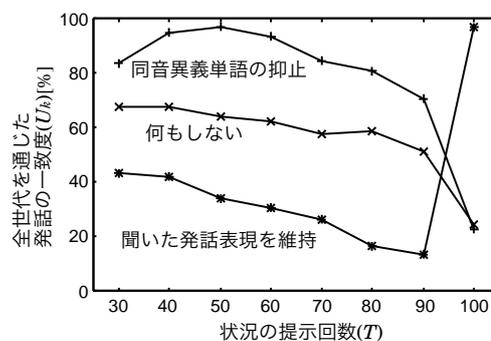


図 2: 状況提示回数の変化に対する世代間での発話の一致度

4. 結論

語彙が変化する原因を人の学習能力に求める場合、状況に対する認識の変化が先行して、それが語彙表現の変化となって現れると考える場合と、文法の変化が先行して、それが結果的に認識を変化させると考える場合の 2 通りがある。本論では特に後者のメカニズムを検討した。合成的な言語の形成過程では、表現できる状況を増やすための文法の汎化 (切り出された単語ルールの他の文ルールへの拡大適用) によって語彙の変化が起こる。これは、実際の言語でも状況認識 (内部表現) の変化が先行せずに語彙の変化が起こり得ることを示唆している。

ではなぜ合成的な言語を形成するのだろうか。本研究の結果から示唆されるのは、合成的な言語の形成によって文法が一般化すると、同じ学習の仕組みを持った者は共通の語彙を持った文法を獲得することができるようになるからである。複数の人間が意思疎通の道具として言語を使おうとするときには、他者の表現をいつまでも維持し続けるより、構文的規則を汎化しやすい方向へ語彙を積極的に変化させた方が、結果的に共通の語彙を持った言語を形成できるのではないかと考えられる。また、その一般化の際には、同音異義語の生成を抑制する能力を持つことが重要である。なぜなら、単語を切り出す際に発生する同音異義語の生成を抑制することで、状況に対する発話表現の曖昧性を低下させ、語彙の共通性を向上させることができるからである。

参考文献

- [橋本 99] 橋本 敬: 動的言語観に基づいた単語間関係のダイナミクス, 認知科学, Vol. 6, No. 1, pp. 55-65 (1999)
- [Kirby 02] Kirby, S.: Learning, bottlenecks and the evolution of recursive syntax, in Briscoe, E. ed., *Linguistic evolution through language acquisition*, pp. 173-203, Cambridge University Press (2002)
- [Smith 06] Smith, A. D. M.: Semantic reconstructibility and the complexification of language, in *Proceedings of the 6th International Conference on the Evolution of Language*, pp. 307-314 (2006)
- [Steels 07] Steels, L.: Language Originated in Social Brains, in Vilarroya, O. and Argimon, F. eds., *Social Brain Matters: Stances of Neurobiology of Social Cognition*, chapter 9, pp. 223-242, Editions Rodopi, Amsterdam (2007)