JAIST Repository
https://dspace.jaist.ac.jp/

Title	言語進化論的アプローチによる文法形成過程のモデル化
Author(s)	中村，誠
Citation	科学研究費補助金研究成果報告書: 1-5
Issue Date	2011-06-13
Type	Research Paper
Text version	publisher
URL	http://hdl.handle.net/10119/9793
Rights	
Description	若手研究（B），研究期間：2008～2010，課題番号：20700239，研究者番号：50377438，研究分野：人工知能，自然言語処理，科研費の分科・細目：情報学・認知科学

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

様式 C-19

科学研究費補助金研究成果報告書

平成２３年　６月１３日現在

機関番号：１３３０２
研究種目：若手研究（Ｂ）
研究期間：2008～ 2010
課題番号：２０７００２３９
研究課題名（和文）　言語進化論的アプローチによる文法形成過程のモデル化

研究課題名（英文）　Evolutionary linguistic approach to modeling the process of grammatical formulation

研究代表者
中村 誠　（ Makoto Nakamura ）
北陸先端科学技術大学院大学・情報科学研究科・助教
研究者番号：５０３７７４３８

研究成果の概要（和文）：本研究は，自然言語の文法が言語使用者間のコミュニケーションによって動的に形成される過程をマルチエージェントモデルによって再現し，学習環境と構文発達の関連を示す定量的なモデルの構築を目的とする．過去の研究成果を元に，通時的および，共時的な言語変化に着目したモデルを提案した．実験により，言語獲得を行うエージェントの学習能力および，それをとりまく学習環境がコミュニティ全体の言語変化に大きく影響を及ぼすことを示した．

研究成果の概要（英文）：This research aims to build multi-agent models which enable us to measure correlation between grammatical development of language and its learning environment by simulating the dynamic process of grammar formulation through communication between language learners. Based on the pre-existing models, some models focusing on synchronicity and diachronicity are proposed. Experimental results showed that languages spoken by the whole community are changed depending on the faculty of language acquisition of agents and its language environment.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
２００８年度	900,000	270,000	1,170,000
２００９年度	700,000	210,000	910,000
２０１０年度	800,000	240,000	1,040,000
年度			
年度			
総　計	2,400,000	720,000	3,120,000

研究分野：人工知能，自然言語処理
科研費の分科・細目：情報学・認知科学
キーワード：進化言語学，エージェント，モデル化

１．研究開始当初の背景

言語の起源と進化に関する研究は，古くから行われており，特に近年の計算機の向上により，シミュレーションによる仮説の検証が可能となった．これまでの文法進化研究として，世代間の継承による通時的な構文構造の学習モデル(Kirby,2001)が挙げられる．これにより，構文形成に対する言語獲得期の「刺激の貧困」(Chomsky,1980)の必要性を論ずるなど，有益な成果を得ている．しかしこのモデルは，語形変化や格助詞の付与を考慮していないために構文の曖昧性を解消できず，自然言語の文法発達を論じる上で不十分といえる．例えば，「私，彼女，好き」という格助詞が欠落した文からは，主語と目的語が明確に決まらず，正常な意味獲得を行うことは困難である．

また，筆者は，言語接触による言語変化，

特にピジン，クレオールのモデリングに関する研究を行っていた．ピジンとは異言語コミュニティ間の接触によって現れる一時的な言語であり，それが母語化したものをクレオールと呼ぶ．これらの言語変化現象に対して，マルチエージェントや数理モデルによるシミュレーションにより，ピジンやクレオールが発生するためのさまざまな条件を導出することに成功していた．しかしながら，モデルがあまりに抽象的であるため，実際の言語変化現象との比較が困難であり，より具体的なモデルの提案を必要としていた．

２．研究の目的

本研究は，自然言語の文法が言語使用者間のコミュニケーションによって動的に形成される過程をマルチエージェントモデルによって再現し，学習環境と構文発達の関連を示す定量的なモデルの構築を目的とする．人間の文法獲得過程を模倣したモデルは，これまでにいくつか提案されているが，本研究においては，特に学習環境や人間の学習能力の設定に着目する．具体的には，言語学習者の他の言語話者との接触状況や，人間に特有とされている認知バイアスを学習機構に導入することで，より現実的なモデルの提案を行い，その検証を行う．また，コーパスを分析することで，実際の通時的，共時的，およびドメインにおける文法の違いについて考察を行う．

３．研究の方法

研究方法としては，大きく２つに分けられる．(1)言語学的調査：主に既存コーパスと言語学書籍から現象分析を行う．ここから，現存言語の文法要素と構文理解との関係を調査する．(2)マルチエージェントモデルの構築とその評価：エージェントが獲得した文法による発話と共通文法の学習をモデル化し，その結果を発話環境と比較することでモデルを評価する．なお，自然言語を処理する能力を持つパーサや文法構造を多数のエージェントに搭載することは実用的ではないことが予備実験の段階で判明したため，これらの実験はそれぞれのモデルで独立して行われることとなった．

マルチエージェントモデルに関しては，文法の共時的変化および通時的変化について，それぞれモデルを構築する．以下にその詳細を述べる．

言語の共時的変化モデル：これまでの言語動力学では，全学習者が全員と同様に相互作用するモデルになっていた．ここで，空間構造を導入し，個々の学習者はローカルに相互作用するようにする．すなわち，その空間内で発話を行うエージェントが配置され，各エージェントの子供は，同じ場所に配置された親と近隣のエージェントに影響を受けながら学習を行う．エージェントは，初期条件として既存の２言語のどちらかを持っていると仮定し，言語の類似性などを考慮した確率分布に従って次世代の言語を決定する．エージェントを結ぶネットワークの形態としては，格子状ネットワークおよび，複雑ネットワークの一種として知られる Barabasi-Albert ネットワークを導入し，既存モデルである数理モデルと比較を行う．

言語の通時的変化モデル：Kirby による繰り返し学習モデル（Iterated Learning Model）(Kirby,2001)をベースとして用いる．これは図１に示すように，基本的な子供の言語発達の能力を仮定し，何世代にもわたる繰り返し学習を行うというモデルである．

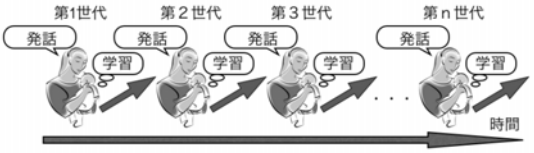

図１：繰り返し学習モデル

本研究においては，認知バイアスの一種である対称性バイアスをエージェントに仮定し，文法獲得に用いる．すなわち，これまで主に語彙獲得に関して有効性が検討されてきた対称性バイアスが，さらに構文獲得においても有効であることを検証する．本研究で提案する拡張モデルと Kirby モデルとの大きな違いは，子どもエージェントが常に正確に発話に対応する意味を獲得できないところである．すなわち通常であれば発話文と，その意味に相当する述語項構造がペアで渡されるはずのものを，発話の文字列のみが渡されるような状況を作る．これは，子どもエージェントにとって次のような状況に相当する．

- 親エージェントからの発話が結局何を意図するものであるか理解できなかった場合
- 親エージェントから指し示しなどによる別モダリティが手に入らず不完全なコミュニケーションに終わった場合

このような場合でも，子どもエージェントは自身が既に部分的に獲得している言語知識によって，親の発話の意図するところを補おうとする．我々はこれを

（述語項構造）→（発話文）

という順方向の含意に対する対称性バイアスと位置づけ，発話文から逆に親の意図するところを補完しようというプロセスを Kirby のオリジナルなモデルに組み込むこととする．これは親の言ったことを自身の不完全な知識により「早とちり」してしまうこと，あるいは勝手な「思い込み」をしてしまうこと

である．したがって，繰り返し学習モデルを改良し，親エージェントの発話はある一定の割合で発話意図を付与しないこととし，意図がわからない文に対して子エージェントは，(I)無視する，(II)ランダムに意味を割り振る，(III)対称性バイアスより意味を付与する，という三種類の方策を取らせることとした．この各方策について世代を追って表現度と文法の規則数，および文法の合成度について比較を行う．

また，コーパスからの言語変化の解析手法として，自然言語処理の分野で用いられる手法，すなわち，コーパスに出現する単語の意味を自動的に弁別する手法を用いる．通常の語義曖昧性解消では，単語の意味(語義)を辞書などによってあらかじめ定義し，特定の文脈中に出現した単語の意味を定義された語義の中から選択する．ところが，単語の意味は日々変化し，新しい意味や用法も生まれている．あらかじめ語義を定義するというアプローチではこのような単語の意味の変化に対応することができない．本研究では，コーパス中に出現する単語を特徴ベクトルで表現し，教師なしクラスタリングによって同じ意味を持つ単語をひとつのクラスタにまとめることで，既存の辞書に依らずに単語の意味を弁別することを目的とする．単語の意味の自動弁別は，単語の新しい意味や用法の自動的な発見につながり，言語変化の研究に役立てることができる．また，その他のコーパスからの分析として，複数言語による法律文の解析を行う．これはドメインを特定した場合，特定表現がどのように利用されるのかを言語ごとに分析を行うことで，環境（ドメイン）変化に伴う普遍的な文法の変化について考察を行う．

４．研究成果

本研究の成果を「言語の共時的変化」,「言語の通時的変化」,「コーパスによる言語変化の分析」に分け，それぞれについて詳細を述べる．

言語の共時的変化に関する研究成果：実験結果の例を図２に示す．各ドットはエージェントを表し，既存の２言語のうちのいずれかを話していた話者が，最終的に新言語であるクレオールを話すようになる過程を表している．複数の言語が共存する準安定状態が存在するなど，既存の空間なしモデルとは異なる振る舞いをすることが観察された．空間構造を用いることの利点として，コミュニティの形成過程を可視化したことが挙げられる．その結果，クレオール化の条件としてコミュニティの形成に関する条件が存在し，それは言語入力量に依存することが確認された．

なお，本研究の成果は，国際会議および論文として発表された．[雑誌論文 3][学会発表 5,6,7]

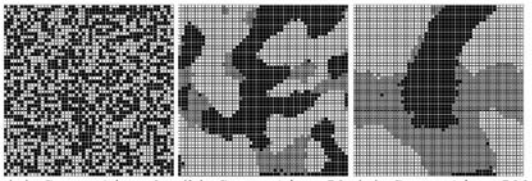

図２：格子状ネットワークを導入したクレオール化のモデル（白，黒：既存言語，グレー：クレオール）

格子状ネットワークからさらに進んで，複雑ネットワークの一種である Barabasi-Albert ネットワーク上に言語獲得と発話を行うエージェントを配置し，新言語であるクレオールが普及するシミュレーションを行った．エージェントは，初期条件として既存の２言語のどちらかを持っていると仮定し，言語の類似性などを考慮した確率分布に従って次世代の言語を決定する．実験により，格子状ネットワークと比較して，ひとつの言語に収束する速度がとても速いことが特徴として得られた．その理由として，リンクが多いエージェントは他のエージェントと比べて周囲に及ぼす影響力が強く，これを起点として言語が普及するためであると考えられる．図３は，その過程において，同じ言語を話すエージェントによるコミュニティを形成している様子を表している．

なお，本研究の成果は，国際会議にて発表された．[学会発表 3]

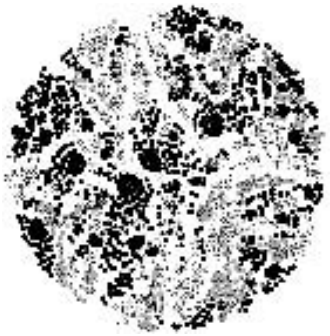

図３：複雑ネットワークを導入したクレオール化のモデル（白，黒：既存言語，グレー：クレオール）

言語の通時的変化に関する研究成果：親エージェントの発話はある一定の割合で発話意図を付与しないこととし，意図がわからない文に対して子エージェントは，(I)無視する，(II)ランダムに意味を割り振る，(III)対称性バイアスより意味を付与する，という

三種類の方策を取らせることとした．この各方策について世代を追って表現度と文法の規則数，および文法の合成度について比較を行った．この結果，対称性バイアスを用いたエージェントは，不明な文を無視するエージェントに対してより早い世代で高い表現度を得ることができ，またランダムに意味を推量するエージェントよりも少ない規則数，すなわち高い合成度で文法を構成することを示した．これにより，子エージェントは親エージェントからすべての発話について意味を付与されなくても，同様に文法を構成できると考えられる．

本研究をきっかけとして，語彙獲得のみならず，構文獲得についても対称性バイアスをはじめとする認知バイアスの効果を示す研究成果が報告されることを期待している．今後の発展として，本モデルを拡張し，屈折や一致現象をはじめとする語形変化の学習を組み込み，認知バイアスの効果を検証することが考えられる．

なお，本研究の成果は，国際会議および論文として発表された．[雑誌論文 1][学会発表 4]

コーパスによる言語変化の分析：大規模コーパスを用いた言語処理の応用として，新語義発見に向けた語義識別の研究を行った．多義性のある対象単語をコーパスからいくつか取り上げ，それらの語義を辞書から推定した．この研究をさらに発展させ，通時的なコーパスを利用することで，語義の変化を追跡調査することが期待される．なお，この研究成果は，国内学会で発表された．[学会発表 1]

また，文法構造の発達の応用例として，法令文を対象とした調査を行った．法令文は独特の表現や文法構造を持っており，これが構文の曖昧性を解消するのに役立っている．このような表現の多様性について日本，ベトナム，米国の法令文を比較し，各国の立法環境と文法構造に関する考察を行った．なお，この研究成果は，国際会議および論文として発表された．[雑誌論文 2][学会発表 2]

以上のように，基礎的なモデルを用いたシミュレーションだけではなく，コーパスの分析による複数言語の文法構造の比較等，現実にある言語資源を用いた実験を行った．

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計３件）

[1] Ryuichi Matoba, Makoto Nakamura, and Satoshi Tojo: Efficiency of the symmetry bias in grammar acquisition, Information and Computation, 査読有, 209(3), 2011, 536-547

[2] Makoto Nakamura, Yusuke Kimura, Minh Quang Nhat Pham, Minh Le Nguyen, and Akira Shimazu: Treatment of Legal Sentences Including Itemization Written in Japanese, English and Vietnamese, 自然言語処理, 査読有, 17(3), 2010, 81-100

[3] 中村 誠, 橋本 敬, 東条 敏: 空間構造を導入した言語動力学モデルによるクレオール化のシミュレーション ―言語入力量によるコミュニティ形成の差異について―, 日本認知言語学会論文集, 査読無, 10 巻, 2010, 762-765

〔学会発表〕（計７件）

[1] 中西隆一郎，白井清昭，中村誠：複数の観点から定義された用例間類似度に基づく語義識別，言語処理学会第 17 回年次大会，2011.3.9，豊橋

[2] Kenji Takano, Makoto Nakamura, Yoshiko Oyama, and Akira Shimazu: Semantic Analysis of Paragraphs Consisting of Multiple Sentences --Towards Development of a Logical Formulation Sysmtem--, JURIX2010, 2010.12.17, Liverpool, UK

[3] Makoto Nakamura, Takashi Hashimoto, and Satoshi Tojo: Self-Organization of Creole Community in a Scale-Free Network, SASO2009, 2009.9.16, San Francisco, USA

[4] Ryuichi Matoba, Makoto Nakamura, and Satoshi Tojo: Efficiency of the Symmetry Bias in Grammar Acquisition, LATA2009, 2009.4.3, Tarragona, Spain

[5] Makoto Nakamura, Takashi Hashimoto, and Satoshi Tojo: Prediction of Creole Emergence in Spatial Language Dynamics, LATA 2009, 2009.4.2, Tarragona, Spain

[6] Makoto Nakamura: Computer Simulation of Grammatical Change, International Symposium on Methodologies in Determining Morphosyntactic Change: Case Studies and Cross-linguistic Applications, 2009.3.6, Osaka, Japan

[7] Makoto Nakamura, Takashi Hashimoto, and Satoshi Tojo: Self-Organization of Creole Community in Spatial Language Dynamics, SASO2008, 2008.10.22, Venice, Italy

〔図書〕（計０件）

〔産業財産権〕
○出願状況（計０件）
○取得状況（計０件）

〔その他〕
ホームページ等

６．研究組織
(1)研究代表者
中村　誠（NAKAMURA MAKOTO）
北陸先端科学技術大学院大学・情報科学研究科・助教
研究者番号：50377438