

Research Proposal for Master's Thesis

氏名 Name	鈴木 啓章	学生番号 Student Number	1310034		
主指導教員 Supervisor	東条 敏	印 Seal	副テーマ指導教員 Advisor for Minor Research	敷田 幹文	印 Seal
副指導教員 Second Supervisor	NGUYEN, Minh Le	印 Seal			

<研究テーマ> Research Title

マルチエージェントシミュレーションによる不規則動詞の規則化に対する人口流入の影響

<研究の目的> Research Aim

本研究の目的は、言語の文化的進化の観点から英語の不規則動詞の減少に関連する人口流入の規模、頻度を検証することである。現在用いられている不規則動詞は[i - a - u]などの母音交替によるものなど、いくつかの規則性が存在する。しかし、このような不規則動詞は長い年月をかけて単純に接尾辞[-ed]を付与する規則動詞の語形変化に統一されていくという現象が見受けられる。大規模なコーパスを用いた統計的な研究^[1]では、この規則化の速さや割合とコーパス中の出現頻度との関係が示されている。

本研究では、複数世代を通じた言語獲得のマルチエージェントシミュレーションを通して不規則動詞の中に存在する規則を発見し、収束していくプロセスを検証する。また、ある段階から英語に不慣れな移民として、どんな動詞にも[-ed]を付与して過去形を形成する人口流入を発生させる。これにより不規則動詞の中で進む規則化と、[-ed]を付与する規則化の間でせめぎ合いが起こり、不規則動詞であったものが規則動詞のように使用される、また混用されるという現象が起こる。本研究では[1]によって示された結果と同程度での速度で規則化が進行するようにシミュレーションを行い、人口流入の規模や頻度による影響を検証する。

<研究の背景・特色> Research Background, Originality

現在の日常的に用いられている不規則動詞はおおよそ 180 語存在し、会話の中に出現する動詞の約 70% が不規則動詞である(be, have, do, etc...)^[2]。これらの不規則動詞は古英語時代[AD 800 頃]に強変化動詞(strong verb)と呼ばれ母音交替によって現在形、過去形、過去分詞が作られていた。また、現代英語で接尾辞[-ed]を付与する規則動詞は弱変化動詞(weak verb)と呼ばれる。このように古英語で母音交替によって語形変化が行われていた不規則動詞が、単純に接尾辞[-ed]をつけるだけの規則動詞に変化している。

[1]では古英語時代から現在まで残っている 177 個の不規則動詞を手で抽出し、古英語から中期英語[AD1200]、中期英語から現代英語[AD2000]における不規則動詞が規則化されていく変遷状況を調査した。その結果、不規則動詞は徐々に規則化され 177—145—98 個と減少していくことが判明した。次に 177 個の不規則動詞についてコーパス^[4]を用いてそれぞれの出現頻度を算出し、頻度ごとに 6 段階に分類した。そしてこの頻度を用いると各クラスの不規則動詞が規則化されるまでの時間は、頻度の平方根におおよそ比例することが明らかになった。コーパス中の出現頻度が低～中程度の不規則動詞に関しては AD800 から 300～2000 年間の期間中に規則化されると予想されている。

本研究ではこれらの予想をもとに、世代を重ねるごとに言語が進化していく文化的進化の観点から不規則動詞が規則化されていく過程のシミュレーションを行う。シミュレーションで用いるモデルは教育者(親)と学習者(子)による文法獲得モデルである繰り返し学習モデル(Iterated Learning Model: ILM)を用いる^[5]。ILM は親から子へ 1 対 1 で発話が伝達されるが、本研究では言語を用いる集団を扱うことから多対多で学習が行えるように拡張を行う。この拡張モデルを用いて[1]で示されている不規則動詞が規則化される予想期間で、規則化が行われるような移民の流入規模、頻度をシミュレーションを通して探索する。特にシミュレーション中で発話される頻度が低い不規則動詞が規則化されるためには、移民流入のタイミングで運よく発話されるか、それ以外の期間でも消滅しない程度で発話される必要がある。コーパスを用いた統計情報だけでは、出現頻度と規則化スピードの関係しか明らかにすることはできない。しかし、本研究では ILM を用いることによって学習過程の中で不規則変化の中のどのような規則が、せめぎ合いを起こすのか、また淘汰されて規則変化するようになるのかといった、規則化までの考察も行うことができると考えられる。

<研究計画・方法>Research Plan, Method

本研究では ILM を用いたマルチエージェントシミュレーションを行う。ILM では大人エージェントが自分自身の言語知識に基づき発話を行う。子エージェントはこの発話と同時に、発話の意味も同時に受け取る。大人エージェントにより渡された発話とその意味を子エージェントは言語知識として蓄え、その知識を汎化し文法とする学習を行う。ある程度の発話が行われると、子エージェントが新たな親エージェントとなり、学習した文法をもとに新たな子エージェントに発話と意味の伝達を行う。親から子への言語の継承のモデルである。これをもとにシミュレーションを 2 つのフェーズに分けて行う。

第 1 段階は不規則動詞の中にも規則性が存在し、それがシミュレーションによって発見され収束に向かうプロセスの検証である。古英語の強変化動詞の中にも語幹母音によって 6 つのクラスに区別される。[sing – sang – sung , eat – ate – eaten]など現在にも残っておりクラスによって母音交替などの規則が異なる。よって現存する不規則動詞中の規則もこれらのクラス分けに対応することが考えられる。

1. 現代英語の不規則動詞の中の規則性の調査

上記で述べたように、古英語に関しては強変化動詞についてクラス分けがなされているが、現代英語に関しても不規則動詞中の母音交替の規則などによってクラス分けがなされているか調査を行う。この調査でのクラス分けを第一段階のシミュレーションの収束目標とする。

2. ILM の拡張および語形変化学習への対応

ILM を多対多に拡張を行う。ILM の拡張に関しては^[6]の研究を参考とする。[6]では複数の大人(親)エージェントと子エージェントを導入し、親子間の発話だけでなく子から子の発話、親子ではない大人から子への発話を導入した。また、フィードバックとして聞き手から発話者への発話が理解できたかどうかという情報伝達も追加された。結果として同世代間の会話の割合が大きくなると言語の収束は加速することが示された。本研究でも、発話された動詞の意味が伝わっているかどうかのフィードバックや、移民を子エージェントとして扱うことで規則化の速度に影響を与えることができるのではないかと考えられる。

3. 拡張した ILM を用いてシミュレーションを行う

まず、シミュレーションで用いる動詞を現存する 98 語を参考に人口動詞として準備する。この人口動詞にはシミュレーション中の発話の中で出現する頻度情報も定義する。シミュレーションを通して 1.で調査したいくつかのクラスに収束する過程を計測する。

第 2 段階は実際に移民として人口流入を考慮したシミュレーションを行う。

4. 3.で利用したシミュレーションモデルに様々な規模、頻度で人口流入を起こす

この人口流入により、第一段階で不規則動詞が収束したいくつかのパターンが失われること、またそのパターン中での母音交替の混用、sing の過去形として sang と singed の両方が用いられるような不規則変化と規則変化の混用も起こる。このような現象を観測した結果、[1]で予想されている規則化の予想期間通りに規則化を起こすには、どのような人口流入の規模、頻度が必要かを検証する。

最後に、以上の研究結果を論文にまとめる。

参考文献

- [1] Erez Lieberman, Jean-Baptiste Michel, Joe Jackson, Tina Tang & Martin A. Nowak, Quantifying the evolutionary dynamics of language. Nature Vol449 (11 October 2007).
- [2] Pinker, S. The irregular verbs. Landfall 83–85 (Autumn issue, 2000).
- [3] Hare, M & Elman, J. Learning and morphological change. Cognition 56, 61–98(1995).
- [4] CELEX database. <http://www.wlands2.let.kun.nl/members/software/celex.html>.
- [5] S.Kirby. Learning, bottlenecks and the evolution of recursive syntax. In Linguistic Evolution through Language Acquisition. Cambridge University Press, (2002).
- [6] Tao Gong, James W. Minett and William S.-Y. Wang. A simulation study exploring the role of cultural transmission in language evolution. Connection Science Vol.22 69-85 (March 2010).

<現在までに単位修得した専門科目> 科目数：9、領域数：5

I114 基礎情報数学, I214 システム最適化(ア), I213 離散信号処理特論(イ),
I223E 自然言語処理論 I(E)(ウ), I236 論理推論と知識表現(ウ), I226 コンピュータネットワーク特論(エ),
I217 関数プログラミング(オ), I234 ソフトウェア環境構成論(オ), I478 システム開発管理(オ)