

修士論文

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関する人口流入の影響

目 次

1	はじめに	2
1.1	研究の目的	2
1.2	本論文の構成	2
2	背景	3
2.1	言語学的背景	3
2.1.1	英語の起源と時代区分	3
2.1.2	不規則動詞とは	4
2.2	歴史的背景	6
3	先行研究	8
3.1	不規則動詞の出現頻度と規則化の速度に関する研究	8
3.2	語形変化に影響を及ぼす力に関する研究	8
4	過去形獲得モデルの提案	8
4.1	実験に必要な知識	8
4.1.1	マルチエージェントとコミュニケーション	9
4.1.2	複雑ネットワークを用いた言語変化シミュレーション	9
4.1.3	人工的な不規則動詞に対する既存の母音活用の適用	9
4.1.4	MetaPhone アルゴリズム	9
4.1.5	遺伝的アルゴリズム	9
5	実験	9
6	実験結果と考察	9
7	結論	9
	参考文献	9

1 はじめに

1.1 研究の目的

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 [1]. これらの不規則動詞は Old English 時代 [AD 800 頃] に強変化動詞 (strong verb) と呼ばれ主に母音が変化して活用していた. Modern English では例外も含め 9 クラスに分類され、クラス内にも細かい分類がある [2]. しかし、不規則動詞には接尾辞 [-ed] をつける規則的な活用に変化しているという現象が起こっている. 英語の歴史的な流れの中では、Old English から中期英語時代における海賊によるイングランドの侵略、ノルマン征服などの人口流入を伴った言語接触により不規則動詞の規則化の誘発、またその加速が起こっている.

本研究ではこの不規則動詞の規則化に対する人口流入の規模、頻度をシミュレーションによって検証することを目的とする. 検証のために遺伝的アルゴリズム [10] (以下 GA) をベースに、エージェントコミュニケーションと変化を進行させるような (外圧) を組み込んだモデルを作成し、複数世代を通したシミュレーションを行う.

1.2 本論文の構成

2 背景

本章では英語の不規則動詞の言語的背景、また言語接触に関する歴史的背景について述べる。言語学的背景では、英語の起源と時代区分について述べる。また Old English 時代に用いられていた強変化動詞について説明を行う。不規則動詞の規則化に関する歴史的背景では、Viking(海賊) やノルマン征服などイングランドにおける侵略と英語に与えた影響について述べる。

2.1 言語学的背景

本節では英語の不規則動詞について英語の起源と時代区分を含めて説明を行う。また、

2.1.1 英語の起源と時代区分

はじめに英語の起源について述べる。英語は Proto-Indo-European を起源とするゲルマン語派言語である [4]。Proto-Indo-European を起源とするその他の言語には、ケルト語派、イタリアック語派などがある。Proto-Indo-European の系統図を図 1 に示す。

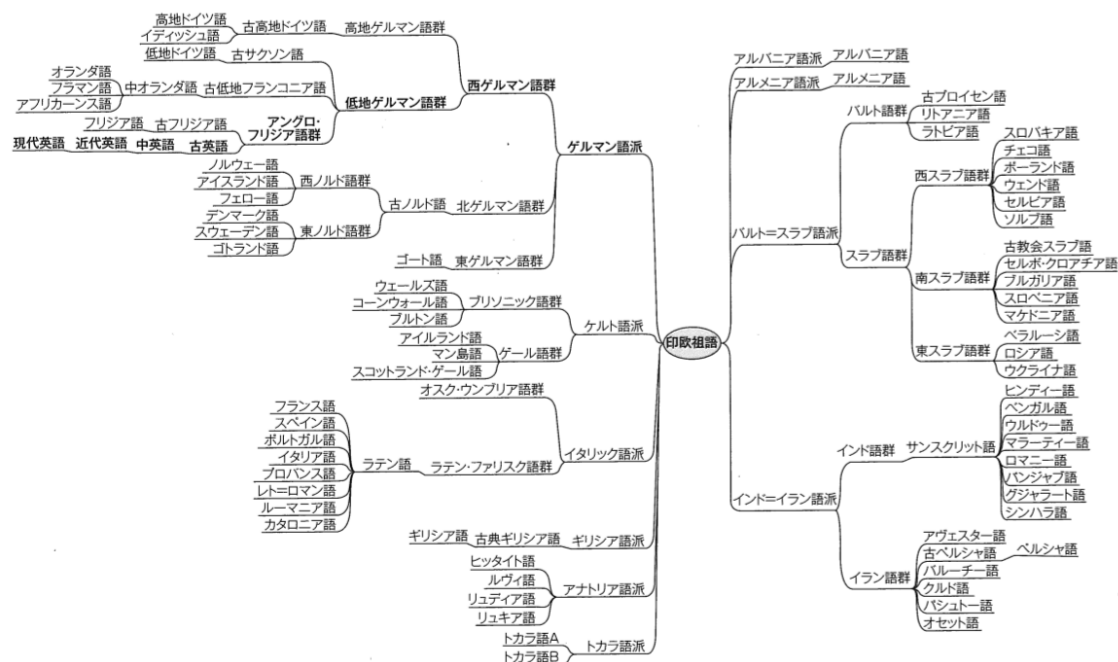


図 1: 印欧語 (Proto-Indo-European) 族の系統図 [4] より引用

Proto-Indo-European 語族には英語を含む多くのヨーロッパ諸国の言語が含まれていることがわかる。Proto-Indo-European は現在は失われているが、紀元前 4000 年ごろに南

ロシアに栄えていたクルガン文化の担い手によって話されていたという説がある [4]. その後、各地で方言化し諸言語へと進化していった.

次に古英語以降の時代区分について述べる. 英語の年代区分は、ノルマン征服など歴史的な事実を区切りに用いるが、3 区切りや 6 から 7 つに区切るモデルも存在する. 表 1 に 4 つの時代に区切るモデル [3] を示す.

表 1: 英語の時代区分

A.D 500 -1150	Old English
A.D 1150 -1450	Middle English
A.D 1450 -1700	Early Modern English
A.D 1700 -	Modern English

各時代について簡単に説明する. Old English 時代は大ブリテン島南部でアングル、サクソン、ジュート族によって言語が確立された時期である. その後、ノルマン征服によってノルマンフランス語との接触による影響が出始めた時代が Middle English 時代である. 活版印刷技術が西ヨーロッパに広がりはじめた時代が Early Modern English 時代、アン女王の時代以降が Modern English 時代となる.

第 2.1 節の区分において Old English 時代に母音交替によって活用していた動詞が不規則動詞である.

2.1.2 不規則動詞とは

本節では現在の不規則動詞について、そのルーツと人間の過去形生成の仕組みについて述べる.

- 強変化動詞

Old English 時代に用いられていた動詞は母音交替によって活用していた強変化動詞、(-t/de) など歯茎音の接辞によって活用する弱変化動詞、過去時制が現在形の意味も表す過去現在動詞やその他少数の不規則動詞がある [6]. 現在不規則動詞と呼ばれるものは強変化動詞がルーツとされる. 強変化動詞の基本構造は [子音+母音+子音] であり、これに *rið-an*(不定詞), *rād*(過去 *i* 形), *rið-on*(過去 *Λ* 形), *rið-ed*(過去分詞) のように活用語尾がつく. ここで、過去 1 形は過去単数 1,3 人称、過去 2 形は過去単数 2 人称および過去複数形である. これら 4 つの形を作る中で母音交替が起こる. 強変化動詞にはその母音交替のパターンによってクラス分けが存在する. 強変化動詞にも基本形と基本形から派生した変異形が存在するが個々では基本形のみ扱うものとする. 表 2 に例を示す.

表 2: 母音交替パターンによるクラス分け [6] を元に作成

類	不定形	母音交代系列
I	rīdan(to ride)	ī,ā,i,i
II	bēodan(to bid)	ēo,ēa,u,o
III.N	bindan(to bind)	i,a,u,u
III.L1	helpan(to help)	e,ea,u,o
III.L2	weorpan(to throw)	eo,ea,u,o
IV	beran(to bear)	e,æ,ǣ,e
V-1	metan(to measure)	e,æ,ǣ,e
V-2	giefan(to give)	ie,ea,ēa,ie
VI	faran(to go)	a,ō,a
VIIa	hātan(to call)	ā,ē,ā, etc...
VIIb	bannan(to summon)	a,ēo,a, etc..

I, II、III 類はそれぞれ長母音、二重母音、母音+亮音 (鼻音 n,m, 流音 l, r) の動詞である。語幹母音が短母音でその末尾の音が流音 l,r、鼻音 n,m である動詞が IV 類、そうでない動詞が V 類である。VI は IV 類と V 類が混合したものである。VII 類はかつてゴード語において畳音 (zigzag などの音を重ねるもの) を用いて過去形を作る動詞であったものである。

- 過去形生成の仕組み

Jackendoff[8] によれば形態論的には規則動詞の過去形生成は生産的ルールであり、不規則動詞の過去形生成は半生産的ルールである。ここで生産的という意味は、動詞に対してまったく規則的に行われるという事である。英語の現在分詞 (-ing) の生成は英語の動詞全てに例外なく行われる。つまり、生産的なルールは発話するときになってその場で適用すればよく、長期記憶に蓄える必要はない。しかし、半生産的なルールはある程度長期記憶に蓄える必要がある。不規則動詞の場合、母音交替にある程度のパターンは見られるものの完全に予想することはできない。そのため、不規則動詞の過去形を使う場合は長期記憶にとどめておく必要がある。

また Pinker らは Word and Rule Theory[9] 提唱した。概要図を図 2 に示す。

図 2 中の Lexicon とは記憶の一部で形態素などを保持している。Grammar は形態素や単純な語を組み合わせる能力である。また Lexicon と Grammar には同時にアクセスできる。例えば hold という語幹が入力された時 (発話しようとした時)、Lexicon の中に hold の過去形とマッチする held が存在する場合、Grammar 側で接尾辞 ed をつけて発話するという作業はストップさせられる。よって発話は held となる。もし、なんらかの理由で held を見つけられない場合 Grammar 側からの発話となり ed がついた過去形が生成される.held

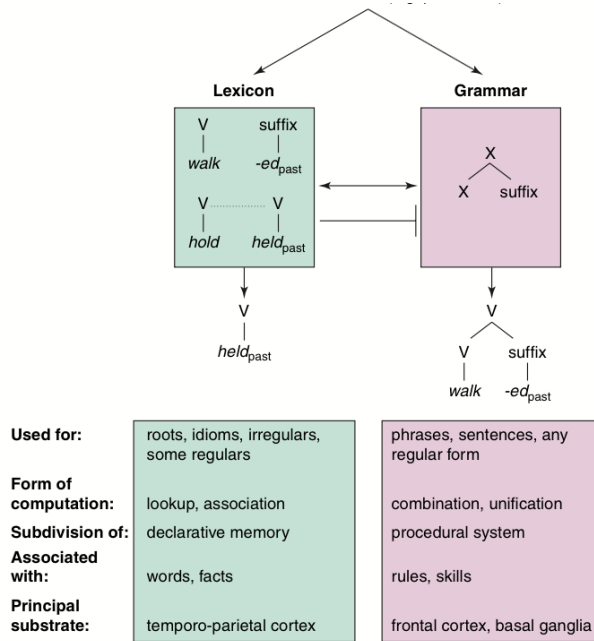


図 2: Word and Rule Theory[9] より引用

を見つけれない理由としては、held の頻度がとても低いか 0 であること、子供のときに異なる形で記憶された、記憶にダメージをおったなどが挙げられている。不規則であった動詞が規則化される (ed をつけて過去形を生成する) 原因としても上記が関係しているのではないかと指摘している。

2.2 歴史的背景

本節では実際の英語史において起こった人工流入や言語接触の影響 [3, 5, 7] について述べる。

Old English は大ブリテン島でケルト人が使用していたとされているが、以下の侵略などの歴史を通して形作られてきた言語である。A.D400-550 頃にゲルマン人の英国侵略が始まり、アングル人 (Angles)、サクソン人 (Saxons)、ジュート人 (Jutes) の部族も侵略に加わった [7]。この影響から Old English 初期の語彙にはゲルマン系の語彙が多い。また、A.D500-700 にかけて英国にキリスト教が伝わった影響でラテン語の借用、またはそのまま Old English に持ち込むといった現象が起こった。apostol (apostle)、cyrce (church) などの教会用語が主である。

A.D800-1066 ごろに Old English に影響を与えたのは海賊 (Viking) である。海賊はスウェーデン、ノルウェー、デンマークに居住していたデーン人である。また海賊は古北欧語 (ノルド語) 話者であった。海賊は主に商売や時には略奪行為を目的とし、英国の東海岸を襲った。また、襲撃を繰り返すだけでなく英国に定住する者もいた。図 3 におおよその地域

(青)を示す。英国のアルフレッド大王は海賊と戦うこともあったが、やがて共存の道を開いていった。その過程の中でノルド語は Old English に吸収されていくが、かなりの数の語を Old English の中に残した。この時取り込まれた語は fellow, sky, to take, などの一般語、they-their-them などの代名詞は今日でも使用されている。この接触によって Old English では三人称単数の語尾に [-s] をつけるようになった。またこの時期に強変化動詞の規則化が始まったとされている。

次に A.D1066-1345 ごろのノルマン征服の影響について述べる。Old English の後期に英国の王位継承をめぐる英国王室と北フランスのノルマンディ公国の間で争いが起こった。争いはノルマンディ側の勝利で終了した(ノルマン征服)。図3にノルマン人によって征服された地域(赤)のおおよその位置を示す。そのため征服以降では、行政や軍事、司法などはフランス語で行われた。Old English に及ぼした影響であるが、一部の同義語の交代や新語の流入が起こっている。表3に同義語の交代の例を示す。

表 3: 同義語の交代例 [7] より引用

以前の表現	以降の表現	
preost	chapellain	僧侶
firð/grið	pais	平和
unfið	werre	戦争
lufu	cherite	愛

表3の語は完全に入れ替わったわけではなく、preost(priest), lufu(love)として Modern English に残っている単語も存在する。新語の流入では baron, champion, army, pawn など爵位、軍事、司法用語などが約1万語ほど流入したと考えられている。また、文法などの言語構造の簡略化や、海賊の影響で始まった規則化傾向も上昇した。ノルマン征服後3世紀ほどで Old Middle English は復活し、英国の話し言葉となった。しかし、フランス語は書き言葉として吸収が続けられた。

以上より、侵略、征服など様々な言語接触を経験していることが英語の特徴であると言える。

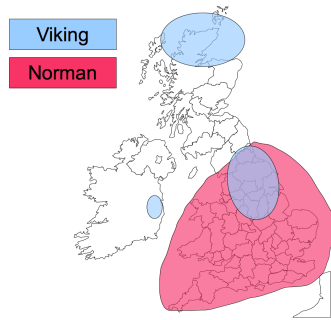


図 3: 海賊とノルマン人の居住、征服地域

3 先行研究

本章では不規則動詞に関するコーパスを用いた統計的な研究について述べる．不規則動詞の出現頻度と規則化の速度に関する研究では、接尾辞 *ed* をつけることによって過去形を生成するようになる規則化がコーパス内での出現頻度と時間の関数として導出した研究について述べる．

語形変化に影響を及ぼす力に関する研究では、単語が古い形から新しい形に変化するとき、変化を推し進めるような 2 つの力 (内圧、外圧) を数値的に求めた研究について述べる．

3.1 不規則動詞の出現頻度と規則化の速度に関する研究

3.2 語形変化に影響を及ぼす力に関する研究

4 過去形獲得モデルの提案

本章ではマルチエージェント環境の中で各エージェントがコミュニケーションを通して共通の過去形を獲得するモデルを提案する．このモデルを使用することで、規則化の速度やその速度を引き起こす原因について考察できると考えられる．

4.1 実験に必要な知識

本節では提案モデルに必要な知識について述べる．マルチエージェントとは何か、またコミュニケーションの目的と言語変化への関わりについて述べる．マルチエージェント環境を現実世界の人間のつながりに近づけるために使用する複雑ネットワークについても説明を行う．また、人工的な動詞の過去形をどうやって生成しているのかアンケート調査を行った研究とそれに付随して単語の音韻的距離を計測するアルゴリズムについて述べる．最後にエージェントの学習機構に用いる遺伝的アルゴリズムについて述べる．

- 4.1.1 マルチエージェントとコミュニケーション
- 4.1.2 複雑ネットワークを用いた言語変化シミュレーション
- 4.1.3 人工的な不規則動詞に対する既存の母音活用の適用
- 4.1.4 MetaPhone アルゴリズム
- 4.1.5 遺伝的アルゴリズム

5 実験

6 実験結果と考察

7 結論

参考文献

- [1] *Pinker, S. The irregular verbs. Landfall 83—85 (Autumn issue, 2000)*
- [2] *Pinker, S, Prince, A. On language and connectionism: analysis of a parallel distributed processing model of language acquisition. Cognition 28,p73-193 (1988)*
- [3] *Tom McArthur, THE ENGLISH LANGUAGES, Cambridge University Press (1998)*, 英語系諸言語, 牧野武彦 監修, 山田 茂, 中本 恭平 訳, 三省堂 (2009)
- [4] 堀田隆一, 英語史で解きほぐす英語の誤解: 納得して英語を学ぶために. 中央大学出版部, (2011)
- [5] *Philip Gooden, THE STORY OF ENGLISH (2009)*, 物語 英語の歴史, 田口孝夫 監修, 悠書館 (2012)
- [6] 岩本忠. ”古期英語強変化動詞の類別定義と韻構造: 基本型と変異型.” (2006)
- [7] 亀井孝, 河野六郎, 千野栄一. 言語学大辞典. Vol. 1. 三省堂, (1996)
- [8] *Jackendoff, Ray. Foundations of language: Brain, meaning, grammar, evolution. Oxford University Press, (2002)*
- [9] *Pinker, Steven, and Michael T. Ullman. ”The past-tense debate.” cognitive processing 5 (2002)*

- [10] 伊庭 斉志, 遺伝的アルゴリズムの基礎-GA の謎を解く オーム社 (1994)
- [11] Erez Lieberman, Jean-Baptiste Michel, Joe Jackson, Tina Tang , Martin A. Nowak, *Quantifying the evolutionary dynamics of language. Nature Vol449 (11 October 2007)*
- [12] CELEX <http://wwwlands2.let.kun.nl/members/software/celex.html>
- [13] Ghanbarnejad, Fakhteh, et al. *Extracting information from S-curves of language change. Cornell University Library arXiv preprint arXiv:1406.4498 (2014)*
- [14] Chris Upward, *Spelling Reform in German. Journal of the Simplified Spelling Society, J21, 1997-1 pp22-24,36*
- [15] Google Ngram Viewer, <https://books.google.com/ngrams>