

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
Author(s)	鈴木, 啓章
Citation	
Issue Date	2015-03
Type	Thesis or Dissertation
Text version	author
URL	<a href="http://hdl.handle.net/10119/12631">http://hdl.handle.net/10119/12631</a>
Rights	
Description	Supervisor: 東条敏, 情報科学研究科, 修士

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  
動詞の規則化に対する人口流入の影響

北陸先端科学技術大学院大学  
情報科学研究科情報科学専攻

鈴木啓章

2015年3月

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  
動詞の規則化に対する人口流入の影響

指導教員 東条 敏 教授

審査委員主査 東条 敏 教授

審査委員 Nguyen Minh Le 准教授

審査委員 白井 清昭 准教授

北陸先端科学技術大学院大学  
情報科学研究科情報科学専攻

1310034 鈴木啓章

提出年月: 2015 年 2 月

## 概要

Lieberman [16] らは統計的な研究で不規則動詞の規則化現象を示した. 本研究の目的は、マルチエージェントシミュレーションでその規則化現象の追試を行うことである.

ドイツ語やラテン語などをはじめ、人間の言語は様々な屈折体系を持つ (屈折言語) ことで他者とのコミュニケーションを円滑に行える. また、日本語のように屈折ではなく膠着的な特徴を持つ膠着言語も存在する. 屈折とは単語自体を変化させて意味を付加し、膠着とは単語に何らかの形態素をつけることで意味を付加するものである. しかし、齊藤 [23] によれば、言語は屈折だけを持つ、膠着だけを持つというように完全に区別することはできない.

英語における時制もその屈折の一つである. 動詞に接尾辞「ed」をつければ過去形を生成できるという規則は、新しく英語に加わる動詞にも例外なく適用できる. 例えば、google はすでに「google で検索する」という動詞として用いられており、過去形は googled である. しかし、母音交替 (Ex. sing - sang) によって過去形を生成する不規則動詞は例外的な扱いになる.

現在用いられている不規則動詞は Old English 時代 (A.D 800) の強変化動詞がもとである. Old English 時代では母音交替によるパターンは何パターンか存在したが、現在の不規則動詞の母音交替のパターンは語形そのものが変化、または不変であるパターンを含めればかなりの数が存在する [2]. このように不規則動詞の過去形には明確なルールが存在しないことは明らかである. 齊藤 [23] によれば、複雑な屈折 (ここでは多数のパターンがあること) の習得過程には多くの精神的負担がかかる. よって習得に負担がかかるような複雑な屈折は時を経るごとに減っていくのではないかと考えられる. しかし、現代まで不規則動詞が生き残っているということもまた事実である. その理由の一つとして、不規則動詞には使用頻度が高いものが多いということがあげられる. 「be, have, come...」などよく登場する動詞は不規則動詞である. 実際、英語に存在する不規則動詞は全動詞の約 3% ほどであるが、会話に登場する動詞の約 70% が不規則動詞である [1].

このような不規則動詞であるが、統計的な研究から使用頻度がそこまで高くない不規則動詞は規則変化をするようになっていくことが明らかになった. 規則変化が進行する速度も時間と出現頻度の関数として示され、過去から未来すべての年代の不規則動詞の数も予想できるようになった.

一方、上記のような言語の歴史的変化を扱う際には人間のコミュニケーションを考慮することが重要になってくる. なぜなら、言語変化とは言語そのものがもつ特徴と人間のコミュニケーションの目的が複雑に絡み合う現象だからである. ここで人間のコミュニケーションの目的とは最小の物理的努力で、明確に意味を伝えることである [12]. 本研究では人間のコミュニケーションと言語の特徴を考慮したマルチエージェントシステムを提案する. 具体的には、各エージェントが持つ動詞の過去形が全く決定されていない状態から、共通の過去形を獲得して行くモデルである. 本研究における言語そのものの特徴とは動詞の出現頻度であり、また屈折が保たれる場合と失われる場合

(膠着的になる) があるという事である. 出現頻度は人間のコミュニケーションによって用いられ、屈折と膠着はコミュニケーションを通じた学習によって決定される.

本研究で提案するモデルでは、マルチエージェントシステムを利用するにあたり複雑ネットワークをエージェント環境に用いた. 複雑ネットワークとはグラフ構造と類似するものであり、ノードとエッジから構成される. 本モデルではノードはエージェントを意味し、エッジはエージェント同士のつながりを意味する. エージェントはエッジがつながっている場合にのみコミュニケーションが可能である. 複雑ネットワークは社会言語学分野で言語変化のシミュレーションに利用されている. また、エージェントの学習機構は遺伝的アルゴリズム (GA) を用いた. エージェントは動詞の過去形を生成するための複数の選択枝 (不規則変化するための母音、規則変化のための接尾辞 ed) と、その動詞の出現頻度を持っている. 選択枝は遺伝子として表現されている. 学習は世代を経るごとに複数の選択枝 (遺伝子) から最も適した選択枝を決定するものである.

上記の提案モデルを用いて行った実験は以下の 2 つである.

- 実験 1

実験 1 の目的は Lieberman [16] の追試である. 追試を行うために次の条件を与える. エージェントが任意の動詞を規則変化させて発話する場合は、遺伝子の適応度をパラメータで抑制する. このパラメータはエージェントが規則変化を受け入れるか、また不規則変化を維持するのか、の割合を表す. パラメータによる抑制は高頻度の動詞に対してのみ行う.

- 実験 2

実験 2 の目的は、複雑ネットワークのサイズ (エージェントサイズ) が規則化の速度に及ぼす影響を検証することである. 複雑ネットワークを用いている以上、ネットワーク構造の観点からも考察が必要である. 高頻度の動詞のみに条件を与える点は実験 1 と同じである.

高頻度動詞に与えられる条件は、コミュニケーションの性質に従い規則化を抑制するためのものである. 結果としては、実験 1 では先行研究の追試を行うことはできなかった. 実験 2 では用いたネットワークの特徴を十分に反映できていることがわかった. また、どちらの実験でもエージェント知識の初期状態に大きく依存することが明らかになった. しかし、本研究で提案したモデルは言語そのものの特徴やコミュニケーションの目的を取り込んだものであり、より現実に近い状況で言語変化をシミュレーションできると考えられる. これらは言語変化シミュレーションのモデルに対する貢献であると考えられる.

今後の課題は、初期化の調整、エージェントに年齢などの概念を加える、また先行研究の追試を行えるように根拠のあるパラメータ調整が必要である.

# 目次

<b>1</b>	<b>はじめに</b>	<b>1</b>
1.1	研究の目的	1
1.2	本論文の構成	2
<b>2</b>	<b>背景</b>	<b>3</b>
2.1	言語学・認知科学的背景	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	5
2.1.3	膠着と屈折	8
2.2	歴史的背景	9
2.3	背景のまとめ	10
<b>3</b>	<b>先行研究</b>	<b>12</b>
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	12
3.2	語形変化に影響を及ぼす力に関する研究	16
3.3	不規則動詞の規則化式の導出	18
3.4	先行研究のまとめとシミュレーションの意義	19
<b>4</b>	<b>実験に必要な知識</b>	<b>20</b>
4.1	マルチエージェントとは	20
4.2	複雑ネットワークと言語伝搬シミュレーション	21
4.2.1	複雑ネットワーク	21
4.2.2	言語伝搬シミュレーション	24
4.3	MetaPhone アルゴリズム	24
4.4	遺伝的アルゴリズム	26
4.5	実験に必要な知識のまとめ	29
<b>5</b>	<b>過去形獲得モデルの提案</b>	<b>30</b>
5.1	コミュニケーションの目標	30
5.2	エージェントの知識	31
5.3	マルチエージェント環境	33
5.4	学習機構の構築	34

5.4.1	遺伝子の生成	34
5.4.2	遺伝子の評価	35
5.5	実装	36
<b>6</b>	<b>実験</b>	<b>37</b>
<b>7</b>	<b>実験結果と考察</b>	<b>39</b>
7.1	実験 1	39
7.1.1	出現頻度 1000 の場合	39
7.1.2	出現頻度 100 の場合	43
7.1.3	出現頻度 10, 1 の場合	44
7.1.4	実験 1 のまとめ	44
7.2	実験 2	45
7.2.1	出現頻度 1000 の場合	46
7.2.2	出現頻度 100 の場合	47
7.2.3	実験 2 のまとめ	47
<b>8</b>	<b>結論</b>	<b>48</b>
8.1	おわりに	48
8.2	今後の課題	48
8.2.1	人口流入の影響	48
8.2.2	初期化依存の問題	49
8.2.3	年齢設定	49
8.2.4	課題のまとめ	50
	参考文献	<b>52</b>

## 目 次

1	印欧祖語 (Proto-Indo-European) 族の系統図 [5] より引用 . . . . .	4
2	Word and Rule Theory[13] より引用 . . . . .	7
3	ヴァイキングの影響地域 [11] より引用 . . . . .	10
4	出現頻度ごとの不規則動詞数 [16] より引用 . . . . .	13
5	不規則動詞の減少時間と減少具合 . . . . .	15
6	言語変化と S-curve . . . . .	17
7	1800 から 2000 年の各語形の変化 [20] より引用 . . . . .	17
8	$I(\omega, t)$ ( $-2000 \leq t \leq 2500$ ) による不規則動詞の分布 . . . . .	18
9	エージェントと環境 . . . . .	20
10	複雑ネットワーク . . . . .	21
11	BA モデルの流れ . . . . .	22
12	BA モデルで生成した複雑ネットワーク . . . . .	23
13	遺伝的アルゴリズムの処理の流れ . . . . .	27
14	交叉 . . . . .	28
15	エージェントの知識とコミュニケーション . . . . .	32
16	遺伝子 . . . . .	34
17	ペナルティの意味付け . . . . .	36
18	実験 1 で用いるネットワーク . . . . .	38
19	実験 2 で用いるネットワーク . . . . .	38
20	規則化が起こった例 出現頻度:1000 高頻度ペナルティ 4.0 . . . . .	40
21	不規則変化を保つ例 出現頻度:1000 高頻度ペナルティ 4.0 . . . . .	40
22	規則化と不規則変化が競合する場合 出現頻度:1000 高頻度ペナルティ 4.0 . . . . .	40
23	類似ペナルティを受ける動詞 出現頻度:1000 高頻度ペナルティ 4.0 . . . . .	40
24	規則化が起こった例 出現頻度:1000 高頻度ペナルティ 5.5 . . . . .	41



25	不規則変化を保つ例	
	出現頻度:1000 高頻度ペナルティ5.5 . . . . .	41
26	類似ペナルティを受ける動詞	
	出現頻度:1000 高頻度ペナルティ5.5 . . . . .	42
27	初期ルーレットの割合に大きな差がある例 出現頻度:1000 高頻度ペナルティ5.5 .	42
28	類似ペナルティを受ける動詞	
	出現頻度:1000 高頻度ペナルティ6.0 . . . . .	42
29	音節ペナルティを受ける例	
	出現頻度:1000 高頻度ペナルティ6.0 . . . . .	42
30	不規則変化と規則変化の逆転	
	出現頻度:100 . . . . .	43
31	初期ルーレットの割合に大きな差がある例 出現頻度:100 . . . . .	43
32	規則変化が起こった例	
	出現頻度:100 . . . . .	44
33	低頻度の場合 出現頻度:10 . . . . .	45
34	超低頻度の場合 出現頻度:1 . . . . .	45
35	規則化が起こった例	
	出現頻度:1000 . . . . .	46
36	不規則変化を保つ例	
	出現頻度:1000 . . . . .	46
37	初期ルーレットの割合に大きな差がある例 出現頻度:1000 . . . . .	46
38	規則化と不規則変化が競合する場合	
	出現頻度:1000 . . . . .	46
39	規則化が起こった例 出現頻度:100 . . . . .	47
40	不規則変化と規則変化の逆転	
	出現頻度:100 . . . . .	47

## 表 目 次

1	英語の時代区分 . . . . .	4
2	母音交替パターンによるクラス分け [9] を元に作成 . . . . .	6
3	Level of Similarity (to “ throw ”) [3] より引用 . . . . .	7
4	同義語の交代例 [10] より引用 . . . . .	9
5	不規則動詞の相対出現頻度 . . . . .	13
6	人工動詞 . . . . .	31
7	出現頻度と人工動詞の数 . . . . .	33
8	実験 1 の設定 . . . . .	37
9	実験 2 の設定 . . . . .	38
10	類似ペナルティ対象の動詞 . . . . .	39
11	類似ペナルティ対象の動詞 . . . . .	45

# 1 はじめに

## 1.1 研究の目的

人間の言語は様々な屈折体系を持つことで、他者とのコミュニケーションを円滑に行えるようにし、かつ無限の生成能力を持つとされる。屈折とは単語自体を変化させて意味を付加することである。英語における時制もその屈折の一つである。動詞に接尾辞「ed」をつければ過去形を生成できるという規則は、新しく英語に加わる動詞にも例外なく適用できる。例えば、google はすでに google で検索するという動詞として用いられており、過去形は googled である。しかし、母音交替 (Ex. sing - sang) によって過去形を生成する不規則動詞は例外的な扱いになる。

現在用いられている不規則動詞は Old English 時代の強変化動詞がもとである。Old English 時代では母音交替 (屈折) によるパターンは何パターンか存在したが、現在の不規則動詞の母音交替のパターンは語形そのものが変化、または不変であるパターンを含めればかなりの数が存在する [2]。このように不規則動詞の過去形には明確なルールが存在しないことは明らかである。齊藤 [23] によれば、複雑な屈折 (ここでは多数のパターンがあることとする) の習得過程には多くの精神的負担がかかる。よって習得に負担がかかるような複雑な屈折は時を経るごとに減っていくのではないかと考えられる。しかし、現代まで不規則動詞が生き残っているということもまた事実である。その理由の一つは、不規則動詞には使用頻度が高いものが多いということがあげられる。「be, have, come...」などよく登場する動詞は不規則動詞である。実際、英語に存在する不規則動詞は全動詞の 3% ほどであるが、会話に登場する動詞の約 70% が不規則動詞である [1]。

このような不規則動詞であるが、統計的な研究から使用頻度がそこまで高くない不規則動詞は規則変化をするようになっていくことが明らかになった。規則変化が進行する速度も時間と出現頻度の関数として示され、過去から未来すべての年代の不規則動詞の数も予想できるようになった。

一方、このような言語の歴史的変化を扱う際には人間のコミュニケーションを考慮することが重要になってくる。なぜなら、言語変化とは言語そのものがもつ特徴と人間のコミュニケーションの目的が複雑に絡み合う現象だからである。本研究では人間のコミュニケーションと言語そのものの特徴を表現可能なマルチエージェントシステムを提案する。そして、シミュレーションを通して、統計的な研究から明らかになった規則化現象の追試を行う。以上が本研究の目的である。

## 1.2 本論文の構成

本論文の構成は以下のとおりである。第2では、英語の不規則動詞について言語学、認知科学、歴史的な背景を述べる。第3章では統計的なデータを対象にした不規則動詞の規則化に関する研究と、本研究の意義について述べる。第4章では提案モデルを構築するにあたって必要な知識について述べる。第5章では実際に提案モデルの説明を行う。第6章で実験を行うための設定について述べ、第7で実験結果と考察を述べる。最後に第8章で本研究のまとめを行う。

## 2 背景

本章では英語の不規則動詞の言語的、認知科学的背景、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・認知科学的背景

本節では英語の不規則動詞について言語学的な背景、人間が過去形を生成する仕組みに関する認知科学的な背景について述べる。

#### 2.1.1 英語の起源と時代区分

はじめに英語の起源と英語の時代区分について述べる。図 1 のような樹形図を作成するためには膨大な言語資料が必要である。樹形図の作成は言語資料を用いて語彙的、音韻的な比較を用いて言語の系統付けを行うことである [6]。Proto-Indo-European はそのような資料が豊富であるため以下のような系統付けが可能である。

堀田 [5] によれば英語は Proto-Indo-European を起源とするゲルマン語派言語である [5]。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。

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 語族は地球最大規模の語族であり、非常に多くの言語を含む [6]。歴史言語学者によれば、Proto-Indo-European の生まれた土地は東欧であるとされているが、はっきりと決定されるには至っていない。また Proto-Indo-European が形作られていく過程にも様々な説が存在する。一例を紹介する。約 5500 年前に中央ヨーロッパに入ってきた土器民族が最初の Proto-Indo-European 語族ではないかと言われている。その後、各地で方言化していき諸言語を形作るに至った。英語と同語派の言語にはドイツ語などが含まれる。

古英語以降の時代区分については以下のとおりである。英語の時代区分は、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 [4] を示す。

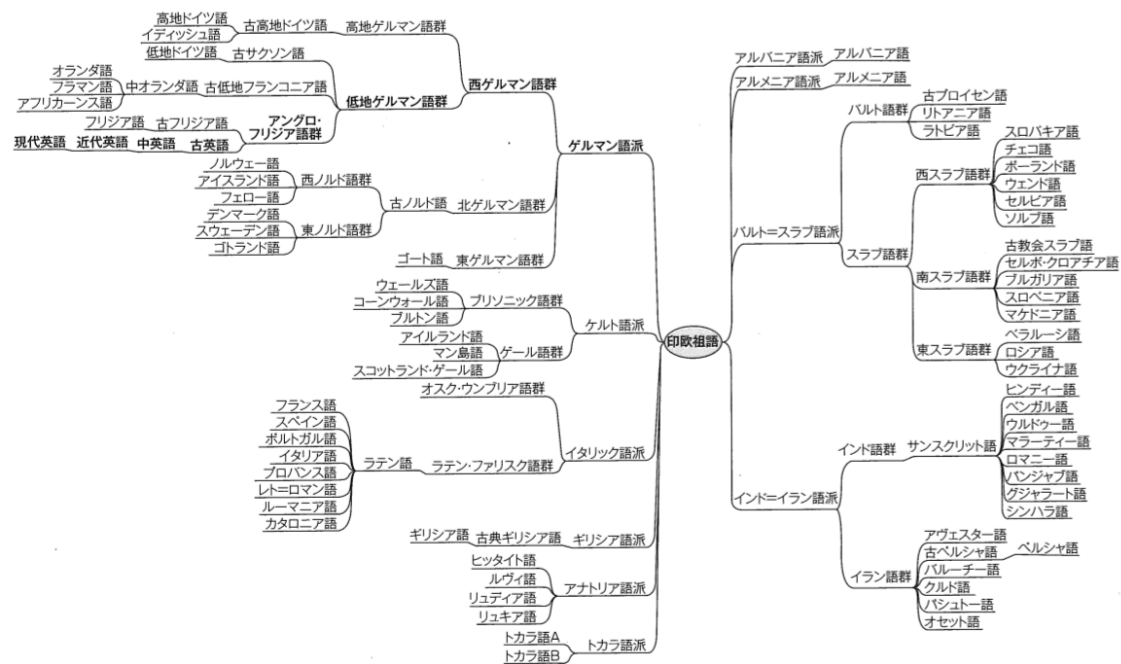


図 1 印欧祖語 (Proto-Indo-European) 族の系統図 [5] より引用

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 Midle 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期が Early Modern English 時代、アン女王の時代以降が Modern English 時代となる。表 1 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現在の不規則動詞のルーツとなる強変化動詞である。

### 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、そのルーツである強変化動詞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 [9]。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 1 形), *rið-on* (過去 2 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ここでは基本形のみ扱うものとする。表 2 に例を示す。

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(東ゲルマン語系の言語である)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。

- 過去形生成の仕組み

Jackendoff [12] 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

表 2 母音交替パターンによるクラス分け [9] 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..

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のである。

また Pinker らは Word and Rule Theory [13] を提唱した。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 を見つけら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。

- Novel Word の過去形生成

Huang ら [3] は Novel Word ([3] 内では人工的に作成した動詞を指す) の過去形生成と音韻的類似度と意味的類似度の影響を研究した。Huang らは既存の不規則動詞を模倣した人工的な動詞を作成した。ベースにした既存の不規則動詞は「swing, sink, lead, blow, bear, throw, read, cling」の 8 語である。例えば「swing」を真似た人工動詞は「spling」であり過去形は「splung」とな



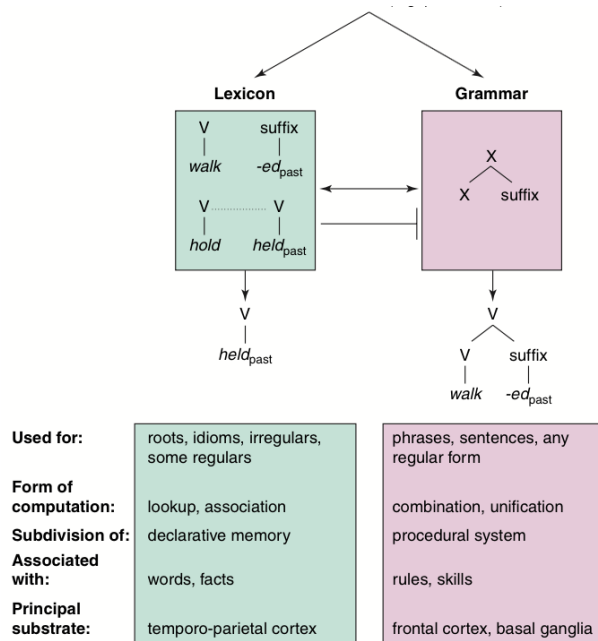


図 2 Word and Rule Theory[13] より引用

る。これらの人工動詞を用いた文章を作成し、不規則変化、規則変化どちらが受け入れやすいのかアンケート調査を行った。受け入れやすさは7段階評価で行なっている。また提示する文章は、表3のようにベースにした不規則動詞との意味の類似性を3段階に分けたものを使用している。

表 3 Level of Similarity (to “ throw ”) [3] より引用

Low	Moderate	Hight
Mike loved to <b>froe</b> elaborate meals for the most ordinary occasions.	The star goalie could <b>froe</b> the puck with any part of his body.	Sam spent the whole summer practicing how to <b>froe</b> a baseball.

統計的な解析を行ったところ以下の傾向が見られることが述べられている。

- 不規則変化の受け入れやすさ
  - － 既存の不規則動詞への音韻的類似度が増加すれば Novel Word の不規則な過去形を受け入れる傾向は増加する
  - － 既存の不規則動詞への意味的類似度が高いレベルでない限り Novel Word の不規則な過去形を受け入れる傾向は変化しない
- 規則変化の受け入れやすさ

- 既存の不規則動詞との音韻的類似度が増加すると不自然に感じる
- Novel Word の意味は関係ない

つまり、既存の不規則動詞と音韻、意味的類似度が高い場合 Novel Word にも不規則変化を適用する傾向がある。しかし、意味的類似度が高いだけでは適用する母音交替パターンを呼び出す事はできない。Novel Word の不規則な過去形を生成するためには音韻的類似度が高いことが重要になっていると言える。

### 2.1.3 膠着と屈折

本節の最後に屈折と膠着について解説を行う。屈折とは英語では、動詞の過去形や名詞の複数形などである。つまり、同一の語の様々な形を作ることである [7]。例えばラテン語は非常に多くの屈折を持つ言語である。ラテン語は文の中で屈折のみで単語の意味などが決まる。ラテン語の例を示す [7](長母音は母音を 2 つ書いている)。

- amaavii 私は愛した。「直接法、能動相、完了、一人称、単数」
- amaaistii あなたは愛した。「直接法、能動相、完了、二人称、単数」
- amaavismus 私達は愛した。「直接法、能動相、完了、一人称、複数」
- anaaveerunt 彼らは愛した。「直接法、能動相、完了、三人称、複数」

このように単語自体を変化させて意味を付加する。このような言語を屈折言語という。上記の例では「am」は「愛」という意味であることはわかるが、人称や完了を示す形態素がどれなのかを区別することが非常に困難である。そのため、融合言語と呼ばれることもある。

膠着も屈折と同様にある語の新しい形を作ることができるが、屈折と異なる点は語幹(語根)と接辞がはっきりと分かれていることである [6]。膠着を用いて文中の意味を決定する言語を膠着言語という。例として日本語の動詞を示す [7]。

- 食べさせられました。tabe(語根) - sase(使役) - rare(受け身) - masi(丁寧) - ta(過去)

語根の後ろにそれぞれの意味を表す接尾辞がついており、その接尾辞ごとに分離できる。

英語においては上記の屈折言語的特徴と膠着言語的特徴が混在している。言語類型論的には孤立語(中国語など)的特徴も混ざっているとされているが本研究では割愛する。孤立語的特徴は文に現れるため、動詞に着目する本研究では関連性が高くないためである。つまり英語の動詞では、不規則動詞は屈折言語的であり、規則変化する場合は膠着言語的であると考えられることもできる。

## 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口流入や言語接触の影響 [4, 8, 10] について述べる.

Old English はグレート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.D 400-550 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 [10]. 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.D 500-700 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 (北ゲルマン語群に属する古ノルド語) 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 に海賊の影響地域を示す. 図 3 より影響地域は東側に広がっていることがわかる. 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. その過程の中で古ノルド語は 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). 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 4 に同義語の交代の例を示す.

表 4 同義語の交代例 [10] より引用  
以前の表現 以降の表現

<i>preost</i>	<i>chapellain</i>	僧侶
<i>firð/grið</i>	<i>pais</i>	平和
<i>unfið</i>	<i>werre</i>	戦争
<i>lufu</i>	<i>cherite</i>	愛

表 4 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。

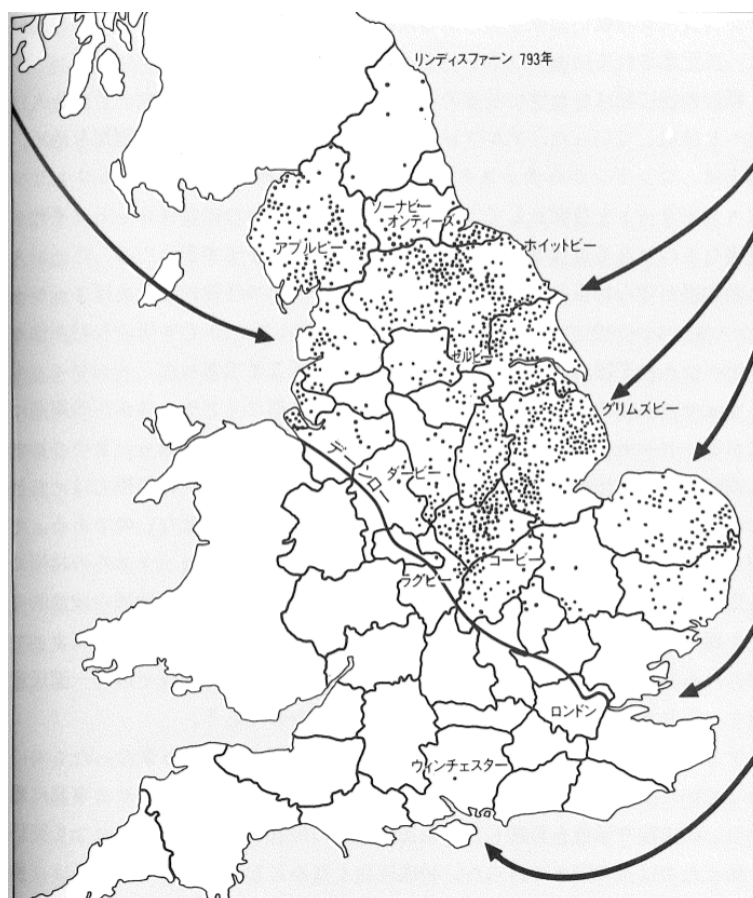


図3 ヴァイキングの影響地域 [11] より引用

### 2.3 背景のまとめ

本章では英語の不規則動詞について言語学的な背景と、不規則動詞の規則化に関する歴史的背景について述べた。言語学的背景では、英語の起源、時代区分とともに現在用いられている不規則動詞のルーツについて解説を行った。また、認知科学的背景では人間が過去形を生成する際の仕組みについても述べた。

歴史的背景では本研究で対象とする不規則動詞の規則化について、イングランドを例にその原因と考えられている言語接触や人口流入について述べた。次章では不規則動詞の規則化に関する統計的な研究とそれを踏まえた上で、本研究の意義について述べる。

### 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表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に、変化を推し進めるような 2 つの力 (内生的要因、外生的要因) 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

Lieberman ら [16] は、CELEX [18] を用いることで、不規則動詞の規則化現象を使用頻度 (コーパス内で) と時間の関数として表した。これにより過去から未来に渡る期間において不規則動詞の数を予測できるようになった。

Lieberman らは統計に使う動詞として、様々な文献から現在まで使用されている 177 の不規則動詞を抽出した。それらの動詞は Old English (A.D 800) では 177、Middle English (A.D 1200) では 145、Modern English (A.D 2000) では 98 に減少していくことを明らかにした。次に、それらの動詞の出現頻度を CELEX を用いて算出した。ここでの出現頻度は CELEX が持つ約 1770 万語の中に出現する動詞、約 331 万語を利用した相対的な出現頻度である。表 5 に不規則動詞と出現頻度を示す。

出現頻度が 1 に近づけば近づくほど、be, have, come などよく使用する単語が現れていることがわかる。これらの不規則動詞について年代別に各出現頻度にいくつの不規則動詞が存在するのかを調べた結果が図 4 である。y 軸が不規則動詞の数、x 軸が表 5 の出現頻度である。図 4 より不規則動詞は中程度の出現頻度を持つものが最も多く、低頻度、高頻度のものが少ない。また、規則化に関しては非常に高頻度な不規則動詞は規則化される傾向はない。約  $10^{-2}$  の出現頻度を持つ不規則動詞は Middle English から Modern English に至るまで全く規則化されていない。しかし、それよりも頻度が低い不規則動詞は年代を経るにしたがって大きく数を減少させている。特に  $10^{-5} - 10^{-4}$  の出現頻度を持つ不規則動詞の数はおよそ 50 から 10 程度までに減少している。つまり 40 あまりの不規則動詞が規則化されたことを示している。

次にある出現頻度を持つ不規則動詞の数が、時間経過によってどれだけ規則化されるかを微分方程式で表すことを考える。これは不規則動詞の規則化を物理的な現象 (例えば原子核の崩壊) として捉えることである。式 1 に原子核の崩壊を表す微分方程式を示す。N は原子核の数、 $\lambda$  は崩壊定数である。崩壊定数は微小時間内に、ある原子核が崩壊する確率である。

表 5 不規則動詞の相対出現頻度

出現頻度	動詞
$10^{-1} - 1$	be, have
$10^{-2} - 10^{-1}$	come, do, find, get, give, go, know, say, see, take, think
$10^{-3} - 10^{-2}$	begin, break, bring, buy, choose, draw, drink, drive, eat, fall, fight, forget, grow, hang, help, hold, leave, let, lie, lose, reach, rise, run, seek, set, shake, sit, sleep, speak, stand, teach, throw, understand, walk, win, write
$10^{-4} - 10^{-3}$	arise, bear, beat, bind, bite, blow, bow, burn, burst, carve, chew, climb, cling, creep, dare, dig, drag, flee, float, flow, fly fold, freeze, grind, leap, lend, lock, melt, reckon, ride, rush shape, shine, shoot, shrink, singh, sing, sink, slide, slip, smoke, spin spring, starve, steal, step, stretch, strike, stroke, suck, swallow, swear sweep, swim, swing, tear, wake, weave, wash, weep, weigh, wind, yell, yield
$10^{-5} - 10^{-4}$	bark, bellow, bid, blend, braid, brew, cleave, cringe, crow, dive, drip, fare, fret, glide, gnaw, grip, heave knead, low, milk, mourn, mow, prescribe, redden, reek, row, scape seethe, shear, shed, shove, slay, slit, smite, sow, span, spurn sting, stink, strew, stride, swell, tread, uproot, wade, warp wax, wield, wring, writhe
$10^{-6} - 10^{-5}$	bide, chide, delve, flay, hew, rue, shrive slink, snip, spew, sup, wreak

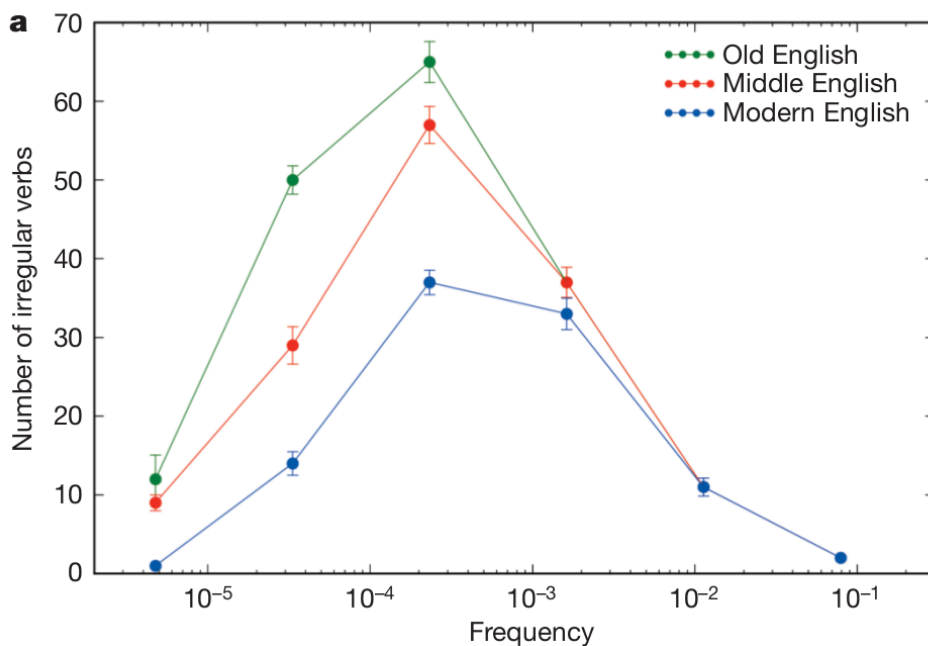


図 4 出現頻度ごとの不規則動詞数 [16] より引用

$$\frac{dN}{dt} = -\lambda N \quad (1)$$

不規則動詞の規則化に関しては  $N$  は不規則動詞の数であり、また崩壊定数のようなパラメータを設定することになる。以降では式 1 と同じ形の不規則動詞の規則化を表す式を作っていく。

原子核の崩壊と不規則動詞の規則化を同様に扱うと以下の考え方が可能になる。表 5 の各頻度に属する不規則動詞の数を原子核の数だと考えてみる。例えば  $10^{-6} - 10^{-5}$  の場合は 12 である。原子核の崩壊を考える場合、物質固有の崩壊定数により半減期 (原子核の数がちょうど半分になるまでの時間) という時間が決まっている。ここでは、物質は不規則動詞の出現頻度だと読み替える。つまり、出現頻度によって固有の崩壊定数が決まっており半減期 (不規則動詞の数がちょうど半分になるまでの時間) が決まっていると考えることができる。

また、原子核の崩壊を考える際に物質固有の崩壊定数 ( $\lambda$ ) を知る必要がある。不規則動詞においても頻度固有の崩壊定数を求めることになる。以降、頻度固有の崩壊定数を [16] に従い Regularization rate (規則化率) と呼ぶことにする。図 5 上には図 4 で示した不規則動詞の数を各年代の位置に示してあり同じ出現頻度を破線で結んである。図 5 下はその破線の傾きであり、出現頻度を  $\omega$  とすると式 2 で表される。

$$d(\omega) \approx \frac{2.133 \times 10^{-6}}{\sqrt{\omega}} \quad (2)$$

図 5 上の  $z$  軸 (Number of irregular verbs) は対数スケールになっている。つまり、破線がほぼ直線になっているという事は指数関数的に不規則動詞が減少している (規則化が進んでいる) ことを示している。さらに式 2 より減少具合を表す指数関数は出現頻度のみに依存し、出現頻度の 2 乗根に反比例する。直感的に述べれば出現頻度が  $1/100$  になれば規則化は 10 倍早く進むことになる。また年代を通して一定である。頻度のみに依存し、一定であるという理由から  $d(\omega)$  は原子核の崩壊定数と同じ扱いができる。よって頻度  $\omega$  の規則化率を  $d(\omega)$  と定めることができる。式 1 の  $N$  に対応する不規則動詞の数は頻度  $\omega$  と時間 (年代)  $t$  の関数  $I(\omega, t)$  として表すものとする。

以上を原子核の崩壊の微分方程式にならい、組み合わせると式 (3) ([17] に記載) になる。式 (3) は微小時間  $\partial t$  で減少する不規則動詞の変化量  $\partial I$  を示す。

$$\frac{\partial I}{\partial t} = -d(\omega) I(\omega, t) \quad t, \omega > 0 \quad (3)$$

式 (3) を解くことで  $I(\omega, t)$  を求めることができる。よって  $I(\omega, t)$  がわかれば全頻度の不規則動詞の数が時間の関数として表すことができるということである。しかし、具体的な解は [16] に



は示されていない. そのため, Supplementary information [17] で提供されているデータとプログラムから数値を抽出し擬似的な解として  $I(\omega, t)$  を導出した. 詳しくは 3.3 で述べる.

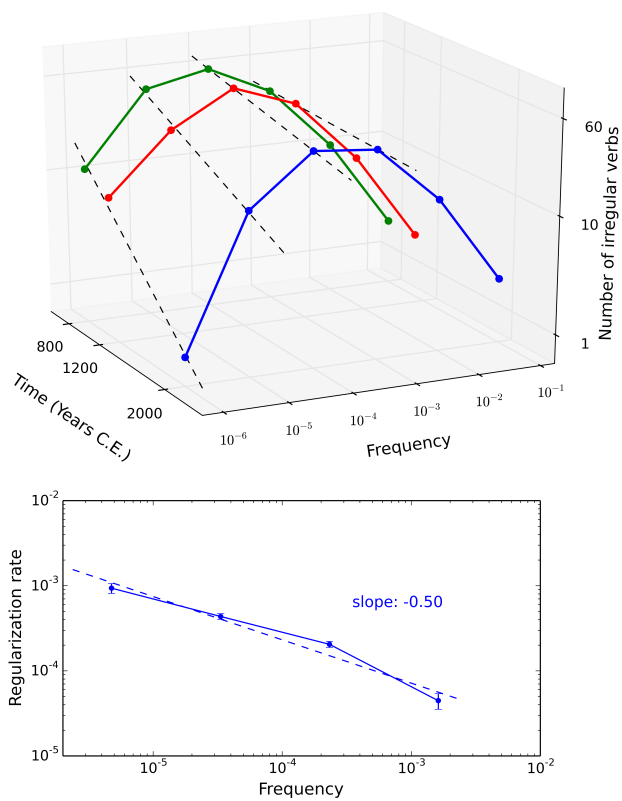


図 5 不規則動詞の減少時間と減少具合

### 3.2 語形変化に影響を及ぼす力に関する研究

言語変化はしばしばシグモイド関数 (S-curve) のように進行すると言われる [19]. (言語変化すべてが必ず S-curve のように進行するわけではない). 図 6 は Innovation 段階において新しい言語が生まれ、時間経過とともに人々に選択され (Select and Propagation)、定着 (Fixation) していく様子をまとめたものである. Ghanbarnejad ら [20] は、正書法の改正、不規則動詞の規則化などを S-curve として捉え、その変化を引きをこす力 (内生的要因、外生的要因) を数値的に明らかにしたものである. 内生的要因とは主に人間同士がコミュニティ内で接触することであり、外生的要因とは変化を推し進めようとする外部からの力のことである. それぞれの要因をパラメータ化し、式 (4) とする. パラメータは外生的要因を  $a$ 、内生的要因を  $b$  としている.

$$\frac{d\rho(t)}{dt} = (a + b\rho(t))(1 - \rho(t)) \quad (4)$$

式 (4) の一般解を実データに近似することでパラメータ  $a, b$  を取り出す [20].

次に近似に用いられている統計データについて述べる. Ghanbarnejad らが対象にした言語変化はドイツ語の正書法の改正 [21] と、不規則動詞の規則化である. ドイツ語の正書法の改正について説明する. 1996 年にドイツ語の [ß] を [ss] と綴るとする改正が行われた. すべての [ß] が改正されたのではなく長母音の後ろに [ß] が来る場合に限りそのまま使用されている. 不規則動詞の規則化はこれまで述べてきたように、母音交替で過去形を生成していたものが接尾辞「ed」をつけて過去形を生成するように変化することである. これらの変化を Google Ngram Corpus [22] (1800 年から 2000 年までに出版された電子書籍データ約 3610 億語) で式 (5),(6) を用いて 1800 年から 2000 年までの各年の統計をとったものである. 式 (5),(6) の  $freq$  は出現頻度を意味する.

$$\rho(t) = \frac{freq(ss)}{freq(ss) + freq(\text{ß})} \quad (5)$$

$$\rho(t) = \frac{freq(regular)}{freq(regular) + freq(irregular) + freq(past\ participle)} \quad (6)$$

以上の式は  $t$  年における変化した形の割合を表している. 不規則動詞の規則化を扱う場合は過去分詞形も入っている.

以上の結果を図 7 に示す. 青のドットが実データ、オレンジの曲線が近似曲線である.  $a, b$  の値からドイツ語の正書法改正においては、大きな外生的要因によって急激な変化がもたらされている

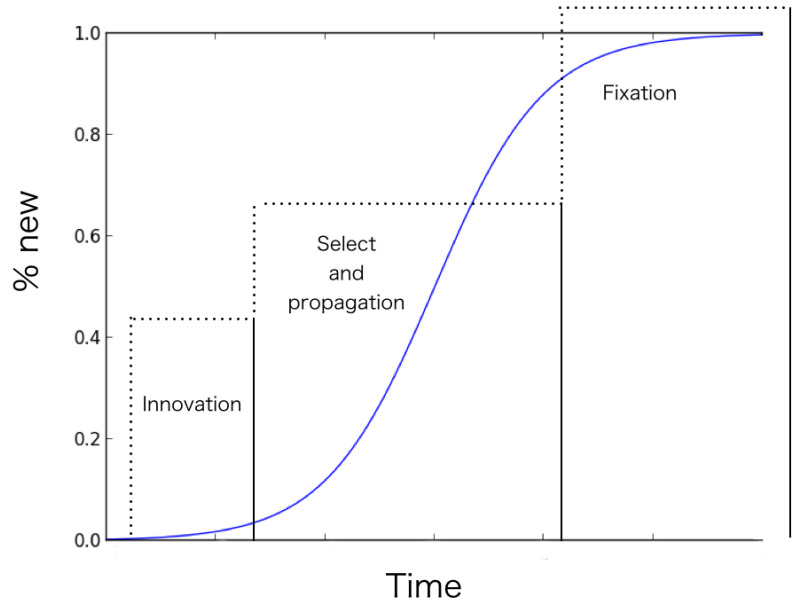


図 6 言語変化と S-curve

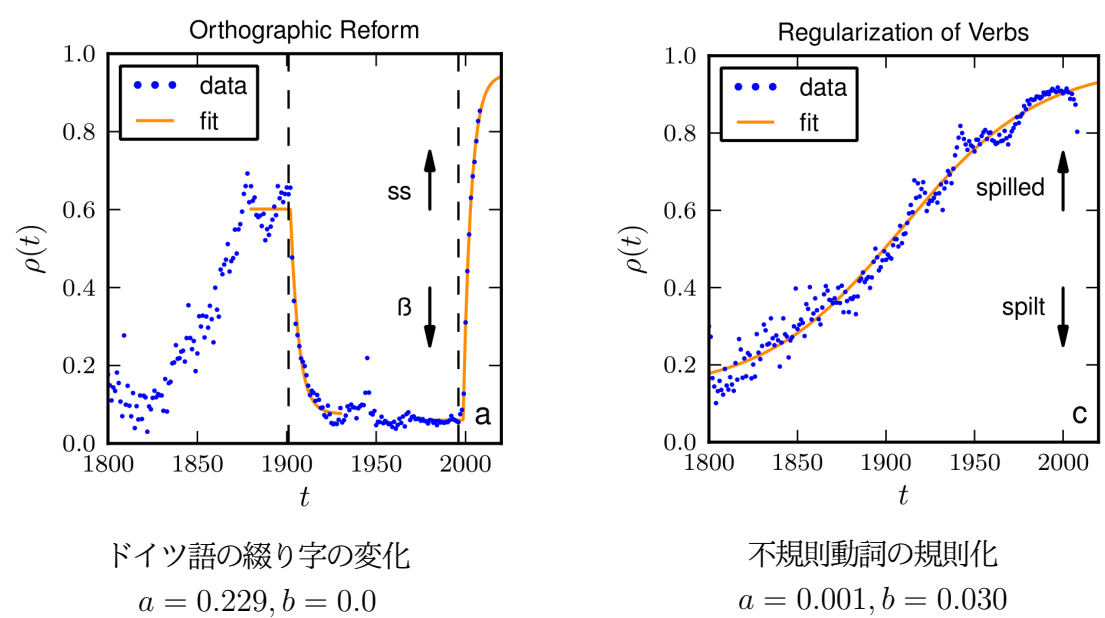


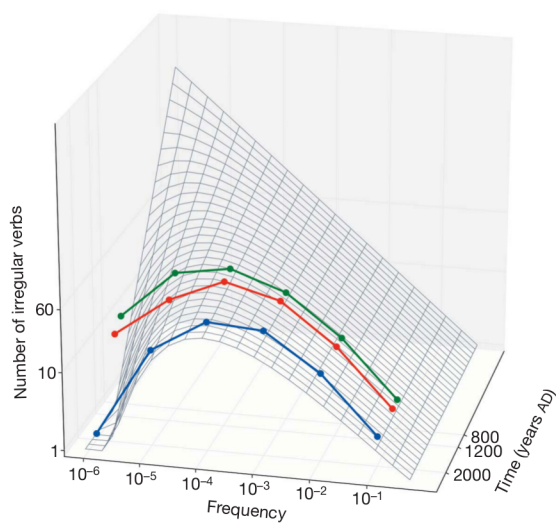
図 7 1800 から 2000 年の各語形の変化 [20] より引用

る. 逆に内部で変化を進めようとする働きは小さい. 不規則動詞の規則化においては, 外生的要因は非常に小さく, コミュニティ内部で変化を進めようとする内生的要因が大きいことがわかる.

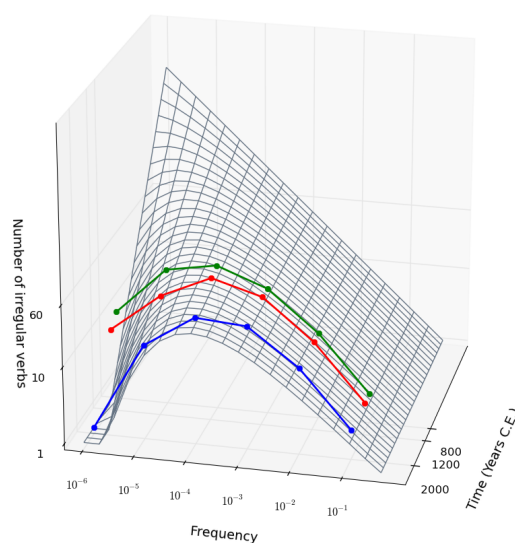
### 3.3 不規則動詞の規則化式の導出

3.1 節で述べた Lieberman ら [16] の研究において式 3 の解は図 8(1) の網目状の部分になると述べられている. しかし, 網目の部分 ( $I(\omega, t)$ ) がどのような式になるかは述べられていない. そのため, 本研究で擬似的な  $I(\omega, t)$  の式の導出を行った. 以下にその式と図 8(2) にプロットした結果を示す.

$$I(\omega, t) \approx \frac{0.4467 \times \exp\left(\frac{-4.9045 \times 10^{-6} \times (2000+t)}{\omega^{0.5088}}\right)}{\omega^{0.7099}} \quad (7)$$



(1) 先行研究 [16] による出力



(2) 式 7 を用いた出力

図 8  $I(\omega, t)$  ( $-2000 \leq t \leq 2500$ ) による不規則動詞の分布

図 8 より, Lieberman [16] らのグラフと同様のグラフが得られることがわかる.

### 3.4 先行研究のまとめとシミュレーションの意義

本章では不規則動詞の規則化に関する統計的な研究について述べた。Lieberman [16] らは不規則動詞の規則化を物理的な現象としてとらえ、コーパスにおける出現頻度と時間の関数として不規則動詞の数が過去から未来すべての時点で予測可能であることを示した。

Ghanbarnejad[20] らは言語変化を S-curve としてとらえ、変化を推し進める内生的要因と外生的要因を数値的に明らかにした。不規則動詞の規則化は外生的要因は小さく、内生的要因が大きいという結果を示した。

しかし、これらの研究は過去の統計的なデータが導く結果であり、本来言語変化とは人間のコミュニケーションや言語的な特徴が絡み合う現象である [12]。つまり、統計的なデータだけでなく現実世界を模したシミュレーション (言語変化と人間のコミュニケーションの関係を組み込んだ) を行う必要がある。また、Ghanbarnejad ら [20] の研究から人間同士のコミュニケーション (内生的要因) を考慮したシミュレーションを行うことが妥当である。

シミュレーションで扱う言語的な特徴とは、第 2 章で述べた屈折、膠着、また出現頻度である。本研究でシミュレーションを行う際の仮定を以下に示す。コミュニケーションに関してはモデルの設定を行う際に詳しく述べる。

- 高頻度の動詞は屈折 (不規則変化) を保つ、これは be, have のように短く発話し、屈折により語形を大きく変化させるほうが、似ている語と意味を取り違える確率が少ないし、発話自体が楽であると考えられる。
- 逆に頻度がそれほど高くない動詞は膠着 (規則変化) 的になっていく。実際にラテン語は、複雑な屈折を諦めるに至った [23]。同様に英語の動詞でも屈折が失われていく。

以上より本研究では Lieberman [16] らの示した規則化の状況 (高頻度の不規則動詞は生き残り、低頻度の不規則動詞は規則化されていく) を追試するためのマルチエージェントシステムを構築する。その後、人口流入についての考察も行う。

## 4 実験に必要な知識

本章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、またマルチエージェント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、単語の音韻的距離を計測する MetaPhone アルゴリズムについて述べる。MetaPhone アルゴリズムはシミュレーションにおいて主にエージェントの初期状態の知識を決定するために用いるものである。

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 マルチエージェントとは

エージェントの定義は山本ら [24] によれば「環境 (environment) の状態を知覚し、行動を行うことによって、環境に対して影響を与えることができる自律的主体である。」とされている。本研究においてもエージェントとは上記の定義を満たすものとして考える。自律的主体であることは、エージェントは人間やソフトウェアなどそれ自体で 1 個体をなすものであり、自身の経験と環境の情報の両方に基づいて意思決定や行動ができるということである (図 9)。

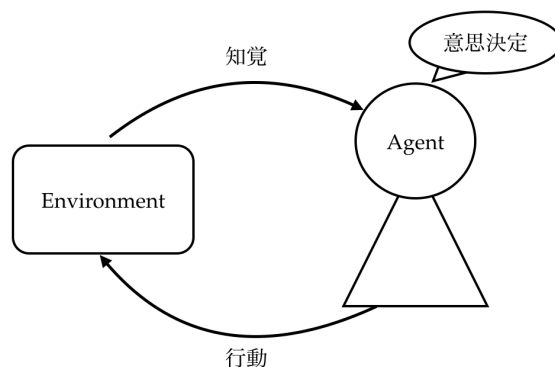


図 9 エージェントと環境

また、図 9 のようなエージェントが複数存在する状況をマルチエージェントシステムという。マルチエージェントシステムでのエージェントは上記の自律的主体性の他にも、他のエージェントと通信できるような社交性を持つ必要がある。これは複数のエージェントと協力して効率よく目的を達成するためである。本研究ではこのマルチエージェントシステムを利用しエージェント同士が会話をしながら、共通の知識を獲得していくシステムを構築する。

## 4.2 複雑ネットワークと言語伝搬シミュレーション

4.1 ではエージェントについて述べたが、本節では本研究のシミュレーションに使用する複雑ネットワーク [26, 27] について説明を行う。複雑ネットワークは本研究におけるマルチエージェントが置かれる「環境」である。また、複雑ネットワークを使った研究として主に社会言語学で行われている言語変化の伝搬シミュレーション [25] について述べる。

### 4.2.1 複雑ネットワーク

複雑ネットワークとは図 10 のように点が線で結ばれたものである。ネットワーク上の点をノードといい、線をエッジという。ノードを人、エッジを関係とすれば人間関係の社会ネットワークであり、ノードをコンピュータと考えればインターネットである [26]。ネットワークはその構造や性質によりレギュラーネットワーク、ランダムネットワークなど様々な種類が存在するがここでは本研究で用いるスケールフリーネットワークについて述べる。

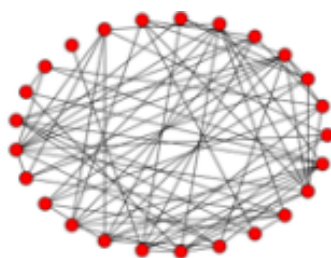


図 10 複雑ネットワーク

はじめに複雑ネットワークを特徴づける概念の一つである次数について説明する。次数とはあるノードからでるエッジの数である。次数は各ノードで異なり、次数が  $k$  であるノードが全ノードに占める割合を  $p(k)$  と書く。全ノード数を  $N$  とすれば、最大の次数は多重辺やループを許さない場合、自分自身以外のすべてのノードとつながっていることであるから  $N - 1$  である。よって  $\{p(k)\} \equiv \{p(0), p(1), p(2), \dots, p(N - 1)\}$  であり、これを次数分布と呼ぶ。上記の式はあるノードが次数  $k$  を持つ確率が  $p(k)$  であることを示す。スケールフリーネットワークのスケールフリーとは次数分布がベキ則になるネットワークモデルである。ベキ則とは  $p(k) \propto k^{-\gamma}$  となることを示す。つまり、非常に大きい  $k$ (次数) を持つノードは非常に少なく、小さな次数を持つノードは非常に多いということである。

次にこのようなスケールフリーネットワークの作成方法 (BA モデル [28]) について述べる。BA

モデルは小規模なネットワークからノードとエッジを追加 (成長) し、追加されるノードのエッジを次数の大きなノードに優先的につなげる (優先選択) という操作を繰り返して生成する. 図 11 にアルゴリズムの概略を示す.

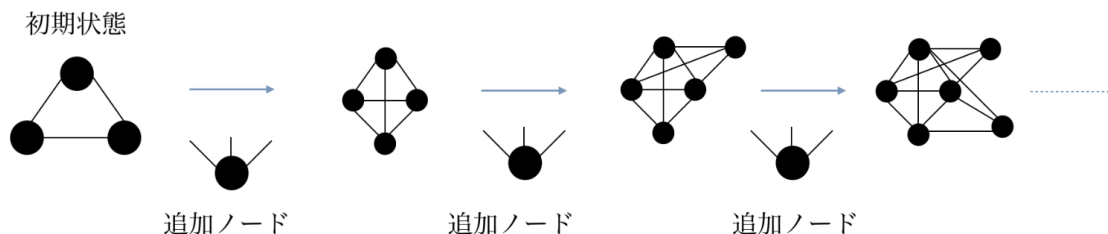


図 11 BA モデルの流れ

アルゴリズムについて説明する.

1. 初期状態で  $m_0$  個のノードを持つ連結なネットワークを作成する. 例えば, 図 11 のように 3 つのノードがすべてつながっている状況である.
2.  $m (< m_0)$  本のエッジを持つノードを一つずつネットワークに追加. この際新しいエッジが既存のノードにつながる確率を以下のように定める. ノードが  $N'$  個あり, 既存のノード  $v_i (1 \leq i \leq N')$  の次数を  $k_i$  とする. 新しいエッジのそれぞれが  $v_i$  に結びつく確率を

$$\Pi(k_i) = \frac{k_i}{\sum_{j=1}^{N'} k_j}, (1 \leq i \leq N') \quad (8)$$

とする. つまり次数が大きいノードほど新しいエッジを受け取りやすくなる.

3. 必要な頂点数になるまで 2 を繰り返す. 図 12 に 1000 ノードの BA モデルを示す.

以上が本研究で用いるスケールフリーネットワーク (BA モデル) である. 上記で示したように BA モデルは成長するモデルであり, 人口流入によってコミュニティが拡大していく様子を表しやすいと考えたためこのモデルを採用した. また BA モデルの特徴として平均距離の短さがあげられる. 平均距離はあるノードから任意のノードまでのノードの数の平均であるが BA モデルの場合, 以下の式で与えられる.

$$L \propto \log N \quad (m = 1) \quad (9)$$

$$L \propto \frac{\log N}{\log \log N} \quad (m \geq 2) \quad (10)$$



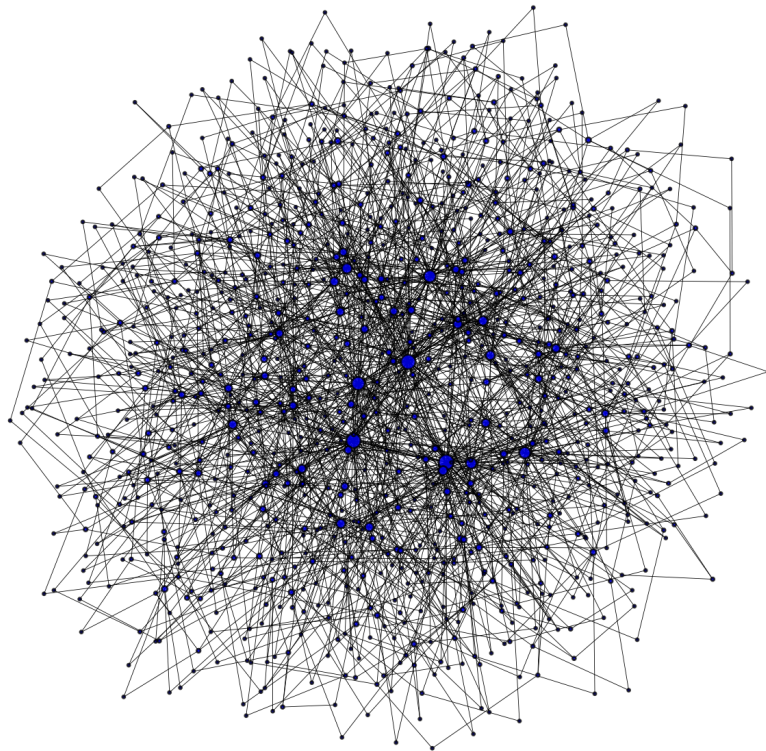


図 12 BA モデルで生成した複雑ネットワーク

平均距離がノードの数  $N$  の  $\log$  に比例することから、ネットワークサイズが大きくなっても平均距離はほとんど変化しないことがわかる。

#### 4.2.2 言語伝搬シミュレーション

本節では人間社会の中で新しい言語が広がっていく様子をシミュレーションした研究 [25] について述べる。ここで扱う言語とは英語、日本語など特定の言語や文法を指すのではなく、U (古い言語), C (新しい言語) といった非常に単純化された記号である。Ke による [25] では、4.2.1 で解説したスケールフリーネットワークをはじめ、様々な複雑ネットワークを用いて人間関係ネットワーク (コミュニティ) を仮定した。つまりノードは人間 (以降エージェントと呼称する) であり、エッジはエージェント同士の関係である。各エージェントはエッジが繋がっていればコミュニケーションが可能である。コミュニケーションは上記の単純化された記号の受け渡しである。また、エージェントは学習機構を持ち、受け取った情報により自分の知識を更新する力を持つ。ただし、エージェントには年齢 (1-5 の 5 段階) が設定してあり、学習できるのは年齢 1,2 (子供) だけである。年齢 1,2 で獲得した言語 (U or C) を以降の年齢 3-5 時のコミュニケーションで使い続けることになる。これは新しい言語が子供の言語獲得によって定着する様子を模倣したものである。エージェントは一定数のコミュニケーションを行った後、世代交代する。世代交代は年齢を一つ上げることが示す (5 に達した場合は 1 に戻る)。世代交代後もエージェント同士のネットワーク構造は変化しない。このように複雑ネットワークとエージェントによる学習を組み合わせたモデルを用いれば、コミュニケーション、学習、世代交代を繰り返すことによってコミュニティの中で新しい言語が広がっていく様子をシミュレーションすることができる。本研究ではこのようなモデルに学習機構として 4.4 で述べる遺伝的アルゴリズムを用いる。

### 4.3 MetaPhone アルゴリズム

本節では実験に用いる単語の音韻的距離 (本研究では発音の近さと定義する) を計測するために用いた MetaPhone アルゴリズム [29, 30] を解説する。MetaPhone アルゴリズムは単語等の文字列を発音記号に似せた擬似文字列に変換するものであり、完全な発音記号に置き換えられるものではない。例えば「spring」を「SPLN」と変換する。なお、MetaPhone は現在多言語対応したバージョン 3 まで存在するが本研究では英語のみに適用できればよいので、バージョン 2 のオープンソースプロジェクト [31] を利用する。以下に基本的 (バージョン、実装によって異なる場合がある) なアルゴリズムを示す。アルゴリズム中ではアルファベットはすべて大文字で表記して

いるが、入力文字列は小文字でも大文字でもよい。一文字ずつ入力文字列を読み込みながら以下のルールに当てはまるものを適応していく。

使用する擬似文字列

子音 : B X S K J T F H L M N P R O W Y

母音 : A I U E O

## アルゴリズム

1. C を除いて 2 つ重なっている文字の 2 番目を削除する
2. もし単語が KN, GN, PN, AE, WR で始まっていたら最初の文字を削除する
3. もし B が M の後に現れてかつ単語の終わりであれば削除する
4. もし CIA, CH のとき C を X に変換する. CI, CE, CY のとき C を S に変換する. それ以外のとき C を K に変換する. (アルゴリズムによっては SCH のとき C を K に変換する)
5. DGE, DGY, DGI のとき D を J に変換する. それ以外のとき D を T に変換する
6. GH(H が単語の最後か母音の前でない場合), GN, GNED のとき G を削除する. I, E, Y の前か GG でないとき G を J に変換する. それ以外のとき G を K に変換する
7. 母音の後でないかつ母音が続かないとき H を削除する
8. CK を K に変換する
9. PH を F に変換する (アルゴリズムによっては PH の P を F に変換する)
10. Q を K に変換する
11. SH, SIO, SIA のとき S を X に変換する
12. TIA, TIO のとき T を X に変換する. TH のとき T を O に変換する. TCH のとき T を削除する
13. V を F に変換する
14. 単語が WH で始まっているとき W を削除する. W が母音に続かない場合 W を削除する

15. 単語が X で始まる時 X を S に変換する. それ以外の時 X を KS に変換する
16. 母音が続かないとき Y を削除する
17. Z を S に変換する
18. 単語の始めでない母音をすべて削除する

変換例を示す.

変換前 ALEXANDRE → ALEKSABTRE → ALKSNTR 変換後  
 X を KS に変換 → D を T に変換 → 先頭以外の母音を削除

上記アルゴリズムを適応後の文字列の編集距離を図ることで音韻的距離とする. 例えば「spling」と「sprink」の音韻的距離は 1 であり、「spling」と「cled」の音韻的距離は 4 となる. 以下距離と表記した場合は、音韻的距離を指す.

#### 4.4 遺伝的アルゴリズム

遺伝的アルゴリズム (GA) は環境に適応して進化する生物のダイナミクスを模倣したアルゴリズムである [15]. GA は制約の中で膨大な数の組み合わせが存在する解空間の中から実時間内に最適に近い解を求める最適化問題などに用いられる. 求める解の候補を遺伝子を保有する個体 (遺伝子の組み合わせ) として計算機上に表現し、生物における選択、交叉、突然変異といった遺伝的操作を行い遺伝子を解に近づけてくという方法である.

GA の遺伝子は、表現型と遺伝子型からなる. 表現型は生物で言えば目で見ることができる特性であり、遺伝子型はその特性を表す数値, 文字列, 木構造などである. 交叉、突然変異はこの遺伝子型に作用する. また表現型が環境の中でどれだけ適応できているかが決定され、その適応具合 (適応度) によって次世代に残される個体が選択される.

図 13 にしたがって GA の処理の流れについて詳しく説明する.

##### 1. 初期集団の生成

ランダムな遺伝子を持つ個体を必要な数 (N 個体とする) だけ生成する.

##### 2. 評価

各個体の適応度が求められ終了条件のチェックが行われる. 終了条件は条件を満たす個体が見つかったか、または一定の世代数まで到達したかどうかを基準となる.

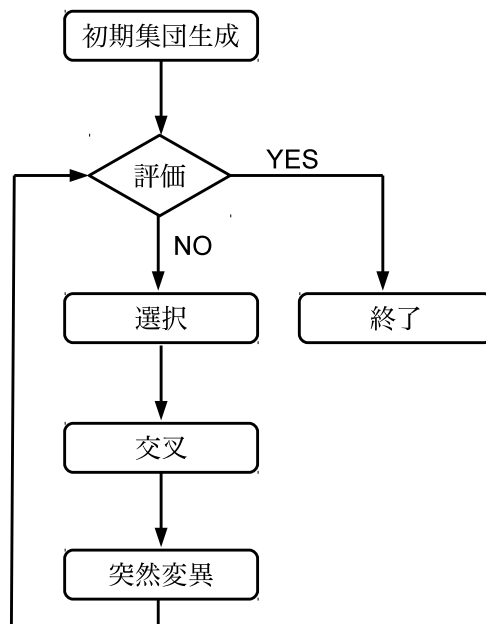


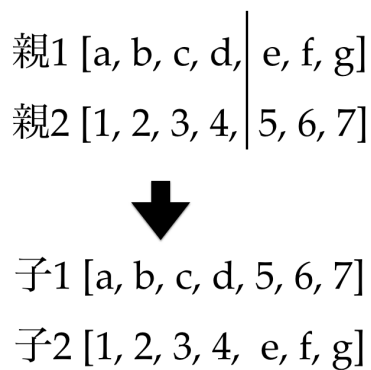
図 13 遺伝的アルゴリズムの処理の流れ

### 3. 選択

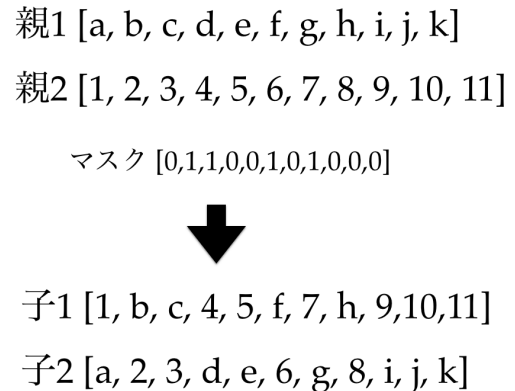
2. で求めた適応度を利用して次世代に残す個体を  $N$  個体選択する. GA における基本的なアイデアは適応度の高い個体を残すことである. 実現するためには様々な選択方法が存在するが、ここでは本研究で使用したトーナメント選択について述べる. トーナメント選択は個体集団の中から、ある個体数をランダムに選び出し、その中で一番良い個体を選択する方法である. ランダムに選ぶ個体数をトーナメントサイズと呼び、トーナメントサイズにより様々な種類が存在する. トーナメントサイズの変更により、個体集団が漸に収束していく速度のコントロールが容易であるため、この方法を採用した.

### 4. 交叉

交叉は選択された 2 つの個体 (親) から子の遺伝子を作る操作である. 親の遺伝子の交叉点を決めて、子供に引き継がれる遺伝子は交叉点の前後どちらかになる  $n$  点交叉と、ランダムなビットマスクを用いる一様交叉がある. 図 14 に概要を示す.



(1) 一点交叉



(2) 一様交叉

図 14 交叉

一点交叉は図 14(1) の線の部分が交差点となり、子の遺伝子が生成される. 一様交叉はマスクのビットによって受け継がれる親の遺伝子が変わる. 子 1 の場合、マスクが 1 の時に親 1 から、0 の時に親 2 から遺伝子を受け継ぐ. 子 2 の場合は逆である.

### 5. 突然変異

突然変異は一定の確率である遺伝子を別の遺伝子に変化させる操作である. 例えばある個体の遺伝子が  $[0,1,1,0,1,1,1,1]$  の時、確率 0.05 で 2 番目の遺伝子を 0 に変える  $[0,0,1,0,1,1,1,1]$  という操作である.

## 4.5 実験に必要な知識のまとめ

本章では実験に必要な知識について述べた。マルチエージェントについて解説を行い、本研究においてマルチエージェントが置かれる環境である複雑ネットワークについて述べた。また、言語伝搬のシミュレーションについても解説を行い複雑ネットワークが言語学のシミュレーションにも活用できることを示した。最後に音韻的距離を測るための MetaPhone アルゴリズムと、エージェントの学習機構である遺伝的アルゴリズムについて述べた。

## 5 過去形獲得モデルの提案

本章ではこれまで述べてきたことをもとに、マルチエージェント環境において共通の動詞の過去形を獲得していくモデルを提案する。本モデルの目的はエージェントがコミュニケーションを行いながら3章で示した規則化を人工的な環境でシミュレートすることである。本モデルでは Pinker ら [14] が作成した人工動詞 60 語を利用する。エージェントたちは、初期状態では上記の人工動詞について特定の過去形を持たないが、時間経過とともに規則化が進行しつつも徐々に共通の過去形を獲得していくことが望ましい。初期状態で特定の過去形を持たないという仮定は、エージェントに過去形の選択肢を多く与えることであり、状況によって様々な過去形の創発を可能にするものである。規則化もその選択肢に含まれている。ただし、どんな人工動詞についても無作為に規則化を起こすわけではない。そこには人間のコミュニケーションの本質と言語変化に関わるパラメータが存在すると考える。

### 5.1 コミュニケーションの目標

本モデルを説明する前に、人間のコミュニケーションの本質について言及しておきたい。Jackendoff[12] はコミュニケーションの目標を以下の2つであるとしている。

1. 話し手と聞き手の双方にとって最小の物理的努力で意味を伝える  
つまり、話を短く省略した形にする
2. 意味をできるだけ明瞭に伝える  
つまり、話を冗長性のある形にする

1と2は互いに矛盾した関係にある。この矛盾は常に存在し、言語の歴史的変化を動機づける要因の一つである。さらに、個々の言語行為および諸言語の文法はこれらの2つの矛盾する目標を反映する形になっていると述べられている。本研究では、不規則動詞の発話でも同じ現象が起こっていると仮定する。これまで不規則動詞は英語の動詞の中では出現頻度が高いものが多いと述べてきた。会話に高頻度で出現する動詞は短い形(不規則変化のまま)にしたほうが話し手の負担を考えればコミュニケーション効率は良い。頻度が高い不規則動詞は記憶されている場合が多いと予想でき、短く明確に意味が伝えられる。逆に、短い形で意味が伝わらない場合は普遍的に適用できる規則変化を行うことが最も的確(単純)であると考えられる。しかし、規則変化を行えば確実に発話する動詞は長くなる(動詞の語尾の発音が t, d である場合音節も増える)。

本モデルは矛盾する2つのコミュニケーションの目標を反映し、高頻度の動詞における発話の長さのみに焦点を当てたものであることを強調しておく。



## 5.2 エージェントの知識

本節では提案モデルの解説入っていく．はじめにエージェントの知識について述べる．

本モデルは 60 語の人工動詞を用いる．Pinker [14] では各人工動詞に過去形も設定されているが本研究では初期化にのみ用いる．表 6 に人工動詞とその過去形を示す．原形-過去形のペアで表示する．

表 6 人工動詞

spling - splung	skring - skrung	sprink - sprunk	cleed - cled
preed - pred	queed - qued	cloe - cloo	froe - froo
plare - plore	quare - quore	fring - frung	ning - nung
frink - frunk	cleef - clef	preek - prek	queef - quef
foa - foo	voa - voo	jare - jore	grare - grorey
trisp - trusp	nist - nust	blip - blup	gleef - glef
keeb - keb	meep - mep	goav - goov	joam - joom
flape - flope	blafe - blofe	plip - plup	glip - glup
brip - brup	gloke - glook	proke - prook	greem - grem
pleem - plem	treem - trem	slace - sloce	nace - noce
brilth - brulth	glinth - glunth	plimph - plumph	ploab - ploob
ploag - ploog	smeeb - smeb	smeeg - smeg	smeej - smej
smaib - smobe	smaig - smoag	frilg - frulg	krilg - krulg
trilb - trulb	ploamph - ploomph	ploanth - ploonth	smeelth - smelth
smeenth - smenth	smeerg - smerg	smairg - smoarg	smairph - smoarph

表 6 の過去形を子音-母音-子音に分解する．子音と母音は深谷ら [32] に記載されている音読のシミュレーションに用いられた始端、終端子音と母音に対応する字素を参考にしたものである (補足資料)．始端子音を表す字素には s, p, t, k などの文字が含まれ、終端子音を表す字素には ch, m, n, ng などが含まれる．母音を表す字素には e, i, o, oo, oa などが含まれる．パターンマッチによる解析の結果、表 6 の過去形に含まれる母音は「e, o, u, ue, uo, oa, oo」の 7 パターンに区別される．エージェントは母音交替を行なって過去形を生成する場合は、この 7 パターンの中からランダムに選択する．その他に規則変化の「ed」を使用して過去形とすることもできる．

図 15 にエージェントの初期知識とコミュニケーションの方法について示した．図 15 の下部「エージェントの知識」にある円が母音交替のパターンである．エージェントがランダムに過去形の母音を選択することから、上記の円は母音ごとに任意の割合 (確率) を持つルーレットである．つまり、ルーレットを回してたまたま選択された母音を過去形の母音として用いるということである．次にこれらの母音交替の選択肢を初期知識としてエージェントに与える方法について述べる．

母音を選択するルーレットは 4.3 節で解説した MetaPhone アルゴリズムを利用して作成される．「spling」

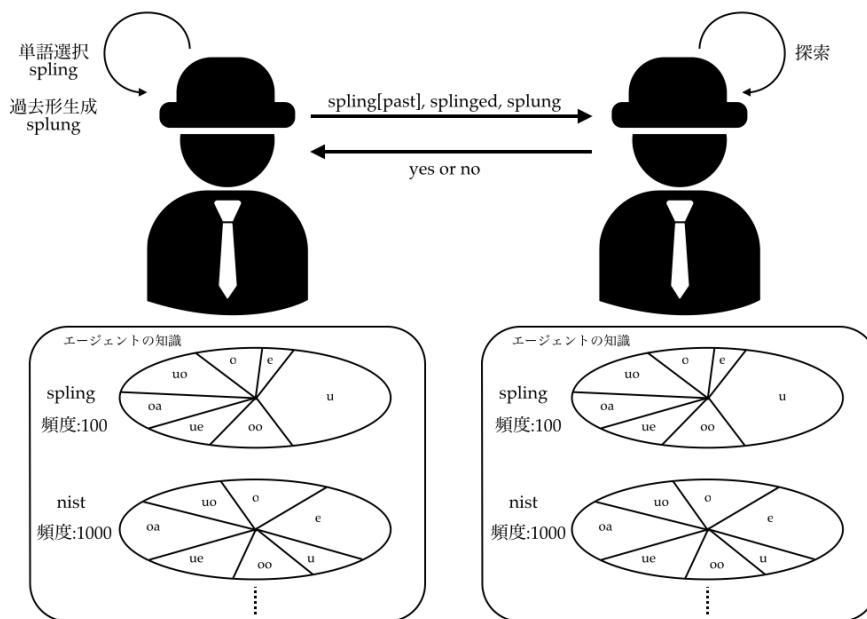


図 15 エージェントの知識とコミュニケーション

を例にとって説明する. まず「spling」と表 6 に示した全動詞の原形(「spling」自身は除く)との距離と、比較した動詞の過去形の母音を得る(「spling」「cleed」の距離は 4 で「cleed」の過去形「cled」の母音は「e」である). その後、過去形の母音ごとに「spling」との平均距離を求める(過去形の母音が「e」である動詞の原形をすべて集め「spling」とそれらの動詞との平均距離は 3 である. この操作を 7 つの母音すべてについて行う). つまり、ある母音があって、その母音を過去形に持つ動詞との平均距離を求めている. この平均距離の比にしたがってルーレットにおける母音の占める割合が決定される. このルーレットは動詞によって一意に決まっており、エージェントは使用する動詞すべてのルーレットを保持している. 初期状態では全エージェントの全動詞のルーレットは共通である.

また、3.1 節で解説したように規則化において動詞の出現頻度は考慮しなければならない重要な要素である. そのためこれら人工動詞にも出現頻度を設定する. 具体的には P13 表 5 における出現頻度を与える. ただし簡単のため  $10^6$  倍して整数に変換して使用する. 出現頻度とその出現頻度を持つ動詞の数を表 7 に示す. それぞれの出現頻度を持つ動詞の数が全動詞数に占める割合は表 5 とおおよそ同じである. これにより出現頻度が高い動詞ほどよく発話され、低い動詞はほとんど発話されないという状況を作り出すことができる. ただし、図 8 から明らかなように非常に高頻度の単語は規則化されるまでかなりの年月がかかる. よって本研究では出現頻度が 100000、10000 の動詞は規則化が起きないものとしエージェントの学習対象から除外することにする. これはエージェントの発話のする動詞の極端な偏りを避けるためにも必要である. 本モデルでは使用する動詞(出現頻度 100000、10000 の動詞を除く)のルーレットと頻度を合わせたものをエージェントの知識と呼ぶ.

表 7 出現頻度と人工動詞の数

出現頻度	人工動詞数
100000	1
10000	4
1000	12
100	22
10	17
1	4

### 5.3 マルチエージェント環境

次に上記の知識を使ったコミュニケーションについて述べる。各エージェントは 4.2.1 節で解説した BA モデル上に配置される。複雑ネットワークのノードが一人のエージェントを表し、エッジがつながっていればコミュニケーション可能である。エージェントコミュニケーションは以下の流れで行われる。

- あるエージェント (発話エージェント) が以下の操作にしたがって、他のエージェント (聞き手エージェント) に任意の動詞の過去形を発話する。
  1. 発話エージェントは出現頻度に従い、発話する動詞 (原形) を一つ選ぶ。
  2. 発話する動詞のルーレットから、ランダムに母音の一つを選ぶ (この母音が過去形の母音と言う意味である)。
  3. エッジがつながっているエージェントから、ランダムに聞き手エージェントを選ぶ。聞き手エージェントに対して [動詞の原形、規則化した過去形、ランダムに選んだ母音] を発話する。ただし、シミュレーションの高速化のため実際に母音交替した過去形は発話せず、母音のみを扱うこととする。
- 発話を受け取った聞き手エージェントは以下の操作を行う。
  1. 発話内容を受信する。
  2. 受信した動詞の原形のルーレットを自身の知識の中から呼び出す。
  3. ルーレットからランダムに一つの母音を選ぶ。
  4. もし、自身が選んだ母音と受け取った母音が同じであれば yes を、異なれば no を発話したエージェントに返す
- 発話エージェントは自身の発話内容と、聞き手エージェントからの返答をもとに学習を行う。

ここで発話エージェントが規則化した過去形を同時に発話する状況は、母音交替で過去形を生成した場合に聞き手エージェントに伝わらず、普遍的に適用できる [ed] を使った過去形を生成し、なんとか意味を伝えようとするものである。Jackendoff [12] に示されているように、コミュニケーションの目標とは意味を

できるだけ明瞭に伝えることも含まれる。よって意味が通じないままコミュニケーションが終了する状況は考えにくい。つまり、本研究の規則化を使って意味を伝えようとする設定は不自然ではない。

## 5.4 学習機構の構築

本節ではエージェントが動詞の過去形を獲得する学習機構について述べる。本モデルではエージェントが学習することにより自分の持つ知識(動詞のルーレット)を変化させることができる。学習には4.4で解説した遺伝的アルゴリズムを用いる。学習はエージェントが持つ動詞ごとに行われる。

### 5.4.1 遺伝子の生成

本モデルにおける遺伝子は動詞のルーレットの特定の母音を表す。遺伝子は7bitのバイナリ文字列である。本研究ではルーレットは図16のように0から99の100個の整数であると考え、つまり各動詞について100個の遺伝子が与えられる。100個の遺伝子はルーレット上の母音の有する割合によって割り振られる。例えば図16の「e」を表す遺伝子は0から11を2進数に変換したものであり、「o」を表す遺伝子は12から32を2進数に変換したものである。100個の根拠であるが、動詞のルーレットを初期化する際にあまりにも遺伝子が少ない場合、各母音の有する割合がすべて同じになってしまうことを避けるためである。

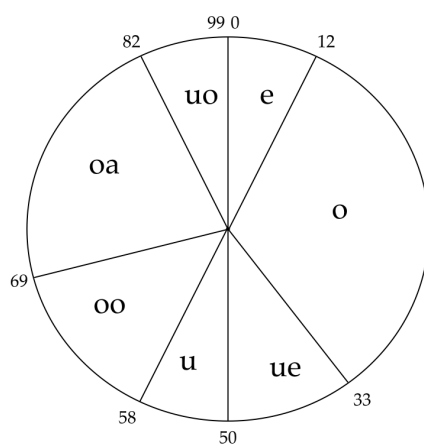


図 16 遺伝子

エージェントのルーレットからランダムに選んで発話するという事は、ランダムに遺伝子を選んで発話することに等しい。また、規則変化するための遺伝子を非常に低い割合で導入する。本研究では遺伝子数の1%から2%に設定してある。この割合はすべての動詞で共通である。適応度の初期値は全遺伝子で一定(0)である。

## 5.4.2 遺伝子の評価

これらの遺伝子の評価は発話エージェントのみによって行われる。聞き手エージェントもランダムに遺伝子を選んでいることになるが評価は一切行わない。

遺伝子の適応度は次の手順で決定される。また遺伝子は 102 個用いるが、遺伝子長は 7bit であるため、何も表現しない致死遺伝子が発生してしまう。致死遺伝子が選ばれた場合はエージェントは何もしない。

### 1. 基本的な評価

- 発話エージェントと聞き手エージェントの選んだ母音が同じであればその母音を表すすべての遺伝子に 10 点
- 発話エージェントと聞き手エージェントが選んだ母音が異なれば規則変化を表すすべての遺伝子に 10 点

### 2. 場合によって以下のペナルティを科す。

- 高頻度 (頻度が 1000) である動詞は規則変化した場合、 $-n$  点 (高頻度ペナルティ)
- 音節が増える場合 -5 点 (音節ペナルティ)
- 超高頻度 (頻度が 100000, 10000) の動詞と距離が 1 以下であれば -5 点 (類似ペナルティ)

### 3. これらの点数は任意の遺伝子が使用されるたびに計算され、世代の中で蓄積されていく。

基本的な評価で加点する 10 点はシミュレーションを繰り返し、意図する結果が現れるよう定めたものであり数値的な根拠はない。

次にペナルティについて説明を行う。高頻度である動詞は規則変化した場合に、 $-n$  点のペナルティを受ける。本研究ではこれをパラメータとして扱う。遺伝子に 10 点の適応度が加点されるという事は、その遺伝子で意味が通じたという事である。しかし、規則変化の場合は 5.1 で述べたように発話の負担が増えることになり、人間は規則変化が不規則変化に比べて 100% 使い勝手が良いと判断することはないと考えられる。つまり、人間の頭の中では、規則変化させれば意味は通じるが (規則変化が適当)、同時に煩わしさも感じているということである。その比を表現するものがペナルティである (図 17)。

図 17 のエージェントは規則変化させて意味が通じたとしても  $10 - n : n$  で実際に規則変化を使うのか、不規則変化を維持するのか自分の中で意識が分かれているのである。

音節が増える場合と超高頻度の動詞と距離が近い場合のペナルティも同様の現象を表現するものである。超高頻度の動詞と距離が近ければペナルティを与える仕組みは 2.1.2 で示した「既存の不規則動詞との音韻的類似度が増加すると不自然に感じる」ということを表現している。これらは高頻度の動詞だけでなく、全頻度に対して一定の確率で科される。本研究では確率は 0.4 とした。ペナルティを -5 点、確率を 0.4 とした理由であるが、高頻度ではないがある程度発話される動詞は規則化が進んでほしいという理由からである。ペナルティを大きくすれば全く規則化は起きないことが予想される。また、確率も同様の理由で低めの値に設定した。

このような学習機構を持つことによって、発話を通して動詞の過去形にどの母音を使うのか、または規則変化させるのか、を世代を経ることで決定できる。

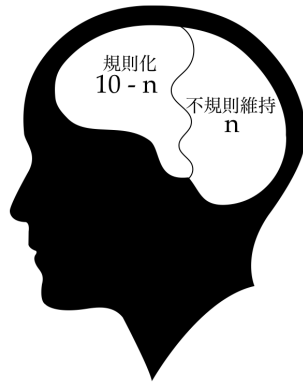


図 17 ペナルティの意味付け

## 5.5 実装

本モデルを Python [33] を用いて実装した。エージェントは独立したプロセスで表現し、コミュニケーションはプロセス間通信で行なっている。エージェントの知識はすべて共有メモリ上に配置してあり、プロセスセーフでアクセス可能である。また世代交代を行う際、GA の選択、交叉などのオペレータはエージェントの持っている動詞ごとに行うが、すべて非同期で行なっている。これらは CPU コア数が増えれば増えるほど高速に処理できる。

## 6 実験

本章ではこれまで提案してきたモデルの実験を行う。

- 実験 1

実験 1 の目的はエージェントに共通の動詞の過去形を獲得させること、また先行研究 3.1 の規則化現象を再現することである。具体的には出現頻度 1000 の動詞は少数の規則化が起き、それ以下の動詞は多くが規則化されるという状況を作り出すことである。そのためにパラメータ  $n$  (高頻度ペナルティ) に関してチューニングを行い有効な範囲を決定した。その結果、 $4 \leq n \leq 6$  とすることとした。

- 実験 2

実験 2 の目的はネットワークサイズが大きくなった時に規則化も含め、過去形が獲得されていく速度について考察することである。

表 8、図 18 に実験 1 の環境を示す。

表 8 実験 1 の設定

パラメータ	高頻度ペナルティ $n$
エージェント数 (ネットワークのノードの数)	12
BA モデルで追加されるノードの持つエッジ数	2
使用する人工動詞	60 語 (うち 5 語は超高頻度の動詞として除く)
人工動詞の出現頻度	表 7 に従いランダムに割り振る
発話数	エッジのつながっている各エージェントに対して 100 発話
母音を表す遺伝子数	各動詞ごとに 100
規則変化を表す遺伝子数	2
GA の世代	30 世代
出力	全動詞の全世代におけるルーレット上の割合の平均 (エージェントの知識の平均)

ただし、パラメータは  $n = 4.0, n = 5.5, n = 6.0$  の 3 パターンで行う。

実験 1 ではペナルティの値のみ変動するものとし、他の設定値はすべて固定した。また、初期化はすべてランダムで行われている。そのため実験を行うたびに人工動詞の頻度は異なる。また、使用するネットワーク構造はすべての実験で共通とした。全エージェントは最低 2 人のエージェントと繋がっている。

次に実験 2 の設定を表 9、図 19 に示す。高頻度ペナルティは 5.5 に固定し、エージェントの人数を 100 に増やした。また、全エージェントが最低 2 人のエージェントとつながっている点は実験 1 と同様である。ただし、実験が非常に長時間になるため各エージェントの発話数は 30 とした。

表 9 実験 2 の設定

高頻度ペナルティ	5.5
エージェント数 (ネットワークのノードの数)	100
BA モデルで追加されるノードの持つエッジ数	2
使用する人工動詞	60 語 (うち 5 語は超高頻度の動詞として除く)
人工動詞の出現頻度	表 7 に従いランダムに割り振る
発話数	エッジのつながっている各エージェントに対して 30 発話
母音を表す遺伝子数	各動詞ごとに 100
規則変化を表す遺伝子数	2
GA の世代	30 世代
出力	全動詞の全世代における規則変化を表す遺伝子の割合

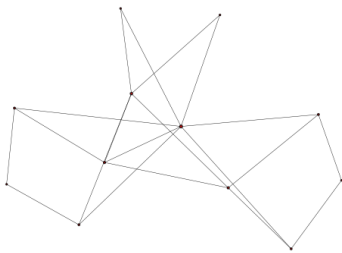


図 18 実験 1 で用いるネットワーク

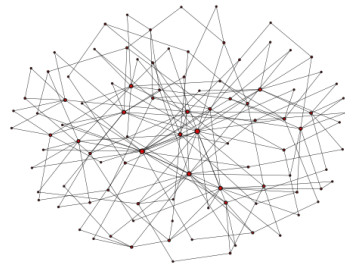


図 19 実験 2 で用いるネットワーク



## 7 実験結果と考察

前節での設定をもとに実験を行った。

### 7.1 実験 1

実験 1 の結果と考察を出現頻度ごとに示す。また、高頻度ペナルティの大きさごとに超高頻度の動詞と、類似ペナルティの対象となる動詞を表 10 に示す。

表 10 類似ペナルティ対象の動詞

パラメータ	超高頻度の動詞	類似ペナルティ対象の動詞
4.0	glinth, fring, trilb, grare, foa	goav, cleef, quare, greem
5.5	queed, cloe, ploag, meep, ploamph	voa, nace, foa, goav, quare cleed, cleef, keeb, plare, ploab, preek blafe, queef
6.0	blip, proke, ploanth, flape, ning	plip, ploag, plare, ploab, blafe glinth, preed, preek

#### 7.1.1 出現頻度 1000 の場合

出現頻度 1000 の動詞は 12 語あるが、すべての結果ではなく一部を抜粋したものである (その他の結果は付録を参照)。図の縦軸 (Activity Ration) は、その母音のコミュニティ内での使用率である。動詞ルーレット上の各母音の割合 (規則変化させる遺伝子も含む) の平均と同じ意味である。横軸は GA の世代である。世代を経るにしたがって特定の母音の使用率が上がっているという事は、別の母音を表す遺伝子が淘汰されコミュニティ内で使用される過去形が決定していく様子を表す。

- 高頻度ペナルティ 4.0 の結果

図 20 では「ning」の過去形の生成に、初期状態でもっともルーレット上の割合が高い「o」すなわち「nong」を過去形として使おうとする動きがある。また「oa」「e」も若干の使用率の上昇を見せている。しかし、10 世代付近で規則変化させる遺伝子に逆転される。以降は「e」の遺伝子が少し使用率をあげようとする動きがあるが、規則変化の進行に抗うまでには至っていない。逆に図 21 では、規則変化の遺伝子は 10 世代目以降ほとんどが淘汰され、不規則変化を保っている。これら 2 つの違いは発話する動詞のランダム性にあると考えられる。「ning」「ploamph」の頻度は同じであるが、全く同じ回数発話されるわけではない。本モデルでは設定上、不規則変化で意味が通じなければ必ず規則変化の遺伝子に点数がつくようになっている。そのためルーレット上に大きな差がない初期の数世代で非常に多く発話されれば、不規則変化では意味が通じないことが多い。よってペナルティの大きさも相まって、規則化は急激に進む。

図 22 では「e」と規則変化の間でせめぎあいが起こっている。「e」「o」を表す遺伝子が最初の数世代で使用率を増加させているが規則変化が進み始めると、少い割合の遺伝子は意味が伝わる確率も下がっていくので徐々に数を減らしていく。ある使用率で停滞してしまう原因としては以下が考えられる。図 22 の状況

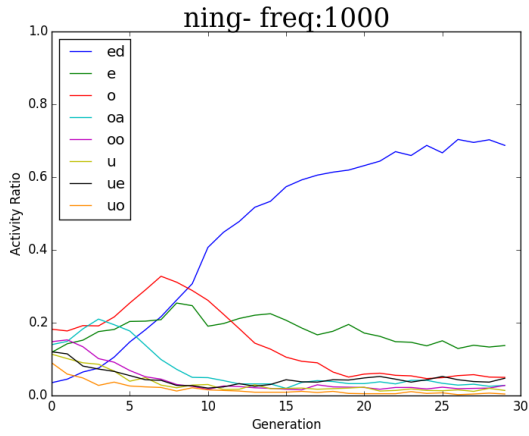


図 20 規則化が起こった例  
出現頻度:1000 高頻度ペナルティ 4.0

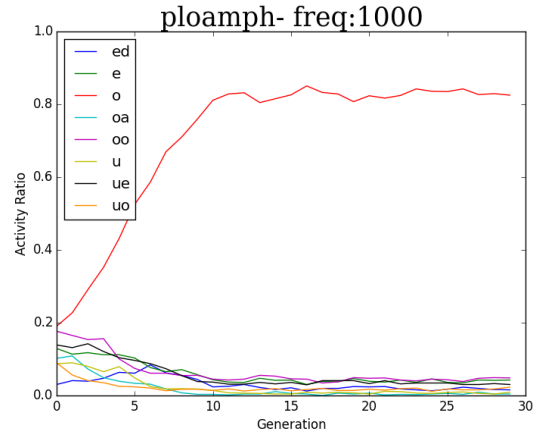


図 21 不規則変化を保つ例  
出現頻度:1000 高頻度ペナルティ 4.0

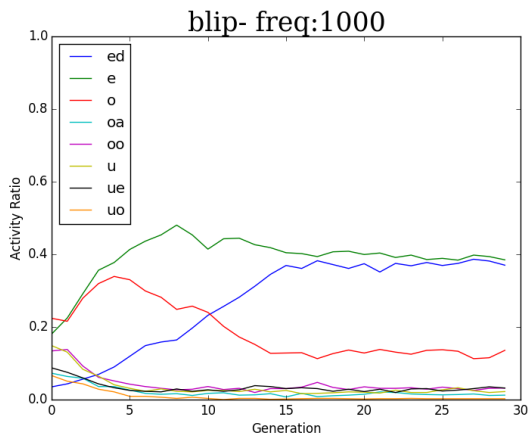


図 22 規則化と不規則変化が競合する場合  
出現頻度:1000 高頻度ペナルティ 4.0

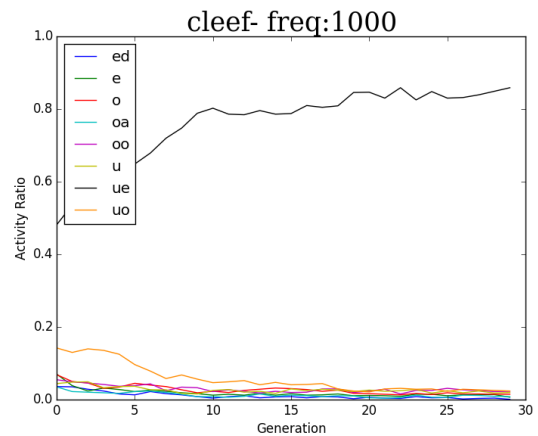


図 23 類似ペナルティを受ける動詞  
出現頻度:1000 高頻度ペナルティ 4.0

では不規則変化の遺伝子「e」「o」を使ってもある程度意味が通じる。よって規則変化させる遺伝子の適応度を高くしない要因が高頻度ペナルティ以外にも生まれる。つまり、それぞれの遺伝子の適応度に明確な差がでなくなり淘汰圧がかからなくなってしまうのである。複数の母音を表す遺伝子が同時に増加していく場合にも同じ現象が見られる。ただし、これは必ずしも悪い結果ではない。人間社会の中で特定の過去形を決定できないこと（規則、不規則の間で揺れ動いていること）を表しているのである。これを解消する方法（GAの中で適応度の変換を行う方法）も考えられるが本研究では行っていない。図 23 は類似ペナルティが科される動詞であるが、ルーレットの初期状態の割合に非常に大きな差があるため、ほとんど効果がみられなかった。

全体では 12 語中 3 語が規則変化の遺伝子が最も多い割合を占めるようになり、それ以外はなんらかの不規則変化の遺伝子の割合が最も多いという結果になった。

- 高頻度ペナルティ 5.5 の結果

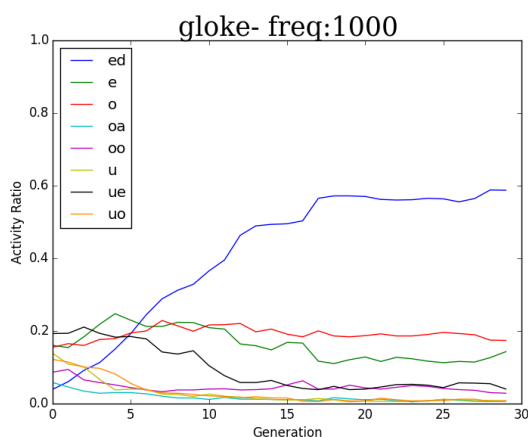


図 24 規則化が起こった例  
出現頻度:1000 高頻度ペナルティ 5.5

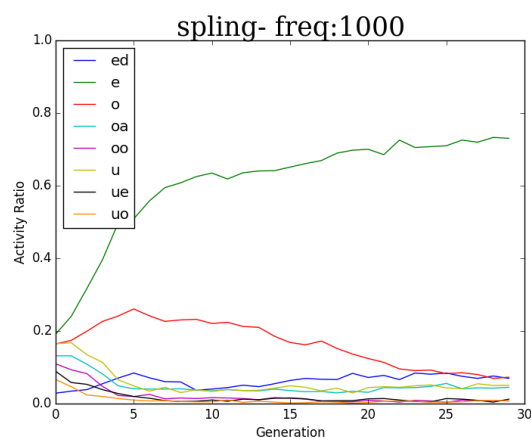


図 25 不規則変化を保つ例  
出現頻度:1000 高頻度ペナルティ 5.5

高頻度ペナルティ 5.5 の場合でも規則変化の遺伝子が多数を占めるような現象が起こった（図 24）。理由は高頻度ペナルティ 4.0 で述べたものと同様であると考えられる。また不規則変化を保つ例（図 25）では「e」の遺伝子の増加とともに「o」の遺伝子が減少していく様子が示されている。図 26 「blafe」は類似ペナルティの効果が見られた例である。6 世代付近まで規則変化の遺伝子が増加していくが、その後減少していく。その代わりに「o」の遺伝子の割合が上昇している。ただし、ここでも複数の母音のせめぎあいが見られる。また、図 27 のようにルーレットの初期状態の割合に非常に大きな差がある場合は、大きな割合を持つ母音に収束する。

全体では 12 語中 1 語（図 24）が規則変化の傾向を示した。

- 高頻度ペナルティ 6.0 の結果

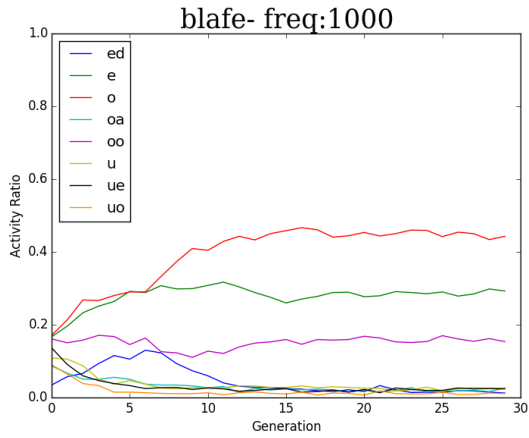


図 26 類似ペナルティを受ける動詞  
出現頻度:1000 高頻度ペナルティ5.5

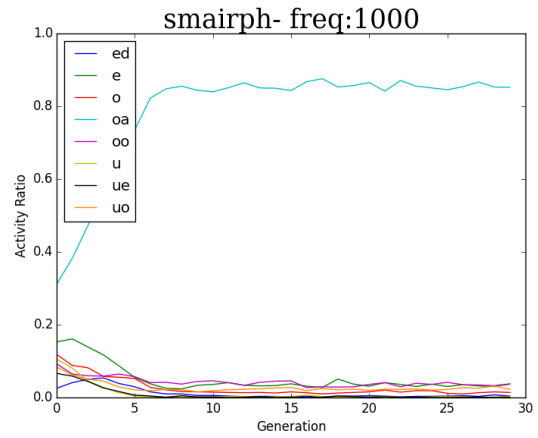


図 27 初期ルーレットの割合に大きな差がある例  
出現頻度:1000 高頻度ペナルティ5.5

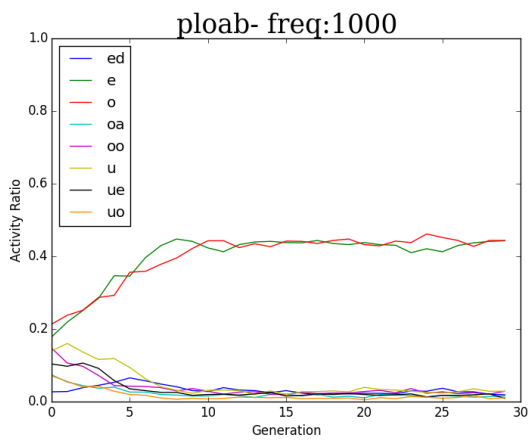


図 28 類似ペナルティを受ける動詞  
出現頻度:1000 高頻度ペナルティ6.0

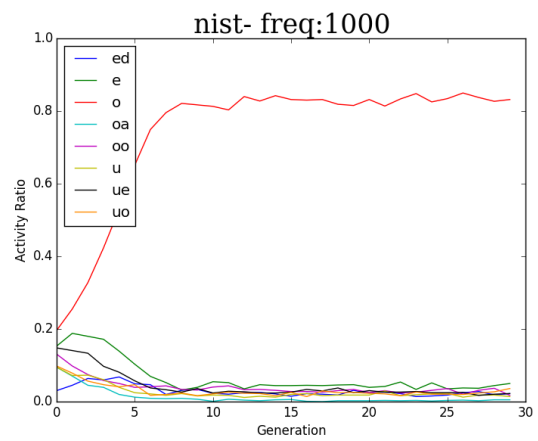


図 29 音節ペナルティを受ける例  
出現頻度:1000 高頻度ペナルティ6.0

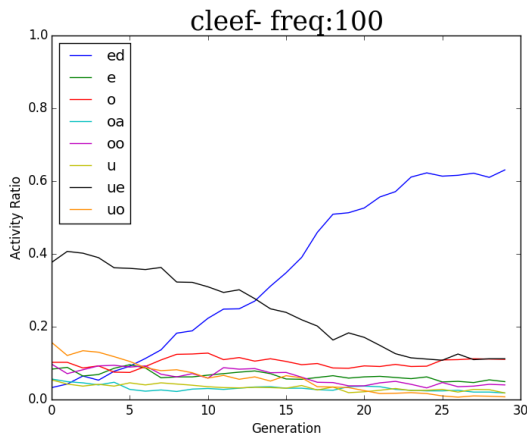


図 30 不規則変化と規則変化の逆転  
出現頻度:100

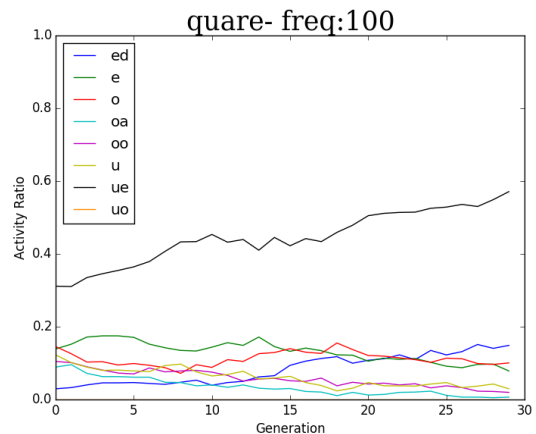


図 31 初期ルーレットの割合に大きな差がある例  
出現頻度:100

高頻度ペナルティが 6.0 の場合でも 4.0、5.5 の場合と同様の傾向見受けられた。ここでは類似ペナルティと音節ペナルティに関する動詞のみ紹介する。

図 28 は類似ペナルティを受ける動詞である。そのため、5 世代目付近で少しだけ規則化傾向を見せるものの「e」「o」がせめぎ合う形に落ち着いている。また「nist」は語尾が「t」であるので規則化した時に音節が増える動詞である。よって音節ペナルティを受ける可能性がある。つまり、高頻度ペナルティと音節ペナルティを同時に受けることもある。つまり、規則化が進む可能性はほとんどないと言える。

全体では 12 語中 2 語が規則変化の傾向を示した。

### 7.1.2 出現頻度 100 の場合

次に出現頻度が 100 の場合について述べる。この出現頻度を持つ動詞は 22 語ある。出現頻度が 100 という事は単純に出現頻度 1000 の動詞の 1/10 の確率で発話されるという事である。この出現頻度の動詞に関しては、7.1.1 節のような高頻度ペナルティは科されない。ただし、類似ペナルティ、音節ペナルティは同様に科す。結果は高頻度ペナルティによらず同じような傾向を示したため、代表して高頻度ペナルティ 5.5 の場合の結果を使用する。以下に一部抜粋した結果を示す。

図 32 は単純に規則変化が起こったことを示す。出現頻度 1000 と 100 の大きな違いは図 30 のような現象が起きることである。初期のルーレット上の遺伝子の割合に大きな差が存在するが、時間が経てば規則変化に淘汰される。これは高頻度ペナルティが科せられないために起きるものである。図 30 の「cleef」は類似ペナルティ対象であるが、確率的に科せられるため規則変化の勢いに抗うことができなかったのである。ここには示していないが「goav」も初期ルーレットがかなり偏っている動詞であり、かつ類似ペナルティ対象であるが規則変化に逆転されている。これは人間で考えるならば、「規則変化させれば不自然に感じる時があるが、使用頻度もそれほど高くなくまた規則変化のほうが意味が通じるので気にせず使う」という方向に変化していくことである。本モデルではこの変化が徐々にコミュニティに浸透していく状況が表現できて

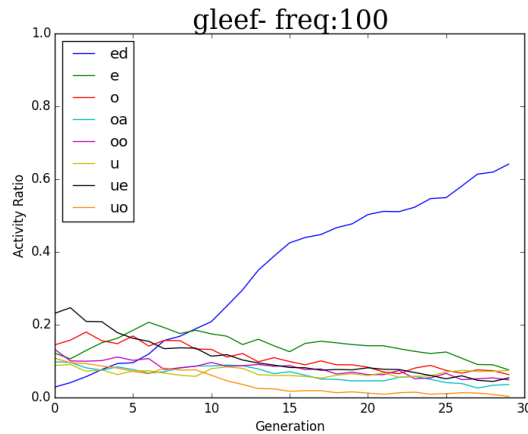


図 32 規則変化が起こった例出現頻度:100

いると言える。

ただし、図 31 のように規則変化に淘汰されずに、コミュニティの中で不規則変化のまま使用され続ける動詞も存在する。この場合は、不規則変化の遺伝子が高い適応度になることは初期のルーレットの偏りに依存しているということも伺える。

最後に全体の傾向を述べる。高頻度ペナルティに関わらず上記の 3 つの図に準ずる変化が起こった。22 語中 16 から 17 語が規則化される結果となった。

### 7.1.3 出現頻度 10, 1 の場合

最後に非常に低頻度である動詞について述べる。これらも出現頻度 100 の動詞と同様に類似ペナルティと音節ペナルティのみが課せられる。しかし、これらの動詞は低頻度ゆえにほとんど発話されない。高頻度ペナルティ 5.5 の場合の結果を図 33、34 に示す。

図 33、34 では微妙な変化が見られるもののほとんど使用率の変化は見られない。発話されないという事は知識の更新ができない。現実世界では忘れ去られていると考えればこれは自然である。

### 7.1.4 実験 1 のまとめ

本節では高頻度ペナルティについて実験結果と考察を述べた。出現頻度 1000 の動詞に関しては高頻度ペナルティが 4 から 6 の間であれば、10%から 25%程度の規則化を起こすことができる。先行研究(図 4 上の  $10^{-3} - 10^{-2}$  の動詞)では 10%である。本研究では動詞の出現頻度を恣意的に固定することはしなかった。確率的に振る舞う言語変化の状況を観察したかったためである。そのため、先行研究と全く同じ結果にすることはできないが、それに近い状況を再現するための高頻度ペナルティの範囲を定めた。その範囲の中では初期が確率的に行われても同じような規則化の割合を再現することができた。その範囲外では、4 以下では規則化の割合が多くなってしまふことが多く、逆に 6 以上にすれば規則化がほとんど起きないような状況であった。

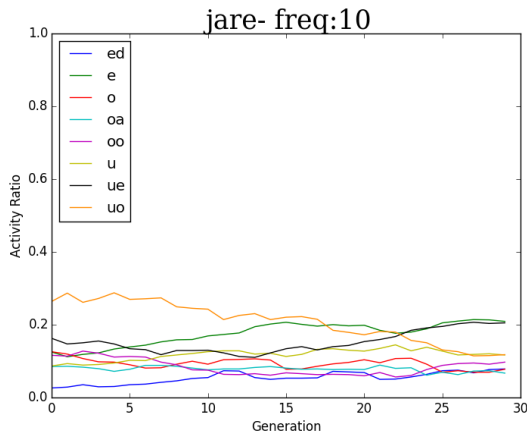


図 33 低頻度の場合 出現頻度:10

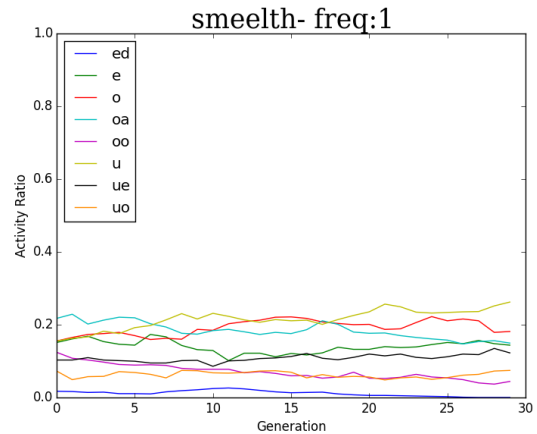


図 34 超低頻度の場合 出現頻度:1

出現頻度 100 の動詞に関しては、本モデルでは 70%程度の規則化が起きるが、先行研究では 43%である。この差は大きいですが、中頻度の不規則動詞が最も規則化されるという状況は本モデルでも再現できている。これ以上の調整は本モデルでは不可能である。規則化を調整するには類似ペナルティと音節ペナルティを用いるしかないが、これらのペナルティは出現頻度に対してではなく個別の動詞に科せられる。ゆえに出現頻度 100 の動詞のみに何かしらの調整をすることはできないのである。

最後に低頻度の動詞についてであるが、先行研究上でほとんどが規則化されていると述べている。しかし、本モデル上では低頻度の動詞は評価できないために規則化を起こすことはできない。恣意的に規則化を起こすような仕組みを入れることも考えられるが、それは不自然である。現実世界で人間が使っていない動詞の知識を更新していくとは考えにくい。よって使われないという事は、忘れられていくと考えたほうが自然であるといえる。先行研究においても低頻度の動詞は不規則動詞であるという認識自体失われており、不規則変化という選択肢が消えた結果であると考えられる。このような理由から低頻度の動詞については先行研究の再現と言うよりも、モデルの自然さを優先した。

ただし、全頻度においてルーレットの初期状態に大きく依存することが明らかになった。

## 7.2 実験 2

実験 1 と同様に類似ペナルティの対象となる動詞を表 11 に示す。

表 11 類似ペナルティ対象の動詞

超高頻度の動詞	類似ペナルティ対象の動詞
jare, gleef, smaib, fring, preed	glip, gloke, goav, blafe, cleef joam, quare, preek, smeej, smeeg, smeeb, meep

## 7.2.1 出現頻度 1000 の場合

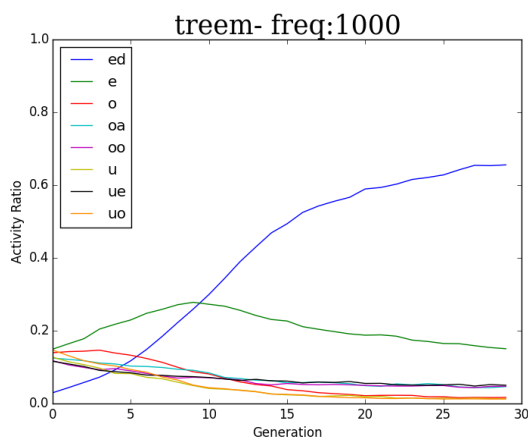


図 35 規則化が起こった例出現頻度:1000

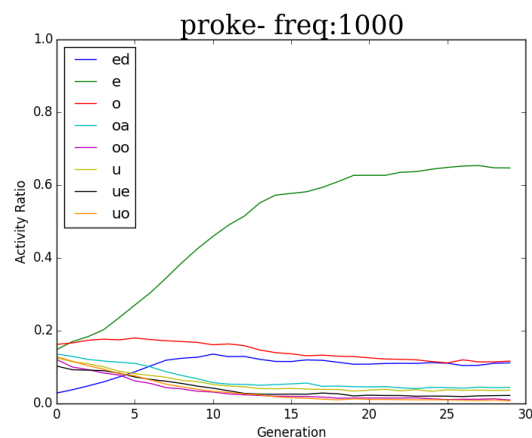


図 36 不規則変化を保つ例出現頻度:1000

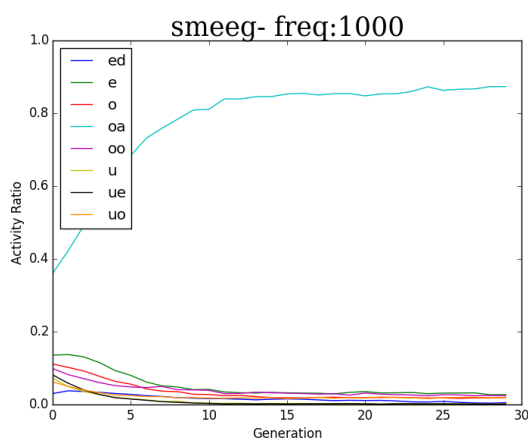


図 37 初期ルーレットの割合に大きな差がある例  
出現頻度:1000

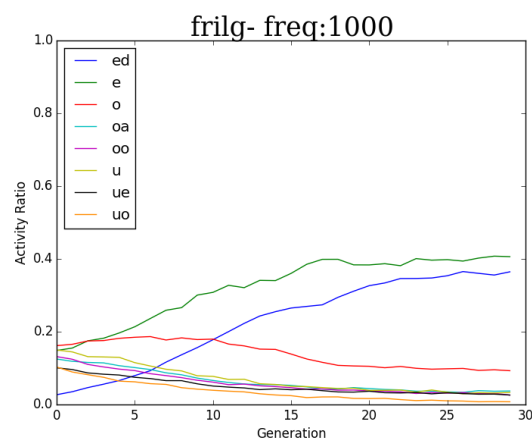


図 38 規則化と不規則変化が競合する場合  
出現頻度:1000

図 35 は規則化が起こった例である。エージェント数が増えたことにもよるが、非常に綺麗な S-curve を描いていると言える。図 36 でも「e」に落ち着いていく様子がわかる。図 37 は初期のルーレットに偏りがあるものであり、ここでも大きな割合を占めるものに収束する。図 38 では「e」の遺伝子と規則変化の遺伝子でせめぎあいが起きている。このように出現頻度 1000 の動詞では実験 1 と同様の傾向が見られた。また実験 1 に比べ、エージェント数が 10 倍ほどになっているが 30 世代でも変化の収束が見られる。これは使用したスケールフリーネットワークの特性によるものである。4.2.1 で述べたように、スケールフリーネットワークではノード (エージェント) の数が増えても平均距離はほとんど変化しない。エージェントが 12 人の



場合の平均距離は式 (10) より約 2.72 であり、100 人の場合は約 3.02 である。平均距離が小さいという事はそれだけ早く変化がエージェント間に広がることを示す。現実の人間関係でも平均距離が大きくなることは非常に少ない [26]。また、Ke [25] においてもスケールフリーネットワークでは、ノードが増えても変化の伝搬スピードに影響を与えないことが示されている。本研究でも出現頻度 1000 の動詞においては、ネットワークの特徴が生かされた実験ができていると言える。

### 7.2.2 出現頻度 100 の場合

出現頻度が 100 の動詞では、図 39,40 のようにほぼ一定の速度で規則化が行われるという現象が起こった。また、図には示していないが初期のルーレットに偏りがある場合はこれまでと同様の挙動であった。7.2.1 で示したように、ネットワークは変化の速度に影響を及ぼさない。しかし、エージェント数が 12 人の場合に比べ、30 世代目で特定の遺伝子の割合が大きな割合を占めるまでに変化は進まなかった。これは発話数を 100 から 30 に大幅に減らしたため、そもそも発話されず十分に学習が行われなかった結果であると考えられる。

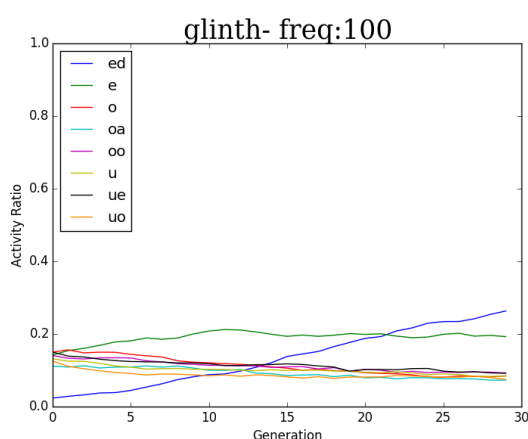


図 39 規則化が起こった例 出現頻度:100

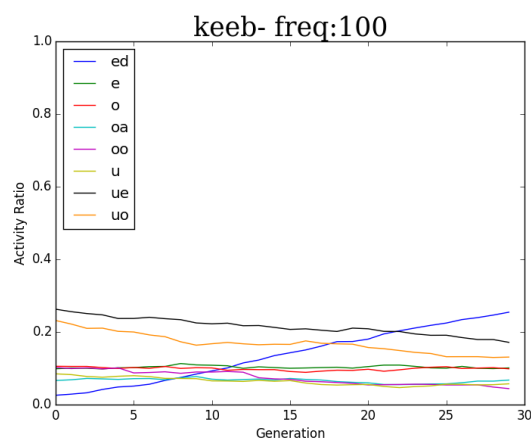


図 40 不規則変化と規則変化の逆転 出現頻度:100

また、出現頻度頻度が 10、1 の動詞は実験 1 と同様に全く変化が起きなかったため割愛する。

### 7.2.3 実験 2 のまとめ

実験 2 ではネットワークサイズと、変化の速度について考察を行った。エージェントサイズを 12 人から 100 人に増やして実験を行っても、出現頻度 1000 の動詞では変化の収束に影響を与えなかった。これは本モデルでも使用したスケールフリーネットワークの特性を十分反映できていることを示す。また、出現頻度 100 の動詞では本モデルでは十分な発話数がエージェントの学習に必要な不可欠であることが明らかになった。

## 8 結論

### 8.1 おわりに

最後に本研究の全体のまとめを行う。本研究では先行研究 Lieberman [16] を追試するためにマルチエージェント環境を用いた不規則動詞の過去形獲得モデルの提案を行った。そのため、不規則動詞に関する言語学、認知科学的な調査を行った。さらに、不規則動詞が規則化される原因についても歴史的背景を調査した。マルチエージェント環境は複雑ネットワークを用いた言語伝搬の研究を参考に構築した。

実験には Pinker [14] で使用されている人工動詞 60 語を用いた。エージェントの知識の初期化は人工動詞の擬似的な音韻的距離を求めることで行った。エージェントはその知識を GA を用いることで学習、更新していく。エージェントの学習に用いた GA では以下の条件のもと適応度計算が行われる。

1. 高頻度の動詞は規則化したくないためペナルティ  $n$  が科せられる
2. 規則化すると音節が増える、また超高頻度の動詞と音韻的距離が近い場合ペナルティが科せられる

実験を繰り返した結果、1 のペナルティは  $4 \leq n \leq 6$ 、2 のペナルティは 5 とした。次に、提案モデルを計算機上に実装し実験を行った。

- 発話されない動詞は学習されないという自然な状況を優先した。そのため、先行研究 [16] の規則化の割合を完全に再現することはできなかった。しかし、高頻度の動詞は不規則変化を保ち、頻度が低くなれば規則化が起こるという状況を再現することができた。
- マルチエージェント環境 (ネットワークサイズ) を大きくしても、言語変化の速度は影響されないというネットワーク特有の特性も十分に反映されていることを確認できた。

本研究において提案したモデルは、言語の持つ特徴と人間のコミュニケーションの目標が組み込まれたモデルである。また Ke [25] のように言語変化を古いか、新しいかの 2 値問題に置き換えるものではない。よって提案モデルはより現実に近い言語変化をシミュレーションできるモデルであると言える。また、同時に複数の動詞を扱える点も大きな貢献がある。エージェント知識の中でそれら動詞の知識がどのように相互作用するのか、を考慮することでエージェントをより人間らしく改良することもできる。

先行研究の完全な追試を行うことはできなかったがこれらは言語変化シミュレーションのモデルに対する大きな貢献であると考えられる。

### 8.2 今後の課題

#### 8.2.1 人口流入の影響

本研究では実際に人口流入に関する実験は行っていないが Fagyal [19], Ke [25] に述べられていることから考察を行い、本モデルでその仕組みが実現可能であることを述べる。どちらも人口流入について明確な議論を行っていないわけではないが、Ke [25] では移民を複雑ネットワークで表現する方法について述べてい

る。言語変化が起こるには、古い言語にとって変わる新しい言語が発生することが必要である。そのような状況を再現するには実験のスタート時に新しい言語を持ったエージェントを多く配置する必要がある。古い言語を持つ人間社会に新しい言語を持つ人間が多く存在するので、大量の移民があったと考えることができるという主張である。しかし、移民は特定の時期に一回だけ起こるのではなく、時代の流れの中で複数回起こっていると考える方が自然である。提案モデルにおいては世代交代が行われるたびに上記の移民が存在するという状況を作成すれば良い。さらにそれを実現する際に、これまで使われてきた実績のある BA モデルのアルゴリズムを用いて人間社会を模倣した人口流入 (移民エージェントの設置) が可能であると考えられる。

また、Fagyal [19] において言語変化を推進するような要因と妨害するような要因が述べられている。例えば以下のようなものである。

- 推進

言語が類似していて、その言語を使う国の文化が発展している場合は語彙の借用が起こる。実際に第 2.2 節で述べたようにフランス語の語彙の流入が起こっている事実もある。

- 妨害

言語ナショナリズムなど自分が使っている言語に固着するような思想傾向

妨害に関しては移民エージェントで実現可能であると考えられる。つまり、移民エージェントはすでにルーレットがほぼ確定したような状態でネットワークに配置され自分の知識はほとんど更新しない (学習しない) ように設定する。そして、この学習の有無がある状態で言語変化を比較すれば、実際に妨害が起こるかどうかが人口流入の影響として観測することができる。

### 8.2.2 初期化依存の問題

実験 1,2 の両方で不規則変化を保つ動詞は初期状態のルーレットの偏りに大きく依存することがわかった。これは本研究で用いた MetaPhone アルゴリズムの特性が関係していると考えられる。MetaPhone アルゴリズムではスペリングが擬似的な発音文字列に変換されるものであり完全な発音ではない。また、MetaPhone に変換後の距離計算には編集距離を用いるため、発音と言うよりもやはりスペリングの距離に近いものとなっている。解消するためにはより発音に近い形にスペリングを変換する必要がある。既存の語であれば

<http://www.speech.cs.cmu.edu/cgi-bin/cmudict>

のようなプロジェクトを利用すればより発音に近い文字列に変換できる。今回のように人工動詞を用いる際は自動で変換はできないので手動で変換するしかない。

### 8.2.3 年齢設定

Ke [25] では各エージェントに年齢が設定されていたが本研究ではモデルの単純化のために設定しなかった。ただし、より自然な (人間的なエージェントという意味で) モデルの構築のためには必要であると考えられる。

#### 8.2.4 課題のまとめ

本研究では人口流入の影響を調査することが大きな目的だったが、提案モデルにも初期依存や年齢の課題があったため考察を行うにとどめた。今後は初期化の調整や、エージェントの年齢などの概念を加えること、また先行研究の追試を行えるように根拠のあるパラメータ調整が必要である。

## 謝辞

研究を進めるにあたり、適切なお助言を頂きました東条 敏教授に深く感謝いたします。日頃から声をかけて下さった佐野 勝彦助教にも感謝申し上げます。また、研究会にてお助言を頂きました富山高等専門学校の的場 隆一准教授、名古屋大学外山研究室の中村 誠特任助教にも深く感謝いたします。

## 参考文献

- [1] Steven Pinker. *The irregular verbs. Landfall* pp83—85, (Autumn issue 2000)
- [2] Steven Pinker. *On language and connectionism: analysis of a parallel distributed processing model of language acquisition. Cognition* 28, pp73-193 (1988)
- [3] Huang, Yi Ting. Steven Pinker. *Semantic, Phonological, and Lexical Influences on Regular and Irregular Inflection* Annual meeting of the Cognitive Science Society. (2005)
- [4] 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)
- [5] 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)
- [6] スティーヴン・ロジャー フィッシャー, *A History of Language*, (2001), 鈴木晶 訳, 『ことばの歴史—アリのことばからインターネットのことばまで』, 研究社, (2001)
- [7] 中島 平三, 『言語の事典』 朝倉書店, (2005)
- [8] Philip Gooden, *THE STORY OF ENGLISH* (2009), 田口孝夫 監修 『物語 英語の歴史』, (2012)
- [9] 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』, 京都産業大学論集, (2006)
- [10] 亀井孝, 河野六郎, 千野栄一, 『言語学大辞典. Vol. 1』 三省堂, (1996)
- [11] ロバート・マクラム, ウィリアム・クラン, ロバート・マクニール, *THE STORY OF ENGLISH*, 岩崎春男ほか 訳, 『英語物語』, 株式会社文藝春秋 (1989)
- [12] Jackendoff Ray. *Foundations of language: Brain, meaning, grammar, evolution. Oxford University Press*, (2002)
- [13] Steven Pinker, Michael T.Ullman. *The past-tense debate. cognitive processing* 5 (2002)
- [14] Prasada, Sandeep, Steven Pinker. *Generalisation of regular and irregular morphological patterns. Language and cognitive processes* 8.1, pp1-56, (1993)
- [15] 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)
- [16] Erez Lieberman, Jean-Baptiste Michel, Joe Jackson, Tina Tang , Martin A. Nowak, *Quantifying the evolutionary dynamics of language. Nature* Vol449 (11 October 2007)
- [17] *Supplementary information*,  
<http://www.nature.com/nature/journal/v449/n7163/supinfo/nature06137.html>
- [18] *CELEX* <http://wwwlands2.let.kun.nl/members/software/celex.html>

- [19] Zsuzsanna Fagyala, Samarth Swarupb, Anna Mara Escobara, Les Gasserc, Kiran Lakkara-jud. *Centers and peripheries: Network roles in language change*. *Lingua*, 120(8), pp2061-2079 (2010)
- [20] Fakhteh Ghanbarnejad, Martin Gerlach, Jose M. Miotto, Eduardo G. Altmann. *Extracting information from S-curves of language change*. *Cornell University Library arXiv preprint arXiv:1406.4498* (2014)
- [21] Chris Upward, *Spelling Reform in German*. *Journal of the Simplified Spelling Society*, J21, 1997-1 , pp22-24
- [22] *Google Ngram Viewer* <https://books.google.com/ngrams>
- [23] 齊藤 渉, 『フンボルトの言語研究:有機体としての言語』, 京都大学学術出版会, (2001)
- [24] 大内 東, 山本 雅人, 河村 秀憲, 『マルチエージェントシステムの基礎と応用:複雑系工学の計算パラダイム』, コロナ社, (2002)
- [25] Ke, Jinyun. *Complex networks and human language* arXiv preprint cs/0701135 (2007)
- [26] 増田直紀, 今野紀雄. 『複雑ネットワーク. 近代科学社』 (2010)
- [27] 増田直紀, 今野紀雄. 『「複雑ネットワーク」 とは何か 複雑な関係を読み解く新しいアプローチ』, 講談社, (2006)
- [28] Barabasi, Albert-Lszl, Rka Albert. *Emergence of scaling in random networks*. *science* 286.5439, pp509-512, (1999)
- [29] Lawrence Philips, *The double metaphone search algorithm*. *C/C++ Users Journal Volume 18 Issue 6*, (June 2000)
- [30] Morfoedro, Approximate string-matching algorithms,  
<http://www.morfoedro.it/doc.php?n=222&lang=en>
- [31] *Fast Python phonetic algorithms* <https://pypi.python.org/pypi/Fuzzy>
- [32] マックレオドプランケット, ロールズ E T, 『認知過程のコネクショニスト・モデル』, 深谷澄男, 喜田安哲, 伊藤尚枝, 向井敦子, 齊藤謁 訳, 北樹出版, (2005)
- [33] Anaconda, CONTINUM ANALYTICS, <https://store.continuum.io/cshop/anaconda/>

## 付録 A

### 人工動詞のパーズに用いた字素

- 始端子音クラスタ

y , s , p , t , k , q , c , b , d , g , f , v , j , z , l , m , n , r , w , h , ch , gh , gn , ph , ps , rh , sh , th , ts , wh

- 母音クラスタ

e , i , o , u , a , y , ai , au , aw , ay , ea , ee , ei , eu , ew , ey , ie , oa , oe , oi , oo , ou , ow , oy , ue , ui , uy

- 終端子音クラスタ

h , r , l , m , n , b , d , g , c , x , f , v , i , s , z , p , t , k , q , bb , ch , ck , dd , dg , ff , gg , gh , gn , ks , ll , ng , nn , ph , pp , ps , rr , sh , sl , ss , tch , th , ts , tt , zz , u , e , es , ed , j



## 付録 B

本文中に示していない実験結果を以下に示す。結果は変化の傾向ごとに全結果ではなく代表例を掲載する。

### 実験 1

#### 出現頻度 1000

- 規則化される例

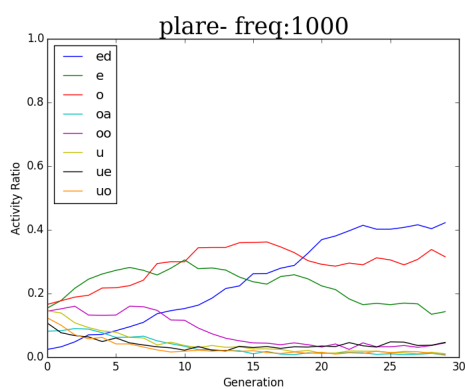


図 B 1 高頻度ペナルティ 4.0 plare

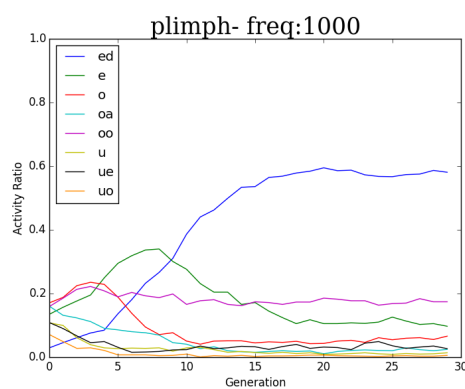


図 B 2 高頻度ペナルティ 4.0 plimph

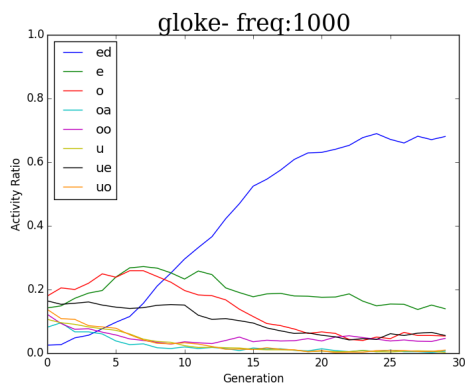


図 B 3 高頻度ペナルティ 6.0 gloke

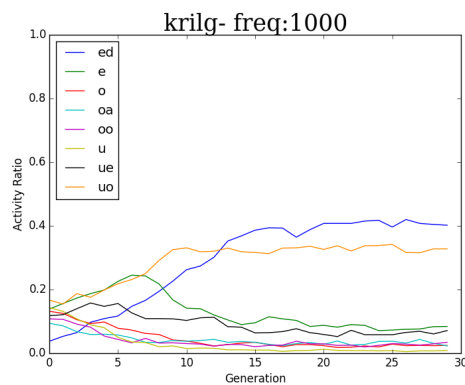


図 B 4 高頻度ペナルティ 6.0 krilg

- 不規則変化を保つ例 (初期状態に偏りがある動詞も含む)

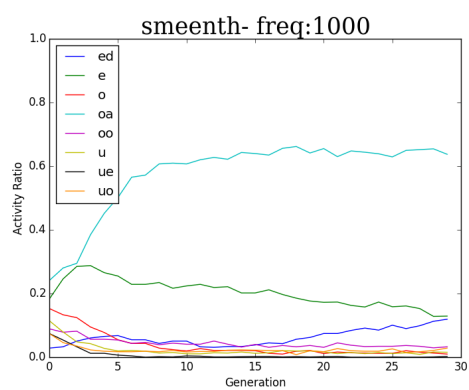


図 B 5 高頻度ペナルティ 4.0 smeenth

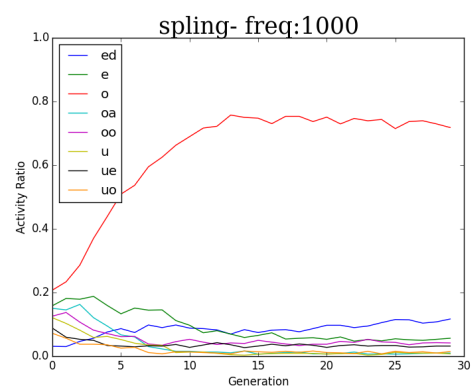


図 B 6 高頻度ペナルティ 4.0 spling

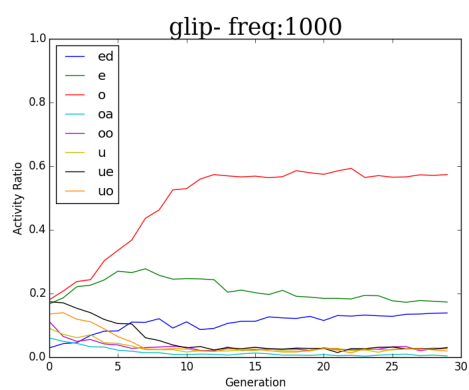


図 B 7 高頻度ペナルティ 5.5 glip

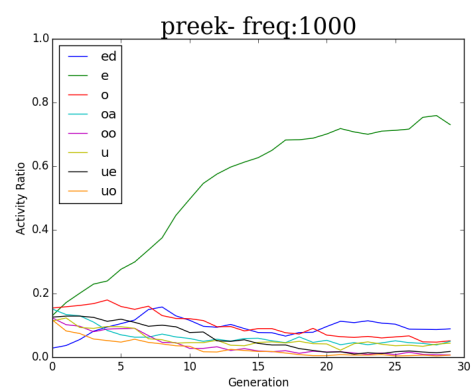


図 B 8 高頻度ペナルティ 5.5 preek

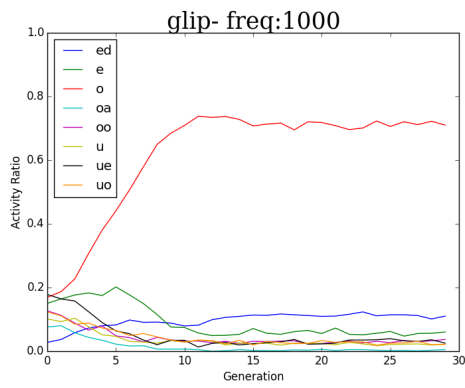


図 B 9 高頻度ペナルティ 6.0 glip

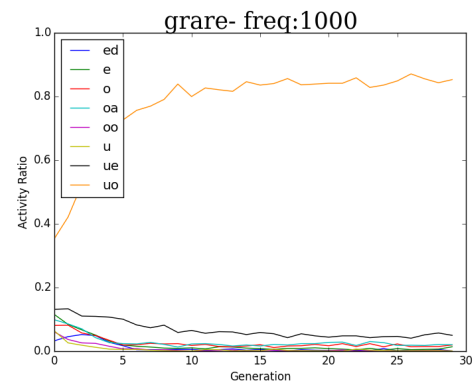


図 B 10 高頻度ペナルティ 6.0 grare

- 競合が起る例

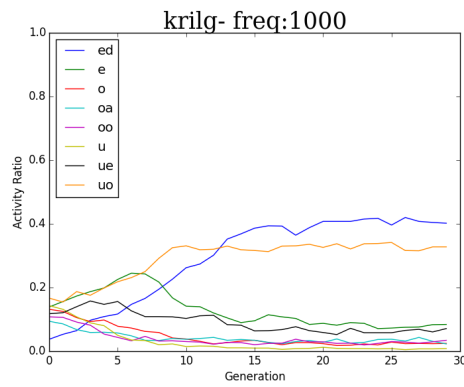


図 B 11 高頻度ペナルティ 6.0 krilg

## 出現頻度 100

- 規則化される例

出現頻度 100 の動詞において特徴的な現象である不規則変化と規則変化の逆転の例も示す。

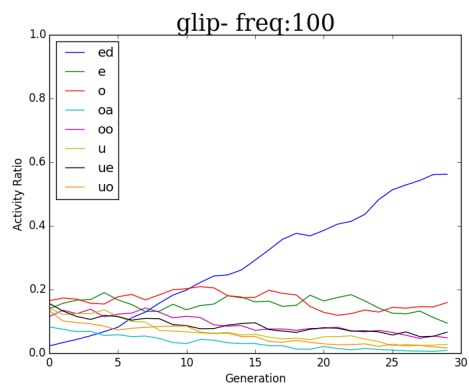


図 B 12 高頻度ペナルティ 4.0 glip

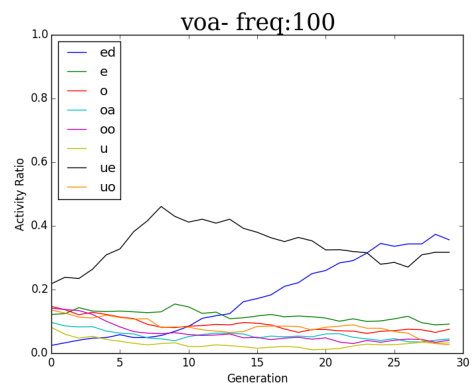


図 B 13 高頻度ペナルティ 4.0 voa 逆転

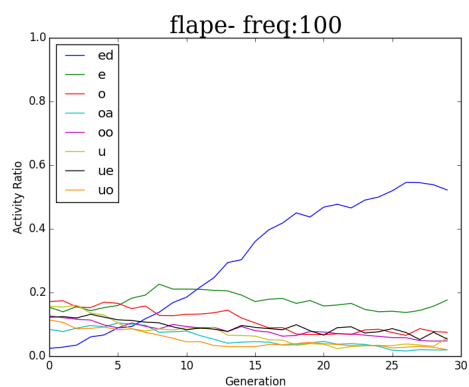


図 B 14 高頻度ペナルティ 5.5 flape

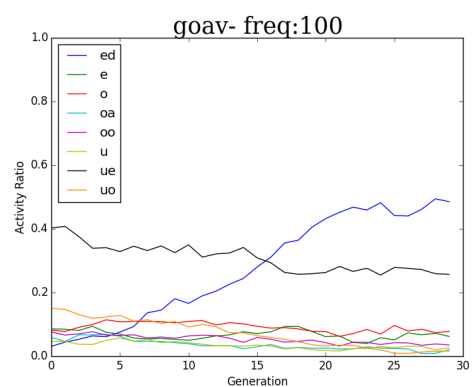


図 B 15 高頻度ペナルティ 5.5 goav 逆転

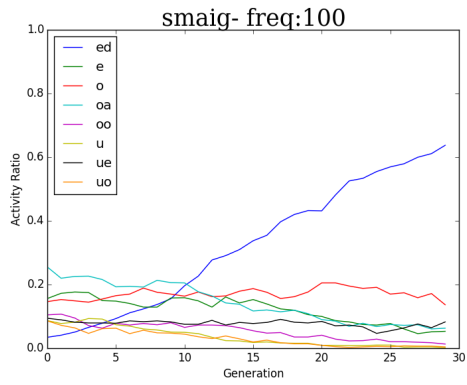


図 B 16 高頻度ペナルティ 6.0 smaig

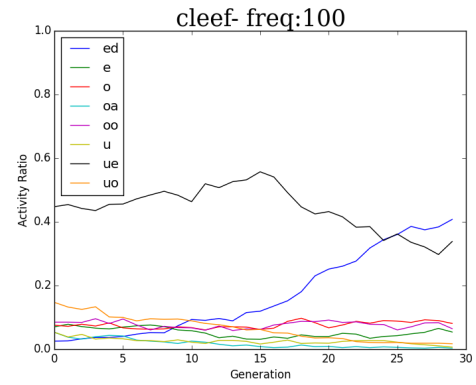


図 B 17 高頻度ペナルティ 6.0 cleef 逆転

- 不規則変化を保つ例 (初期状態に偏りがある動詞も含む)

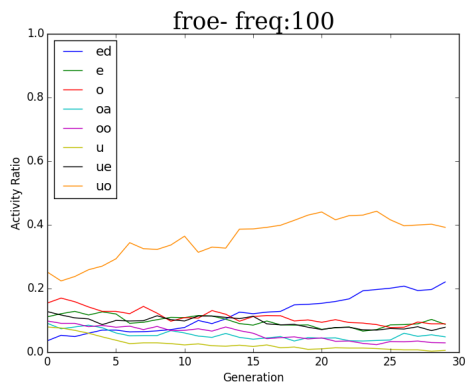


図 B 18 高頻度ペナルティ 4.0 froe

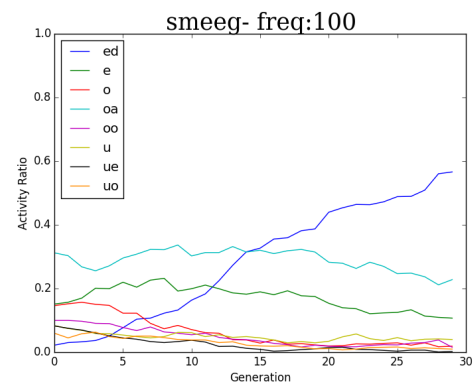


図 B 19 高頻度ペナルティ 5.5 smeeg

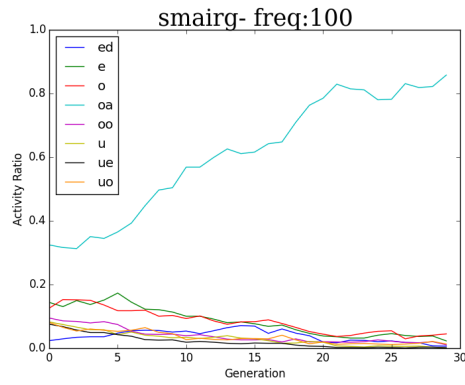


図 B 20 高頻度ペナルティ 6.0 smaigr

## 出現頻度 10 および 1

出現頻度が 10 および 1 の動詞は、ほとんど変化が見られなかった。

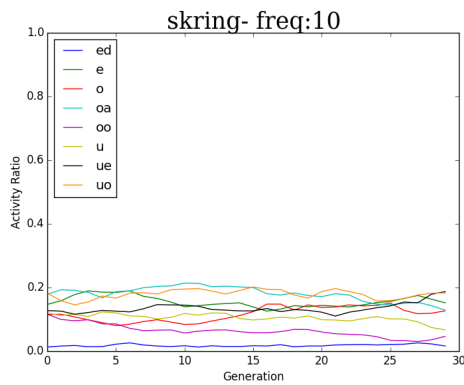


図 B 21 高頻度ペナルティ 4.0 頻度 10 skring

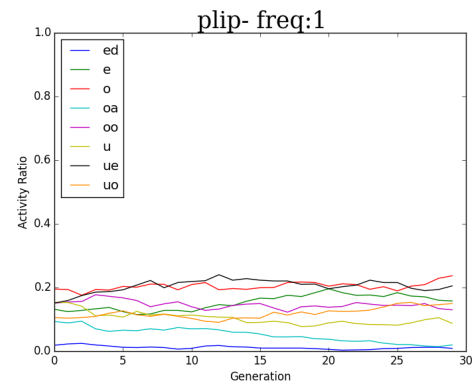


図 B 22 高頻度ペナルティ 4.0 頻度 1 plip

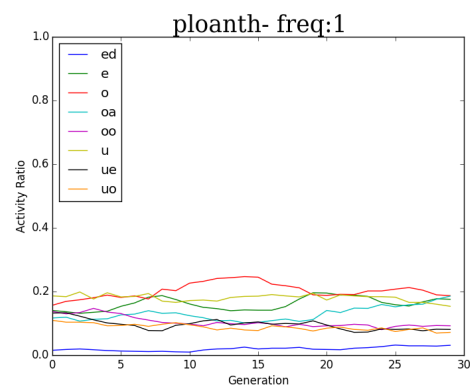
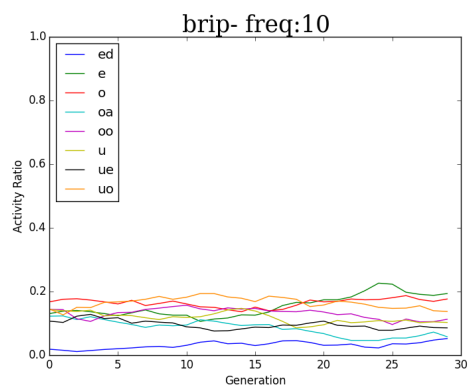


図 B 23 高頻度ペナルティ5.5 頻度 10 brip

図 B 24 高頻度ペナルティ5.5 頻度 1 ploanth

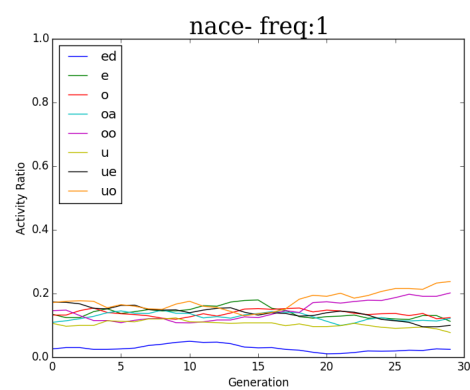
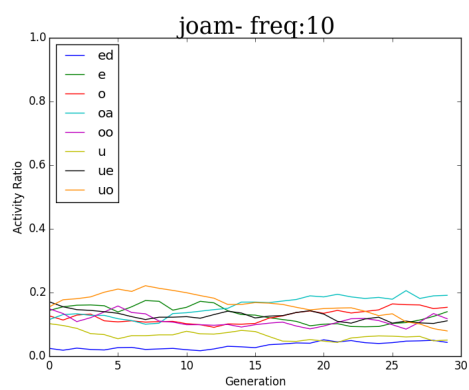


図 B 25 高頻度ペナルティ6.0 頻度 10 joam

図 B 26 高頻度ペナルティ6.0 頻度 1 nace

## 実験 2

出現頻度 1000

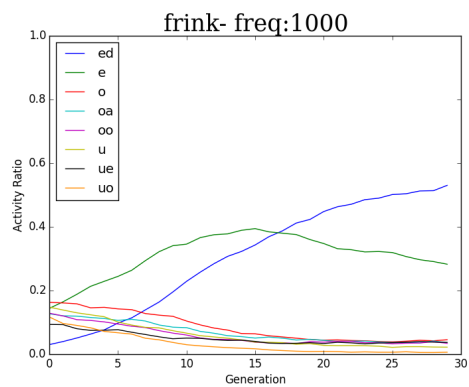


図 B 27 規則化される例 frink

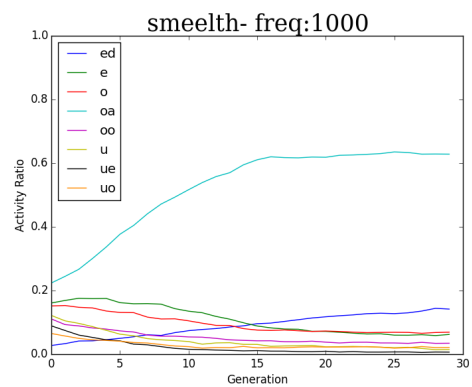


図 B 28 不規則を保つ例 smeelth

出現頻度 100

発話数を大幅に減らしたため、変化のスピードは遅くなっている。どちらも規則化傾向にあるが、規則変化の使用率は低いままである。

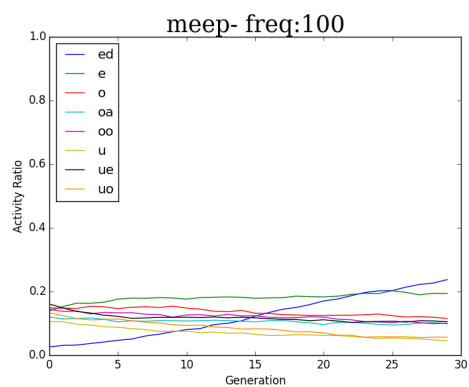


図 B 29 規則化傾向にある例 meep

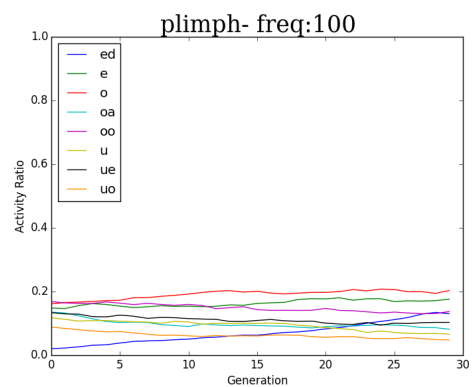


図 B 30 規則化傾向にある例 plimph



## 出現頻度 10 および 1

やはり、これらの出現頻度の動詞も実験 1 と同様にほとんど変化が見られなかった。

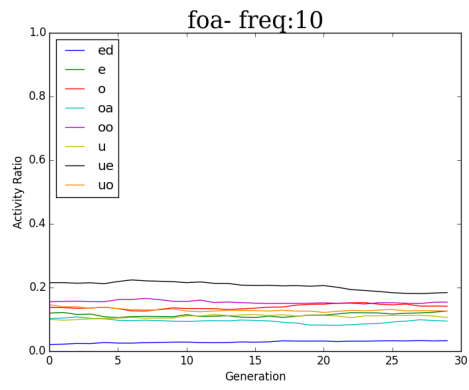


図 B 31 頻度 10 foa

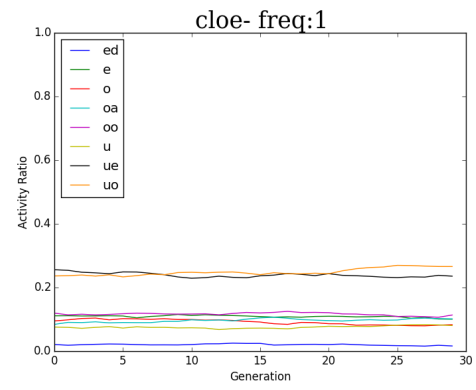


図 B 32 頻度 1 cloe