

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

東条研究室

鈴木啓章

2014年9月1日

発表の流れ

- ・ 背景
- ・ 目的
- ・ 先行研究
 - ・ 不規則動詞の出現頻度と規則化速度
 - ・ 語形変化と外圧の影響
- ・ シミュレーションベースの構築
- ・ まとめと今後の課題

研究背景

- 英語の時代の区分と不規則動詞

Weak Verb(弱変化動詞) : -ed(-t / de)で活用



規則動詞

Strong Verb(強変化動詞) : 母音交替によって活用



不規則動詞

sing – sang – sung

wed – wed – wed

現在ではおおよそ180語存在[2]

英語の時代区分[1]

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

[1] 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)

[2] Pinker, S. The irregular verbs. Landfall 83–85 (Autumn issue, 2000).

研究背景

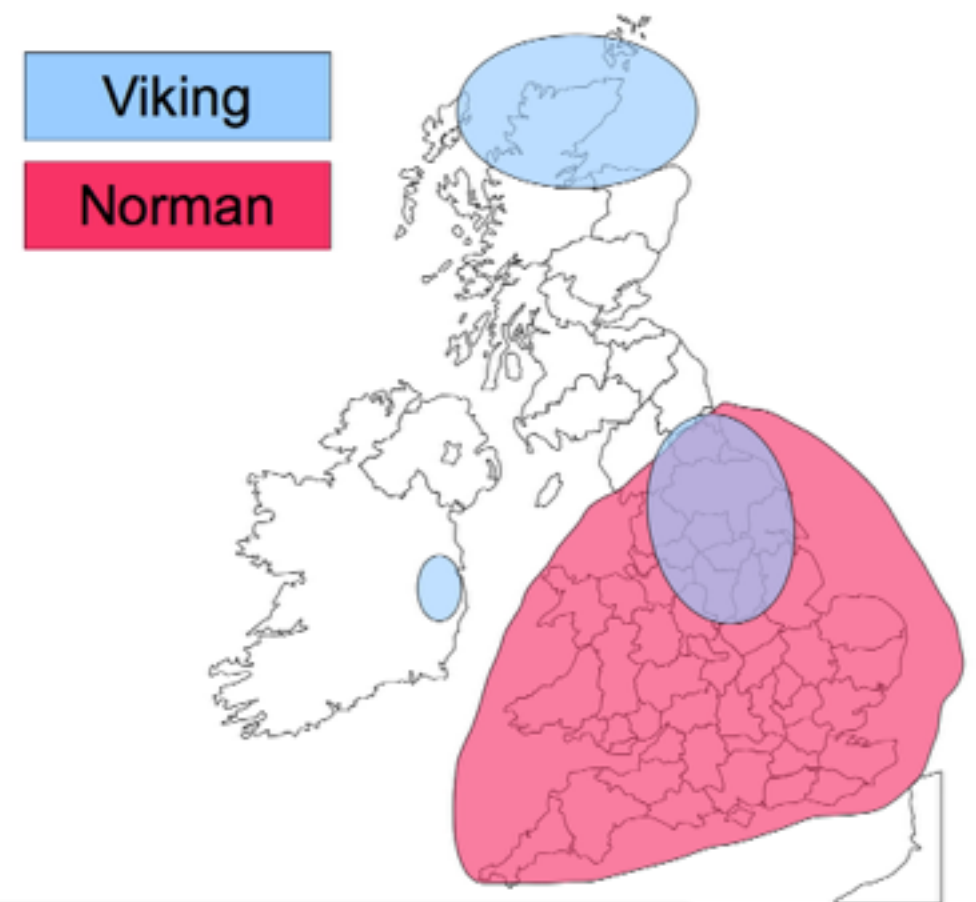
歴史から見る人口流入と言語接触

- A.D 800-1066 Viking (スウェーデン、ノルウェー、デンマーク)

- 3人称単数の語尾に [-s] をつける
- 不規則動詞の規則化 [接尾辞 -ed]が始まる

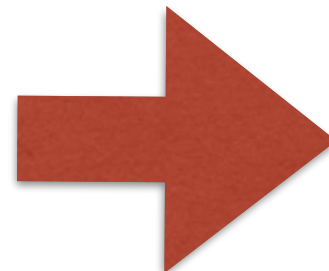
- A.D 1066-1345 ノルマン征服

- 大量の語彙の流入(約 1 万語)、言語構造の簡略
- 不規則動詞の規則化傾向も上昇



英語の特徴

様々な言語接触を経験



不規則動詞の規則化

目的

- 進化計算的方法を利用し、現実の規則化を再現できるマルチエージェントのモデルを構築

人工的な動詞を含むコミュニティを仮定

- 不規則動詞の規則化に関連する人口流入の規模、頻度 を検証



- 規則化の速度 = 動詞の出現頻度と時間の関数(先行研究)
- その速度が保たれる条件となる人口流入の規模、頻度

発表の流れ

- ・ 背景
- ・ 目的
- ・ 先行研究
 - ・ 不規則動詞の出現頻度と規則化速度
 - ・ 語形変化と外圧の影響
- ・ シミュレーションベースの構築
- ・ まとめと今後の課題

出現頻度と規則化

- 規則化の速度 = 動詞の出現頻度と時間の関数として導出[6]

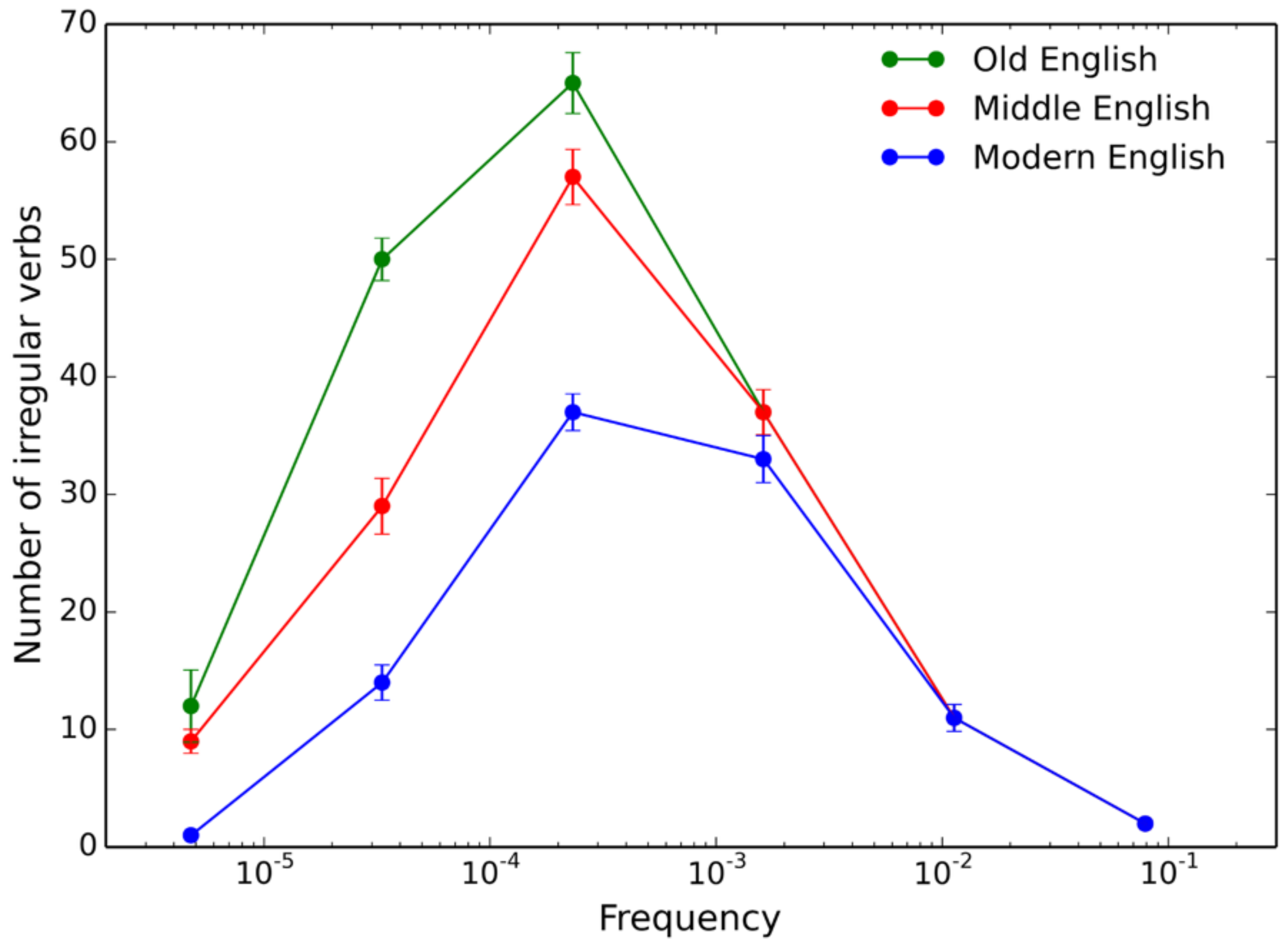
- CELEX database[7]を用いた統計

Old English中の不規則動詞 (177 verbs)

出現頻度	動詞
$10^{-1} - 1$	be, have
$10^{-2} - 10^{-1}$	come, do, find, get, give, go..
$10^{-3} - 10^{-2}$	begin, break, draw, drink, help , reach
$10^{-4} - 10^{-3}$	teach, throw, walk , win, work , write, bake
$10^{-5} - 10^{-4}$	wake, wash , weave, weep, weigh , wind, yell , yield
$10^{-6} - 10^{-5}$	bide , shrive , slink, snip , spew , sup, wreak

[6] Erez Lieberman, Jean-Baptiste Michel, Joe Jackson, Tina Tang & Martin A. Nowak, Quantifying the evolutionary dynamics of language. Nature Vol449 (11 October 2007).

[7] CELEX database <http://wwwlands2.let.kun.nl/members/software/celex.html>



各年代の頻度ごとの不規則動詞の分布[6]


規則化速度

- 時間経過を考慮した規則化速度

- ・ 相対頻度 ω と時間 t の関数
- ・ パラメータ推定を用いて近似

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

- $I(\omega, t)$ の計算例

$10^{-5} - 10^{-4}$ の頻度を持つ 5 0 0 年後の不規則動詞の数 = 9 語  37語から9語に減少

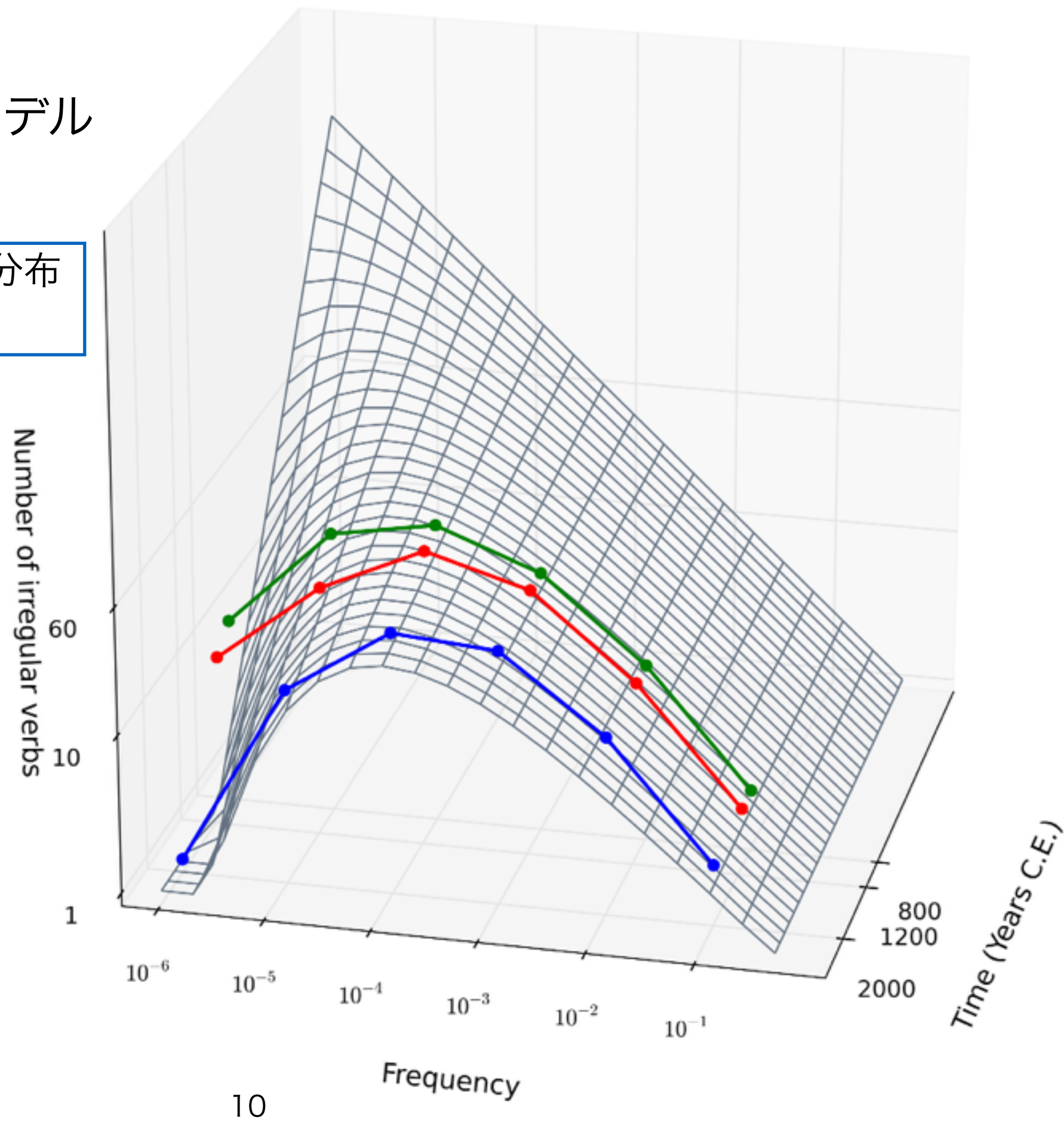
be, haveのどちらかが規則化されるまで場合38800年以上

- 本研究の目的となるモデル

過去一未来の不規則動詞数の分布
規則化までの時間



予測可能



語形変化とSigmoid

- 不規則動詞、綴り字の改正など語形変化に関する統計的研究

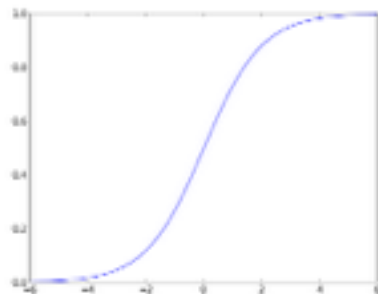
- 不規則動詞の場合(規則化された割合のような)

$$\rho(t) = \frac{n(-ed)}{n(-ed) + n(irregular) + n(pastparticiple)}$$

- Google Ngram Corpus[8]を利用
- 1800~2007年まで各年で計算

- Sigmoid曲線に近似 $\frac{d\rho(t)}{dt} = (a + b\rho(t))(1 - \rho(t))$

a: 変化に影響を与える外圧
b: 接触確率



外部からの力が加わる
内部で起こった要因のみ

近似値を取り出す

[7] Ghanbarnejad, Fakhteh, et al. "Extracting information from S-curves of language change." arXiv preprint arXiv: 1406.4498 (2014).

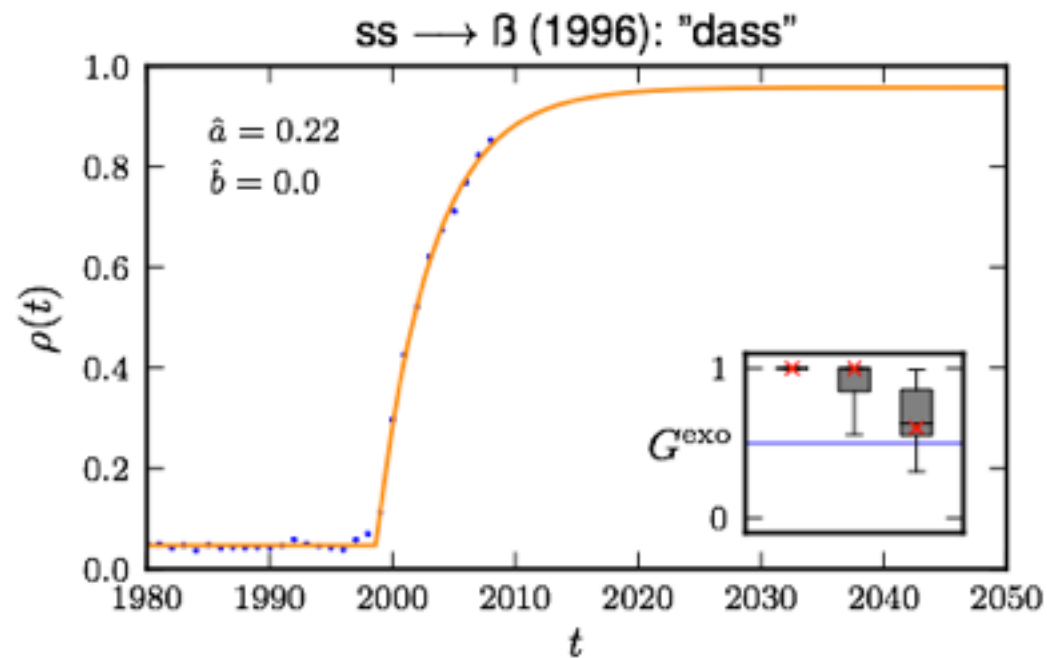
[8] Google Ngram Viewer , <https://books.google.com/ngrams>

外圧の影響

大きな外圧の影響で急激に変化

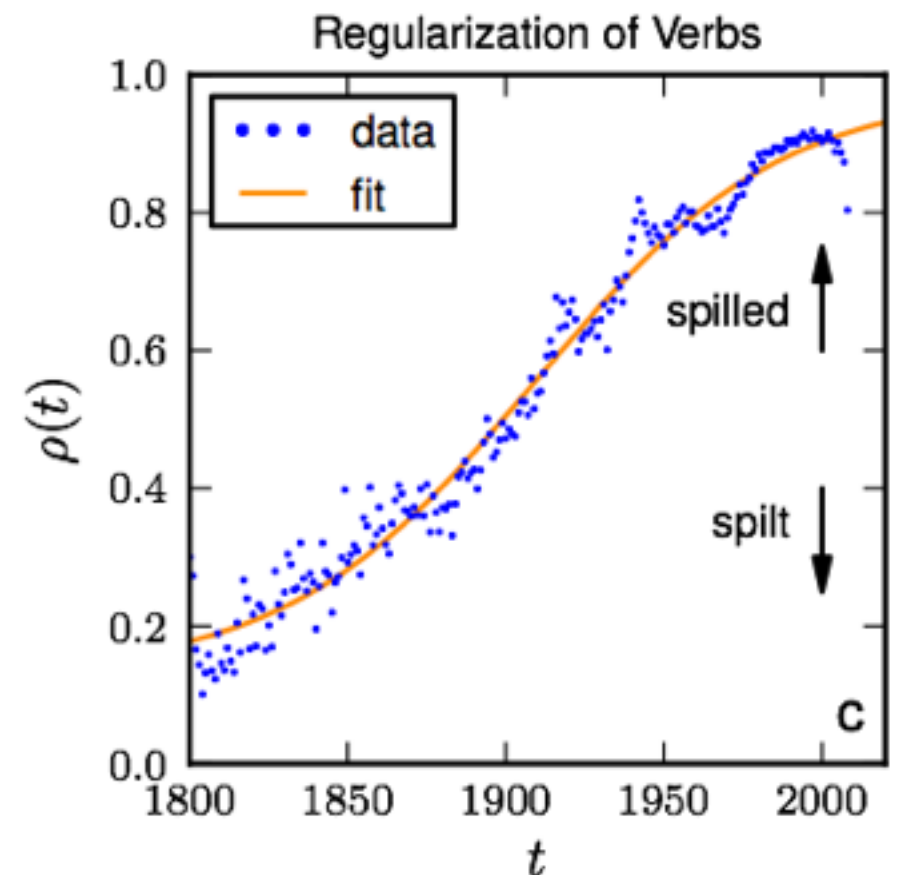
- ドイツ語の綴りの改正

1996: ss → β



$a=0.22$
 $b=0.0$

緩やかな変化では外圧が小さい



$a=0.001$
 $b=0.030$

シミュレーションの意義

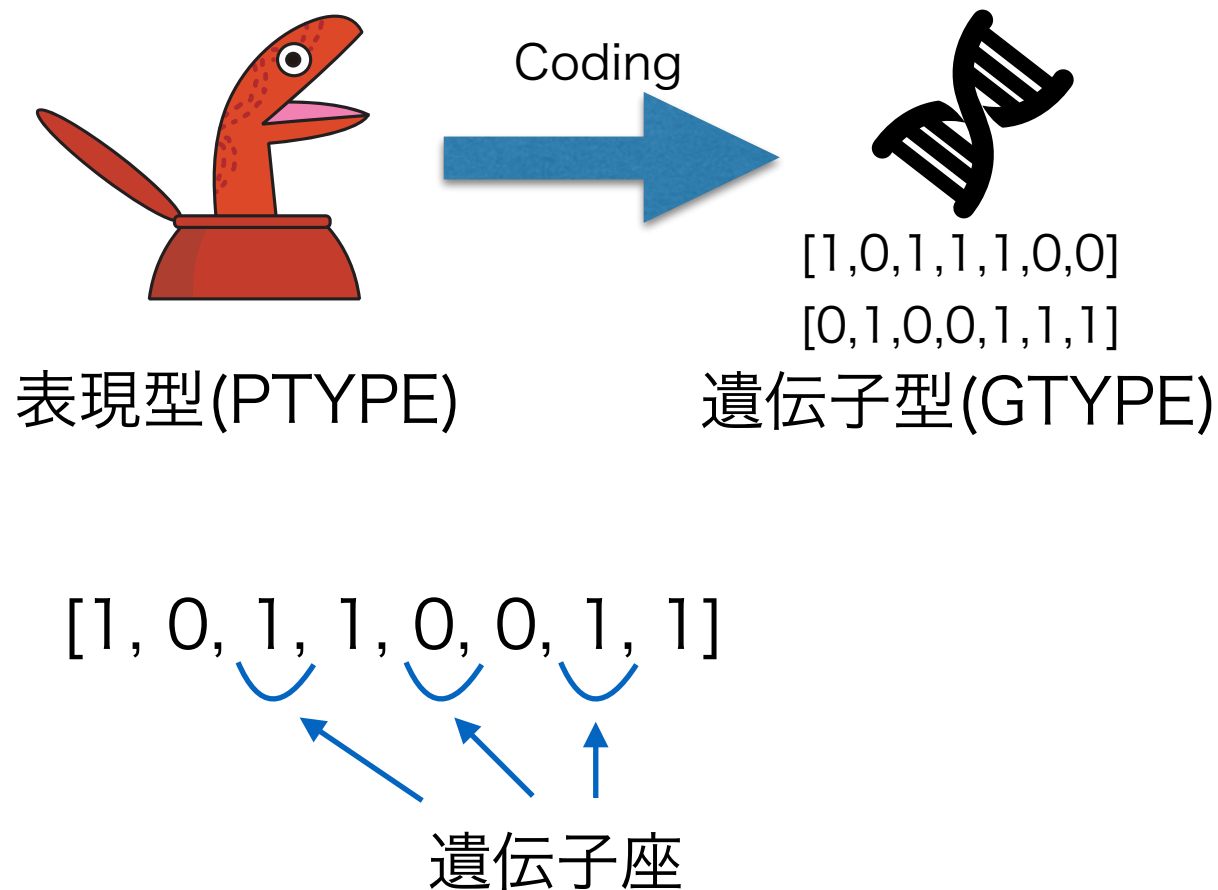
- 先行研究では実際の統計データ

発表の流れ

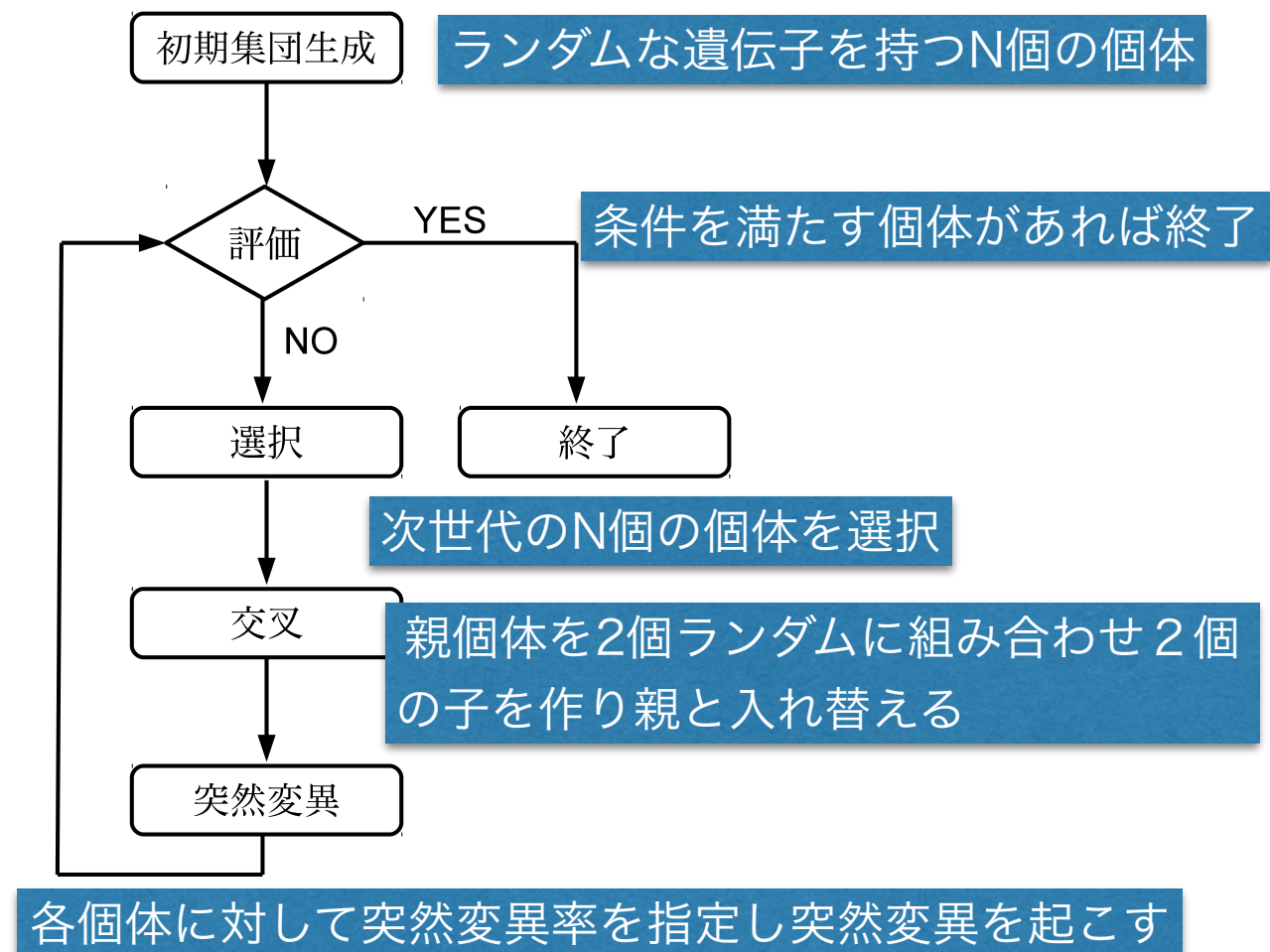
- ・ 背景
- ・ 目的
- ・ 先行研究
 - ・ 不規則動詞の出現頻度と規則化速度
 - ・ 語形変化と外圧の影響
- ・ シミュレーションベースの構築
- ・ まとめと今後の課題

遺伝的アルゴリズム(GA)

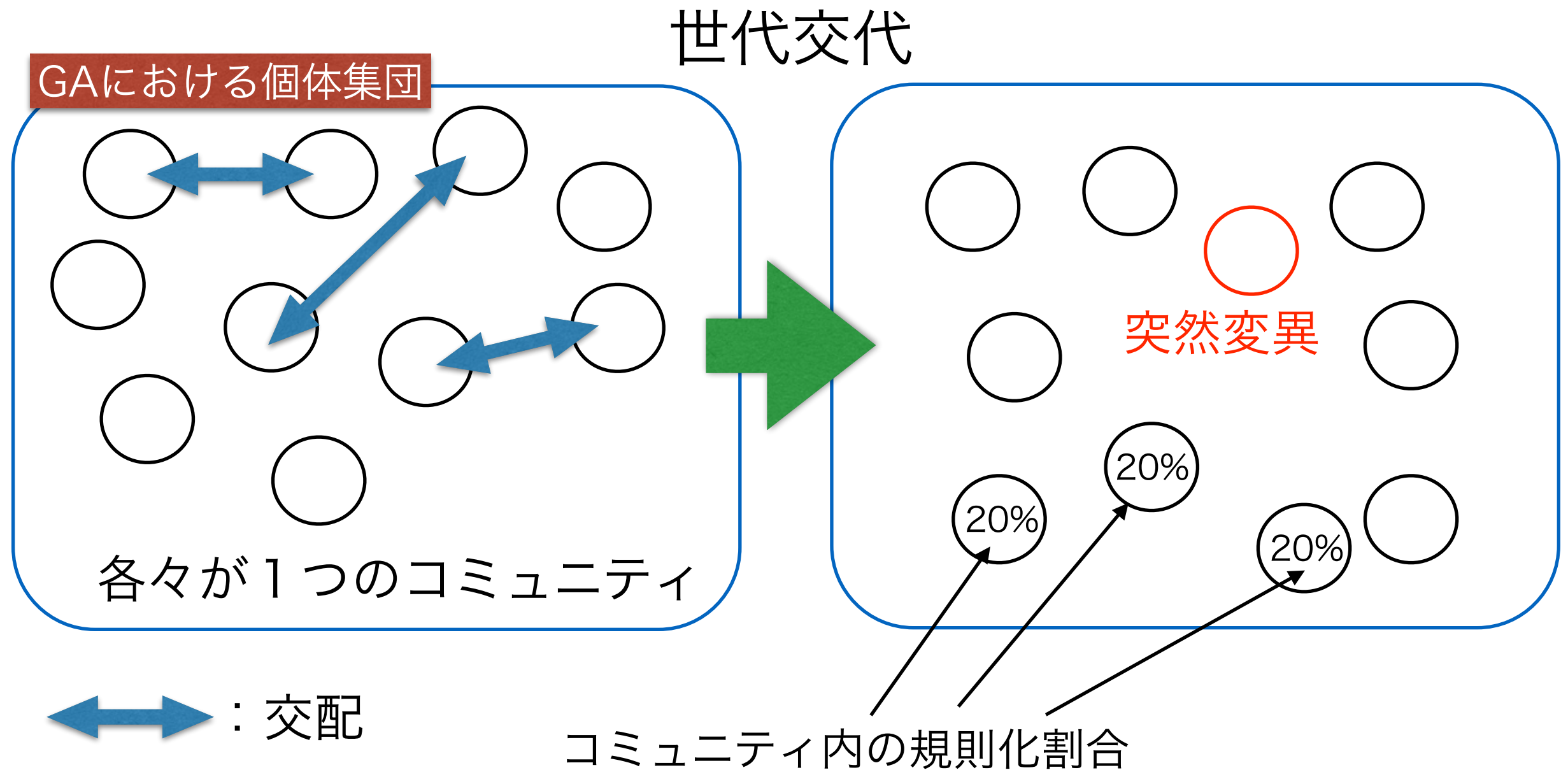
- Coding(Encoding)



- Evolution flow



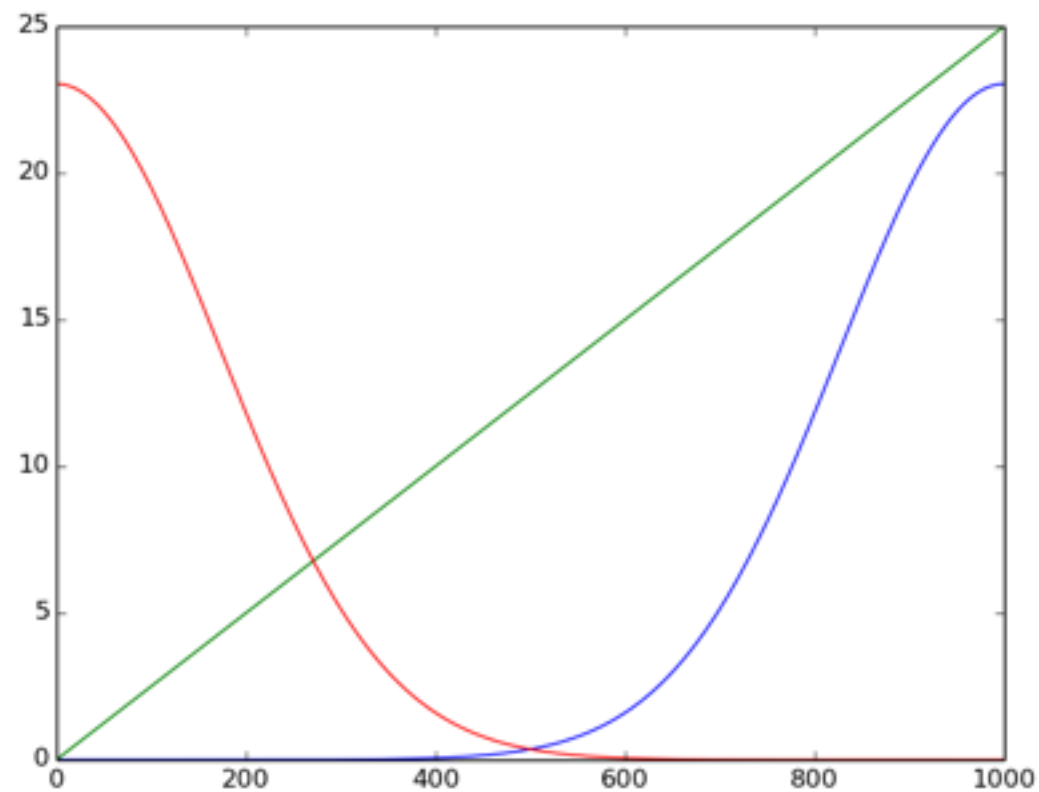
GAを用いた規則化進行の表現---1/4



GAを用いた規則化進行の表現---2/4

● 適応度

ある動詞の使いやすさを表すような関数

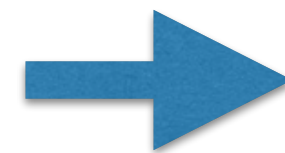


不規則変化のまま

規則化

● 関数の形を決める要因

- ・ 使用頻度
- ・ 会話をしたときの伝わりやすさ
- ・ 音韻論における制限、ストレス



常に変化に晒される

発音しにくい



発話

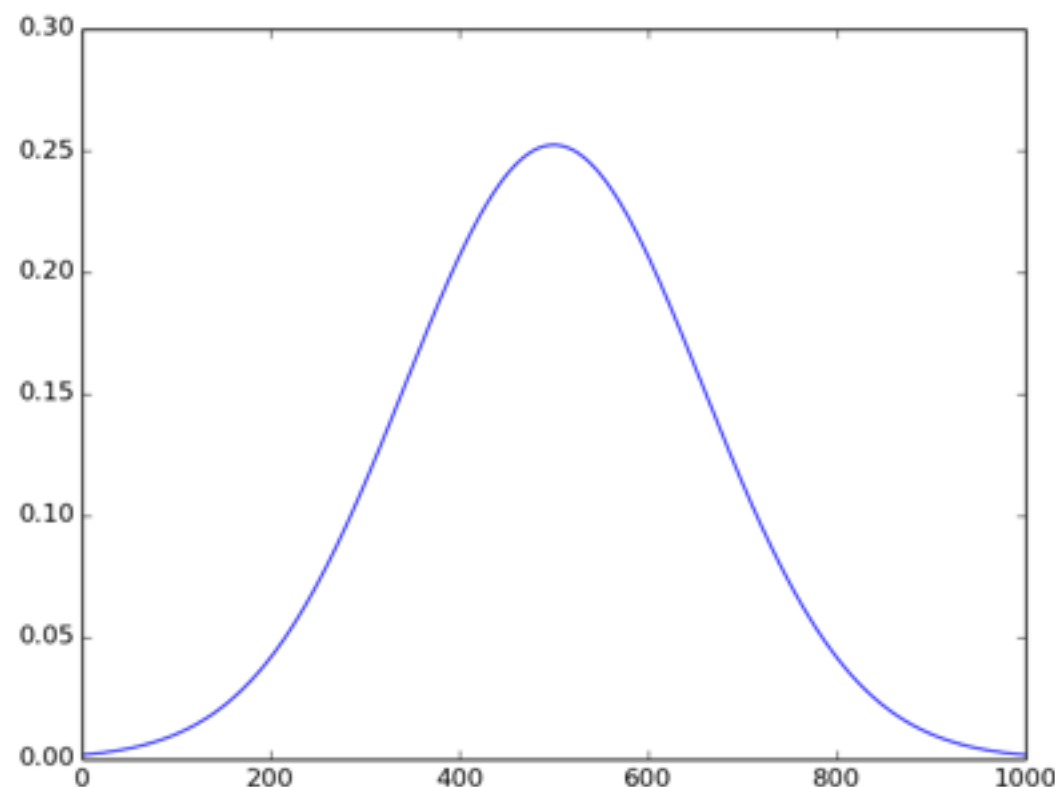


?



GAを用いた規則化進行の表現---3/4

- バイアス(デモで使用)



- ・ 正規分布を模倣した形

変化の始まりと終わりは加速させる力は弱く
変化の途中であるほど大きい

- GA内でのバイアスの実現

突然変異オペレータを変更

- ・ 外圧をトリガーとして発動
バイアスに従い整数を発生させ、
その分各個体のPtypeを上昇
- ・ 外圧の影響が徐々に浸透していき
浸透しきれば影響は少なくなる

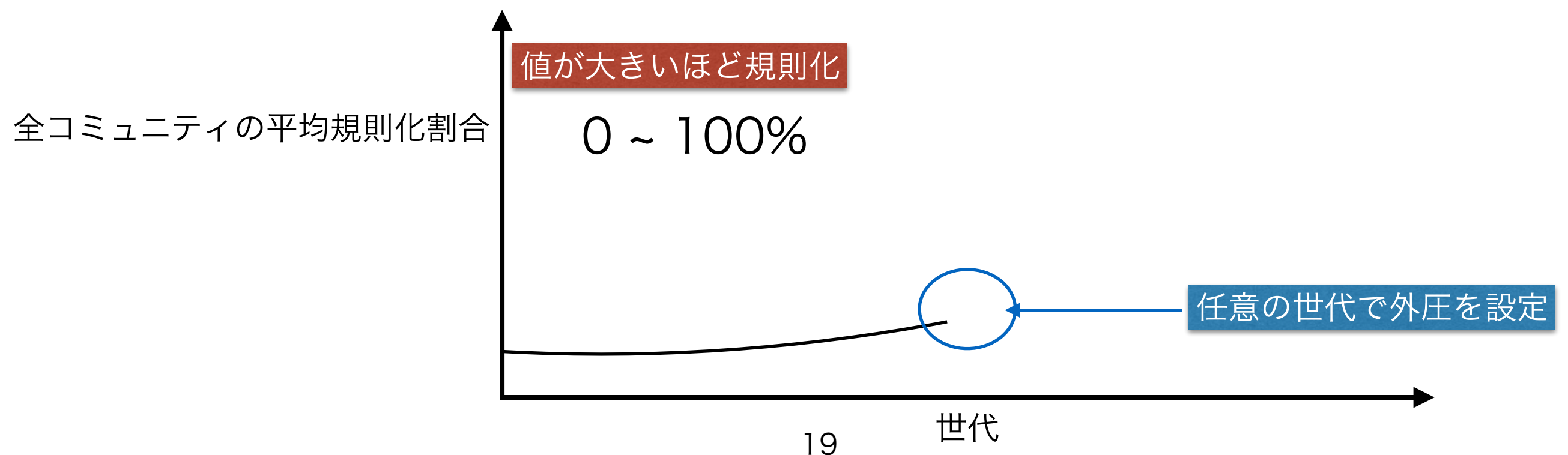


数世代にわたって影響

GAを用いた規則化進行の表現----4/4

- 設定のまとめ

- ・ 個体はコミュニティを表現している
- ・ 個体の適応度は規則化割合or留まる力
- ・ 外圧がかかればバイアスによって変化が加速される



シミュレーション

- 設定

- ・ 個体

PTYPE : 200(20%)

GTYPE : 2進数(10bit)

集団数 : 2500

- ・ 繰り返す世代 : 150世代

- ・ 選択 : トーナメント方式(サイズ2)

- ・ 交叉

交叉確率 : 0.6

方法 : 一様交叉

- ・ 外圧がかかる世代 : 50世代目

- ・ 外圧の影響が及ぶ期間 : 20世代

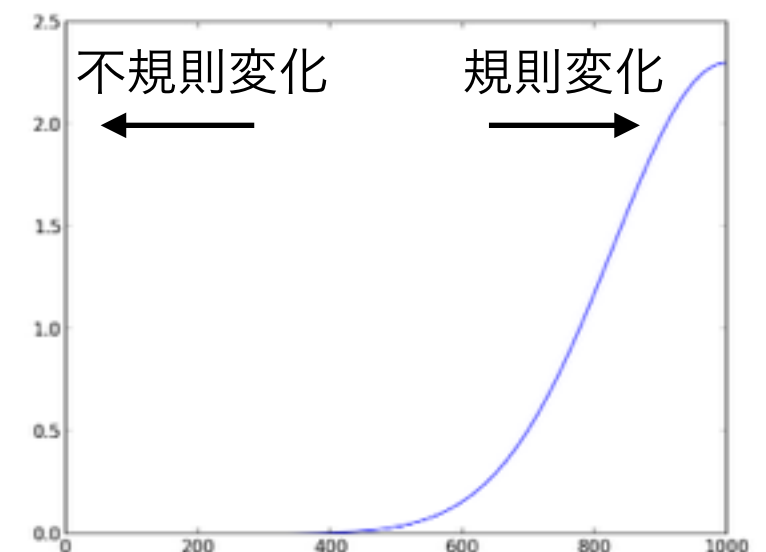
- ・ 適応度

正規分布

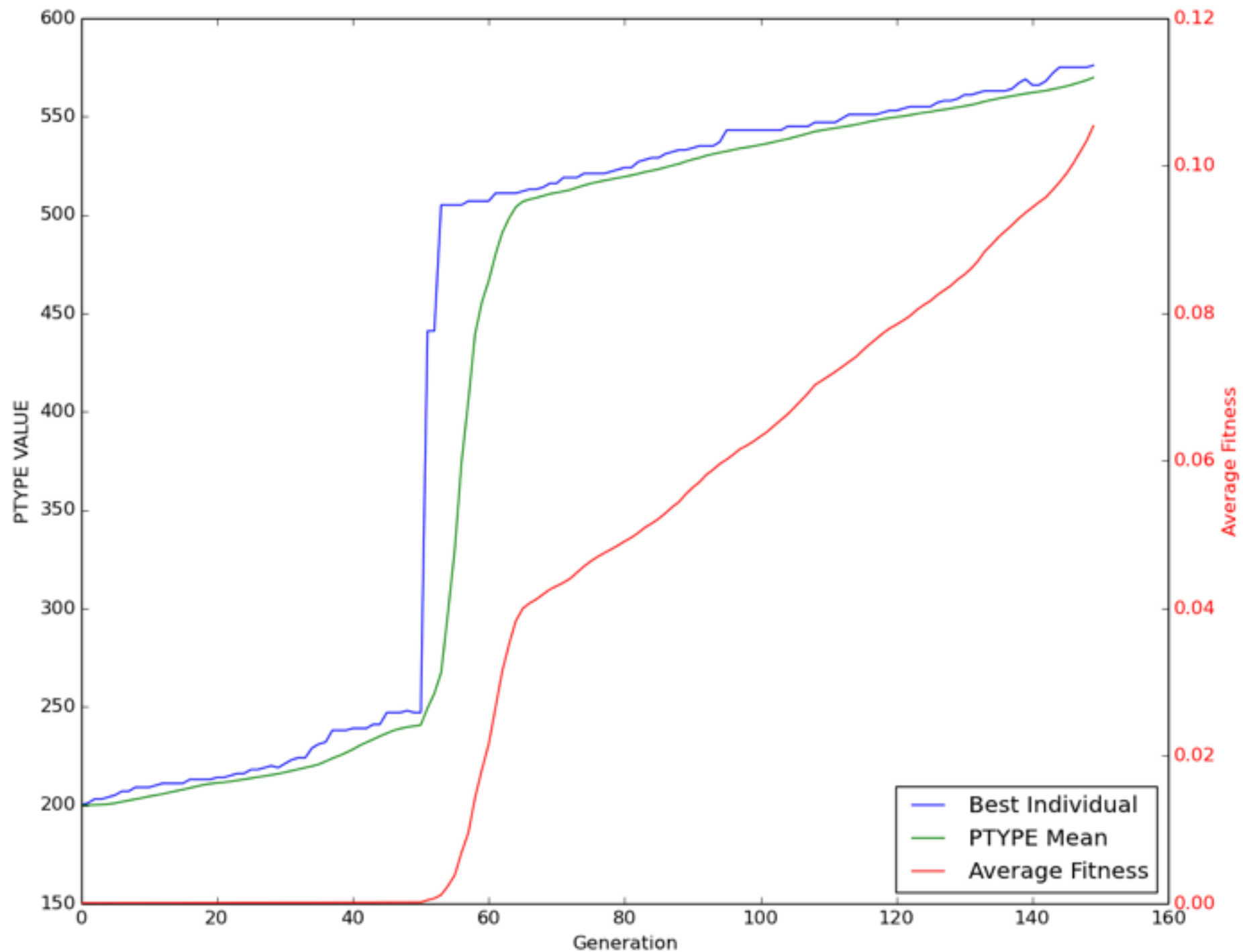
分散:30000

平均:1000

規則化された方が適応度が高い



100回繰り返した平均



発表の流れ

- ・ 背景
- ・ 目的
- ・ 先行研究
 - ・ 不規則動詞の出現頻度と規則化速度
 - ・ 語形変化と外圧の影響
- ・ シミュレーションベースの構築
- ・ まとめと今後の課題

まとめと今後の課題

- シミュレーションベースの作成
 - ・ 外圧をGAの中の意図的な突然変異として表現
 - ・ 抽象的な設定で不規則動詞が外圧によって急速に規則化されていく様子をシミュレートできた

課題

- ・ 不規則化への対応とバイアスの動的な設定
- 人工的な動詞を含んだコミュニティの設定
 - ・ 人工動詞の作成
 - ・ 外圧と人口流入の設定
 - ・ エージェントコミュニケーションによる適応度関数作成