



日本語形容詞の通時的な 意味変化の傾向

～日本語の変化を可視化する～

兵庫県立姫路西高等学校

2年7組

改發尚雅 三木駆人 渡邊佳歩



目次

Table

1.研究背景

2.研究目的と意義

3.研究方法

4.結果と考察

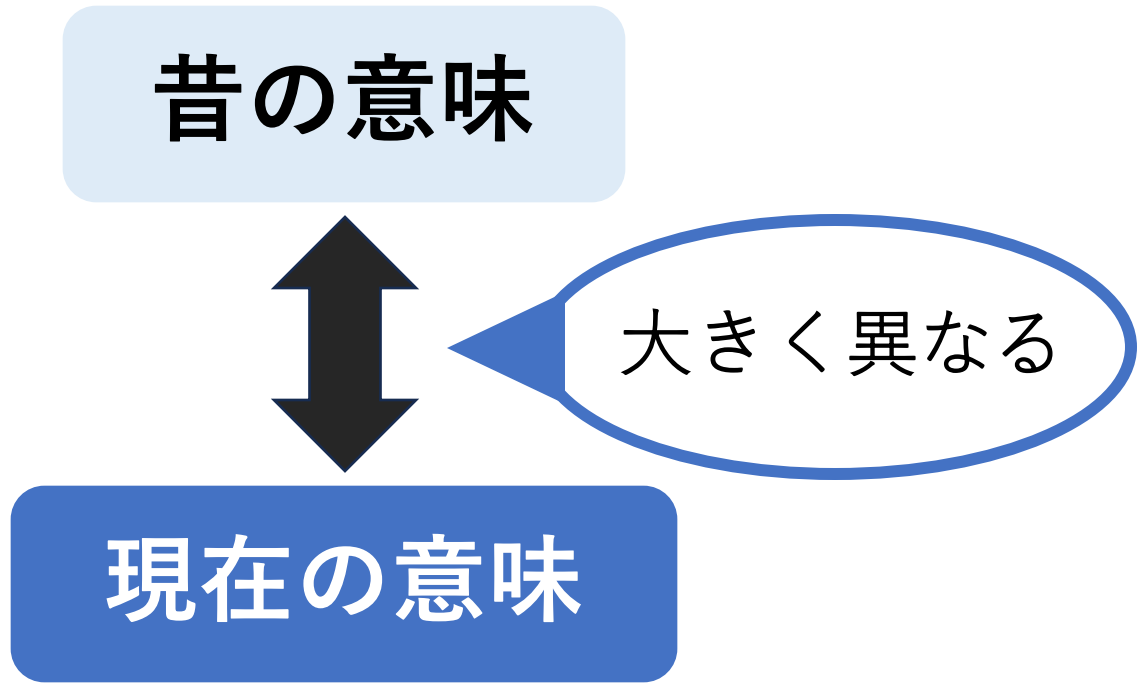
5.結論と展望



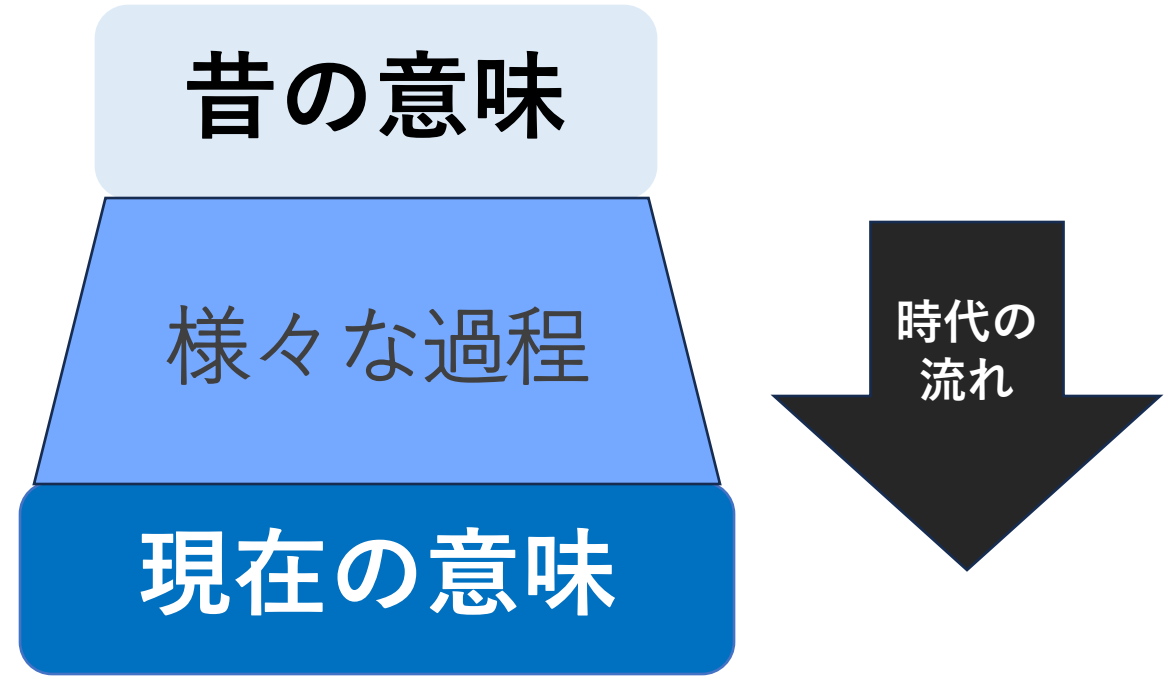
研究背景
Background



研究の動機



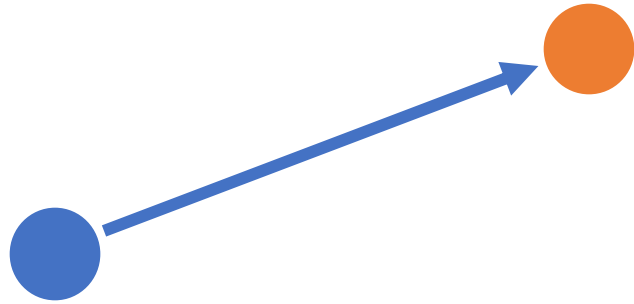
単に**意味の違い**→よく注目される
古典の授業でも学ぶ



どのように**変化したか**→注目されない
ここに興味を持ち研究を始めた

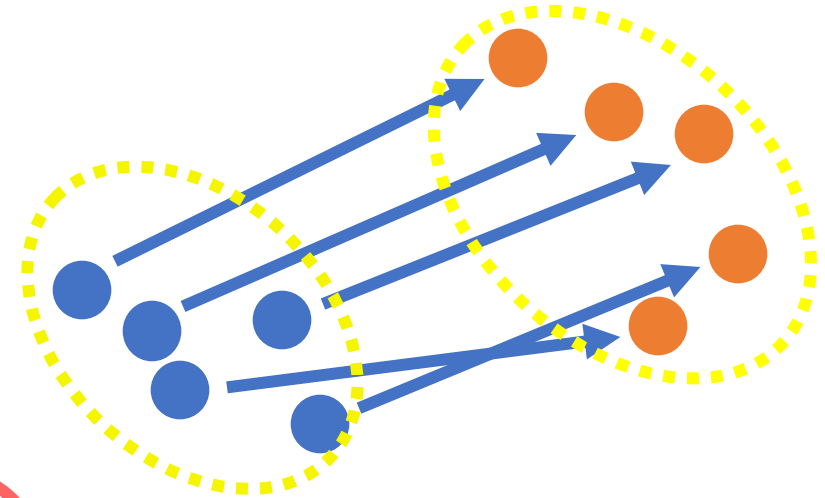
研究の動機

単語一つの意味変化



様々な単語に当てはまる意味変化の傾向を調べられない

複数の単語の意味変化



一般的な意味変化の傾向を調べることができる

線としての変化→面としての変化



研究の動機

「機械学習を用いて言葉の意味を解析したい」

先行研究

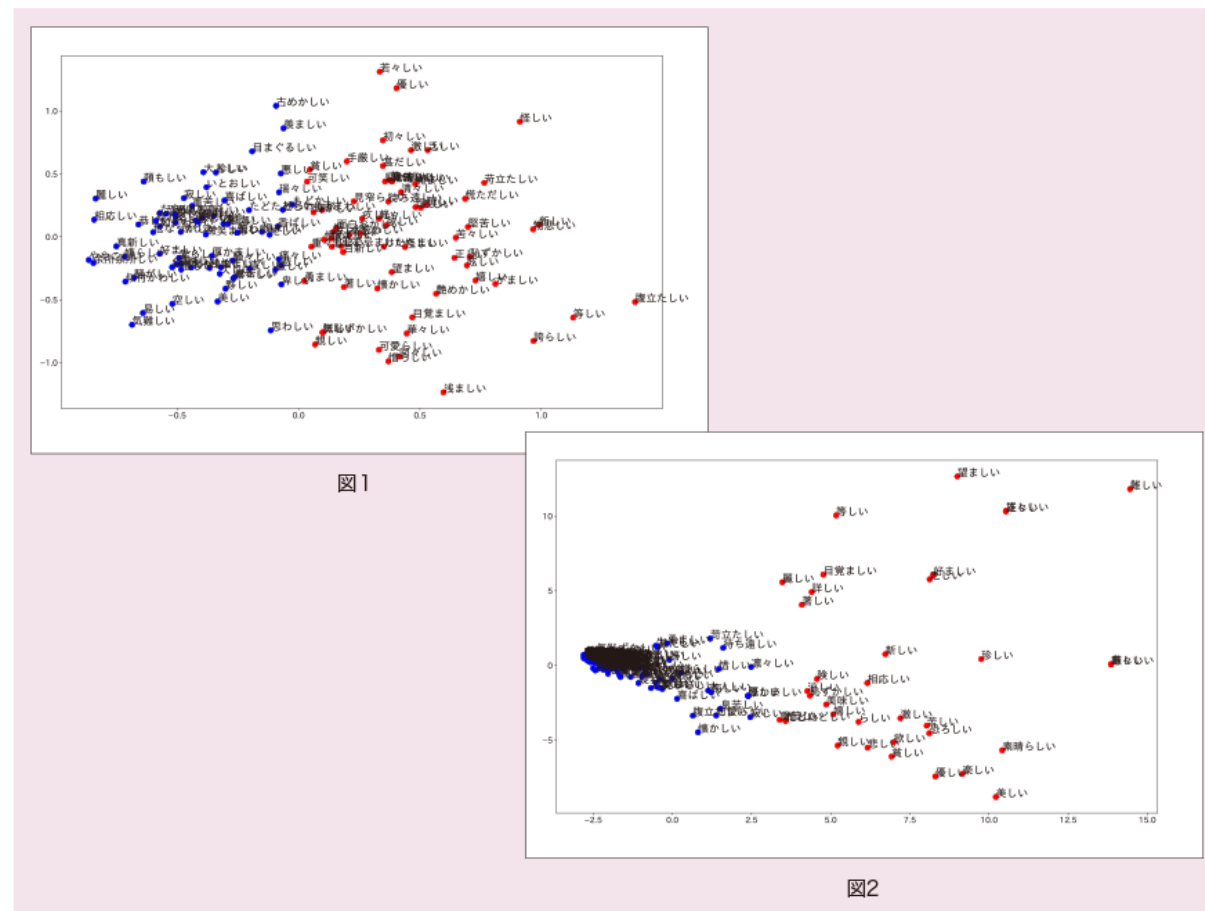
先行研究

近藤泰弘 (2022)

源氏物語で使われているシク形容詞の意味をWord2vecを使ってベクトル化し、現代での意味と比較考察している。

図1：源氏物語のシク形容詞をベクトル化した散布図

図2：現代語の、シイを語尾とする形容詞をベクトル化した散布図





単語の分散表現とは

文章中の単語の意味を複数の次元（50から300次元が多い）のベクトルとして表現すること

単語	球技	個人競技	食べ物	...
テニス	0.93	0.88	0.14	...
野球	0.88	0.13	0.07	...
陸上競技	0.24	0.81	0.11	...



Word2vecとは

文章中で一緒に使われた単語から、機械学習によってある単語の意味をベクトルとして表現し、数値化する

(例) わたしは赤いりんごを食べます。

わたしは赤い「 」を食べます。

「」の中はりんご？いちご？



研究目的と意義
Purposes and Significances



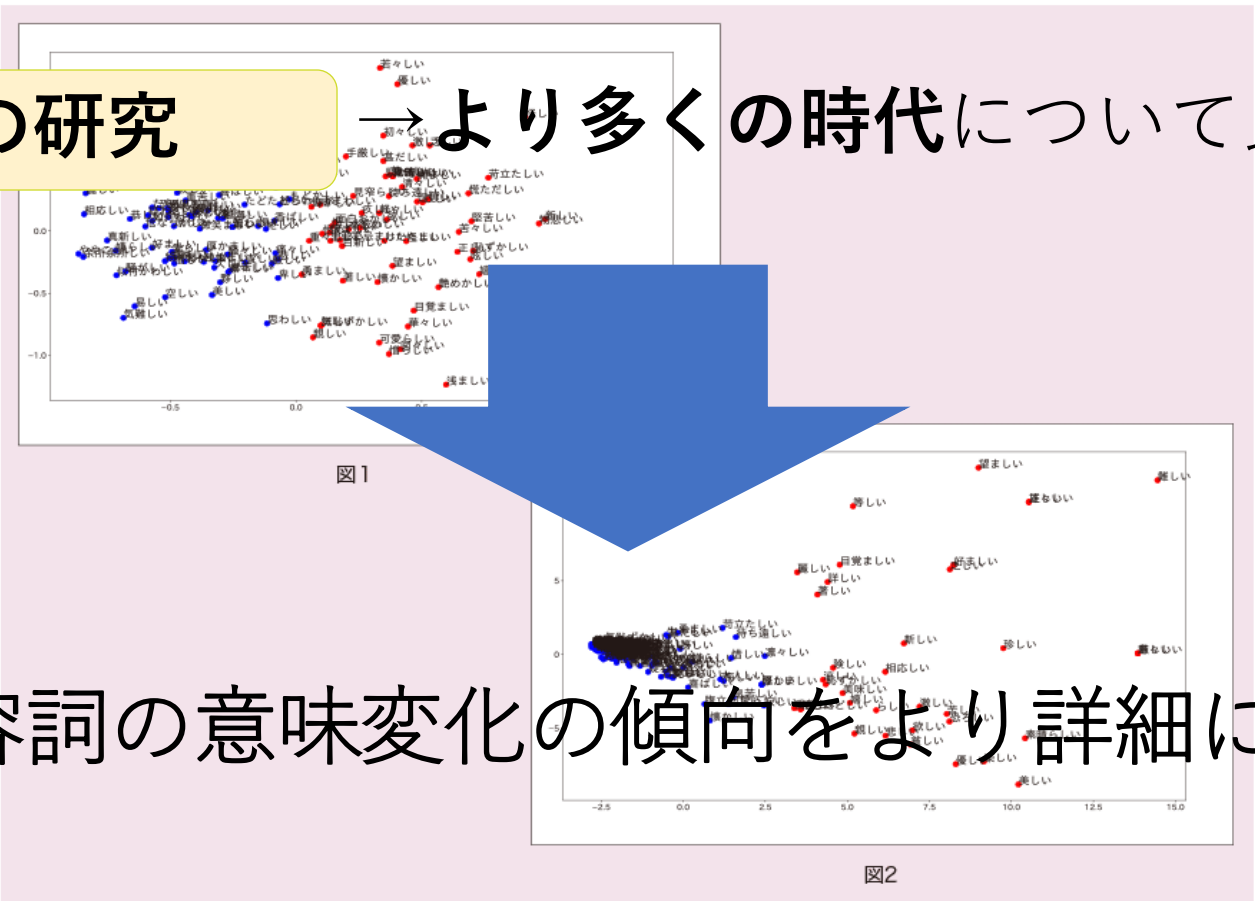
研究目的

近藤泰弘(2022)

→二つの時代について比較

今回の研究

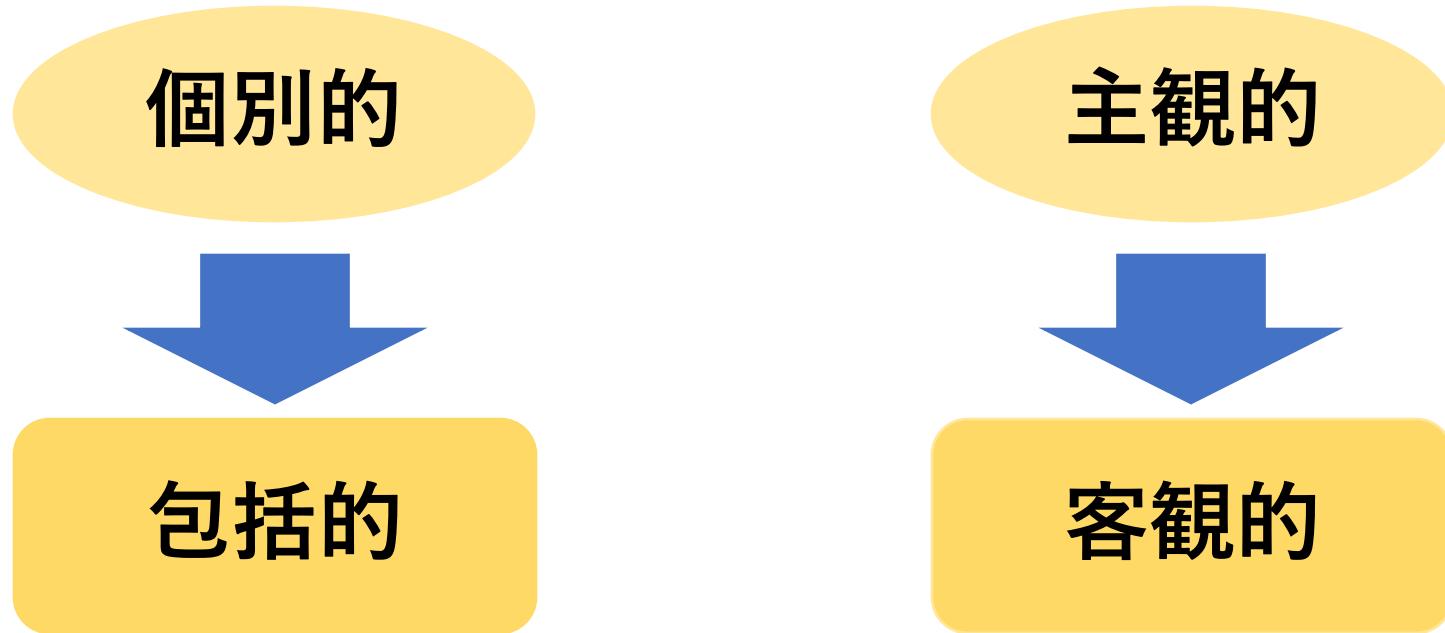
→より多くの時代について比較



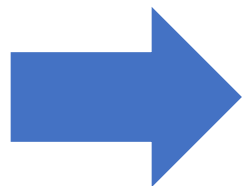
日本語形容詞の意味変化の傾向をより詳細に調べたい



研究意義



機械学習を用いた日本語研究を発展させる



- ・ 古典学習を新しい視点から作り上げる
- ・ 限りのある古典資料だけに頼る必要がなくなる

大きな可能性



研究方法
Methods

研究手法①資料の取得

11世紀～21世紀から4つの期間をとり、単語ベクトルを作るための資料となる文献を取得する。

中古	11世紀～12世紀（1001年～1200年）	平安時代～鎌倉時代
中世	14世紀～15世紀（1301年～1500年）	室町時代～戦国時代
近世	17世紀～18世紀（1601年～1800年）	江戸時代
現代	20世紀～21世紀（1901年～）	現代

- 文献は、インターネット上で公開されているものを使用



時代 (n=モデルの総語数)	使用した資料(成立時期)
中古 (n=263930)	<ul style="list-style-type: none">・更級日記(1059年ごろ)・源氏物語(1008年ごろ)
中世 (n=321707)	<ul style="list-style-type: none">・風姿花伝(1400年)・徒然草(1330年ごろ)・とはずがたり(1306~1313年)・太平記(1349年ごろ)



近世

(n=160048)

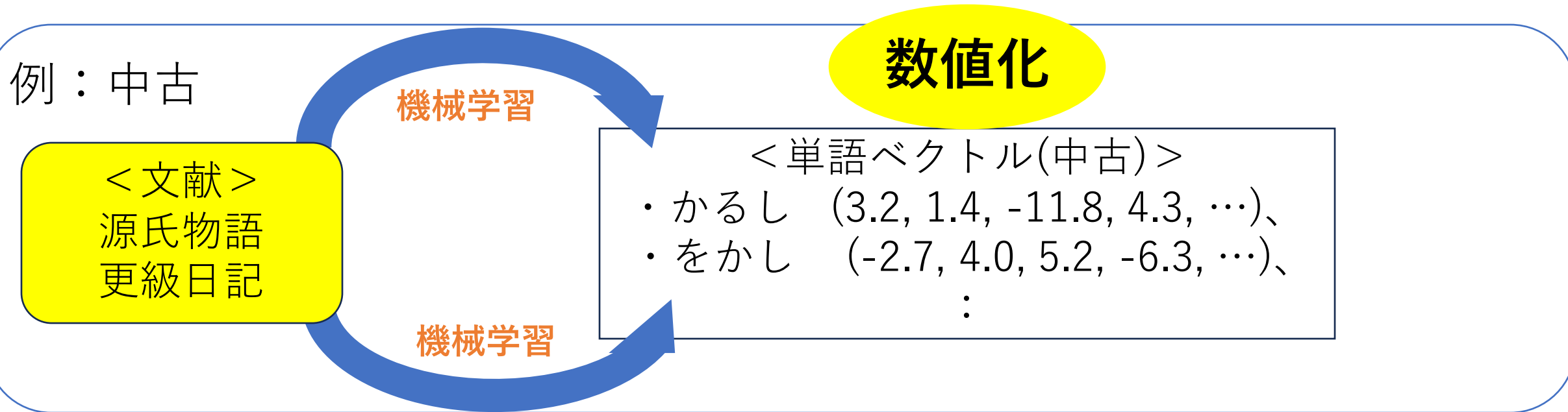
- ・ 醒睡笑(1623年)
- ・ 奥の細道(1702年)
- ・ 雨月物語(1776年)
- ・ 養生訓(1712年)
- ・ 駿台雑話(1732年)
- ・ 世間妾形気(1767年)
- ・ 諸道聴耳世間猿(1766年)
- ・ 仮名手本忠臣蔵(1748年)

現代

学習済み単語ベクトル(Wikipediaエンティティベクトル)を使用したため資料はなし

研究手法②単語の分散表現

1. 取得した文献をもとに単語の分散表現を行う



2. どの時代でも5回以上使われている形容詞を抽出

研究手法③: 散布図に可視化

単語ベクトルの次元数を主成分分析によって2に減らす

例: 中古

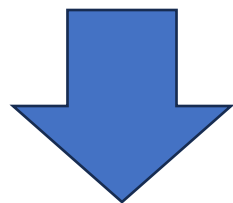
・ かるし (3.2, 1.4, -11.8, 4.3, ...)

200次元

・ をかし (-2.7, 4.0, 5.2, -6.3, ...)

⋮

⋮



主成分分析

・ かるし (-0.2, -2.0)

2次元

・ をかし (2.1, 8.5)

⋮

⋮

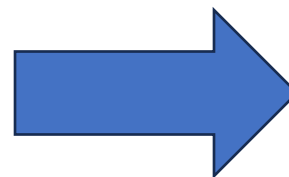
主成分分析とは

主成分分析…計算によって、多くの変数を持つデータから特に重要な要素を取り出し、少ない変数に要約する手法

例：テストの点数

	国語	数学	英語	社会	理科
Aさん	53	33	68	68	30
Bさん	74	75	82	75	81
Cさん	67	82	56	76	88
⋮	⋮	⋮	⋮	⋮	⋮

主成分分析



	総合学力	文理
	第1主成分	第2主成分
Aさん	-1.95	0.99
Bさん	2.81	0.22
Cさん	2.34	-1.74
⋮	⋮	⋮

ただし、主成分(軸)の持つ意味は分析者が考察する必要がある！

研究手法③: 散布図に可視化

単語ベクトルの次元数を 主成分分析 によって 2 に減らす

例：中古

- ・ かるし (3.2, 1.4, -11.8, 4.3, ...)
- ・ をかし (-2.7, 4.0, 5.2, -6.3, ...)
- ：　　　　：



主成分分析

- ・ かるし (-0.2, -2.0)
- ・ をかし (2.1, 8.5)
- ：　　　　：

2次元



散布図に表示
クラスタリングを行う

この操作を
各期間で行う！

200次元

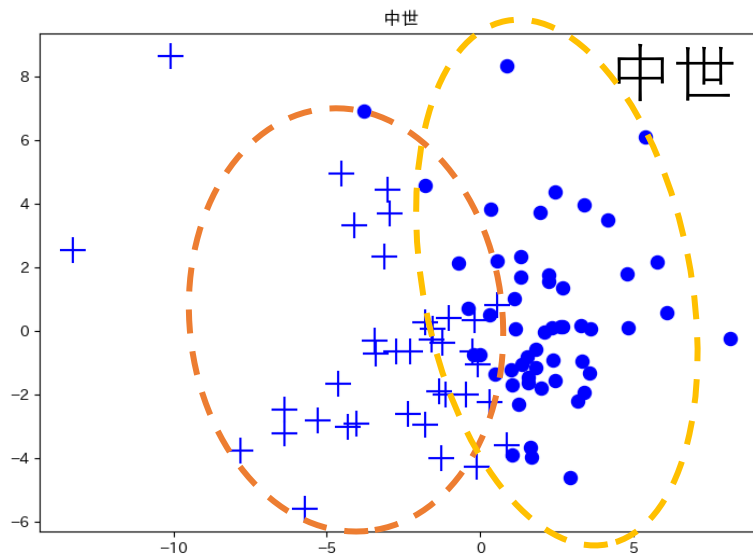
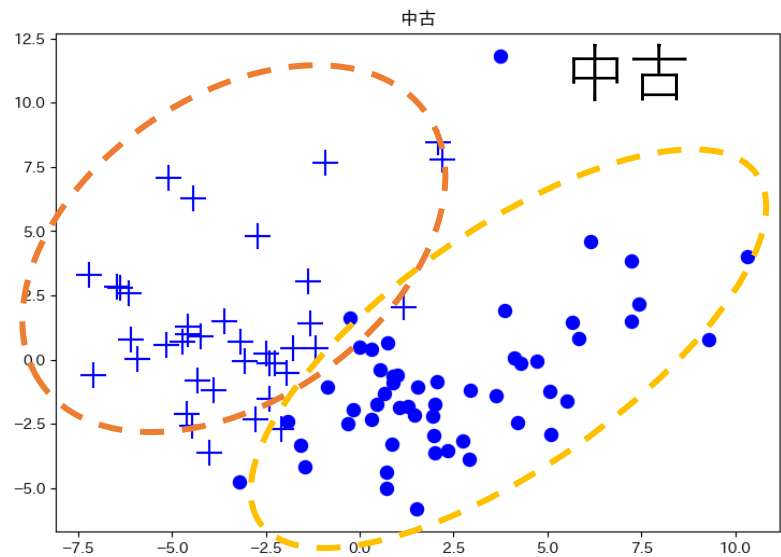


結果と考察

Results and Considerations

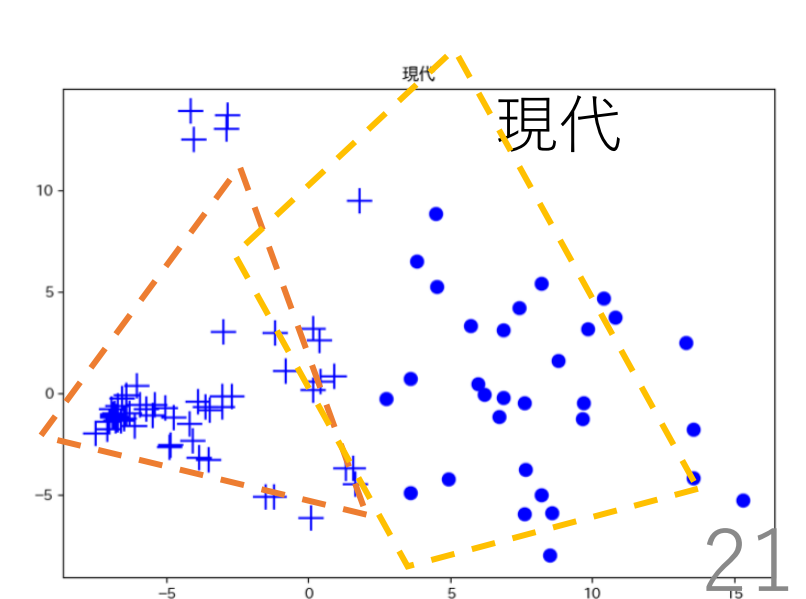
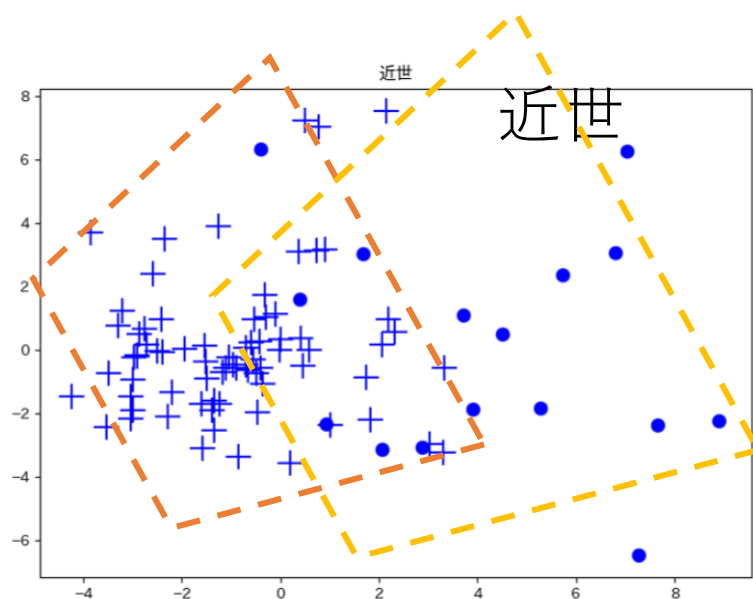


結果



共通の軸(主成分)
が見出せた

左側 (+)
→主観的な意味を持つ
形容詞が多い
右側 (● ☒)
→客観的な意味を持つ
形容詞が多い





結果からの読み取り: グラフの横軸と縦軸

具体的
形のあるものを表す 例: 長い

例: 悲しい

例: 少ない

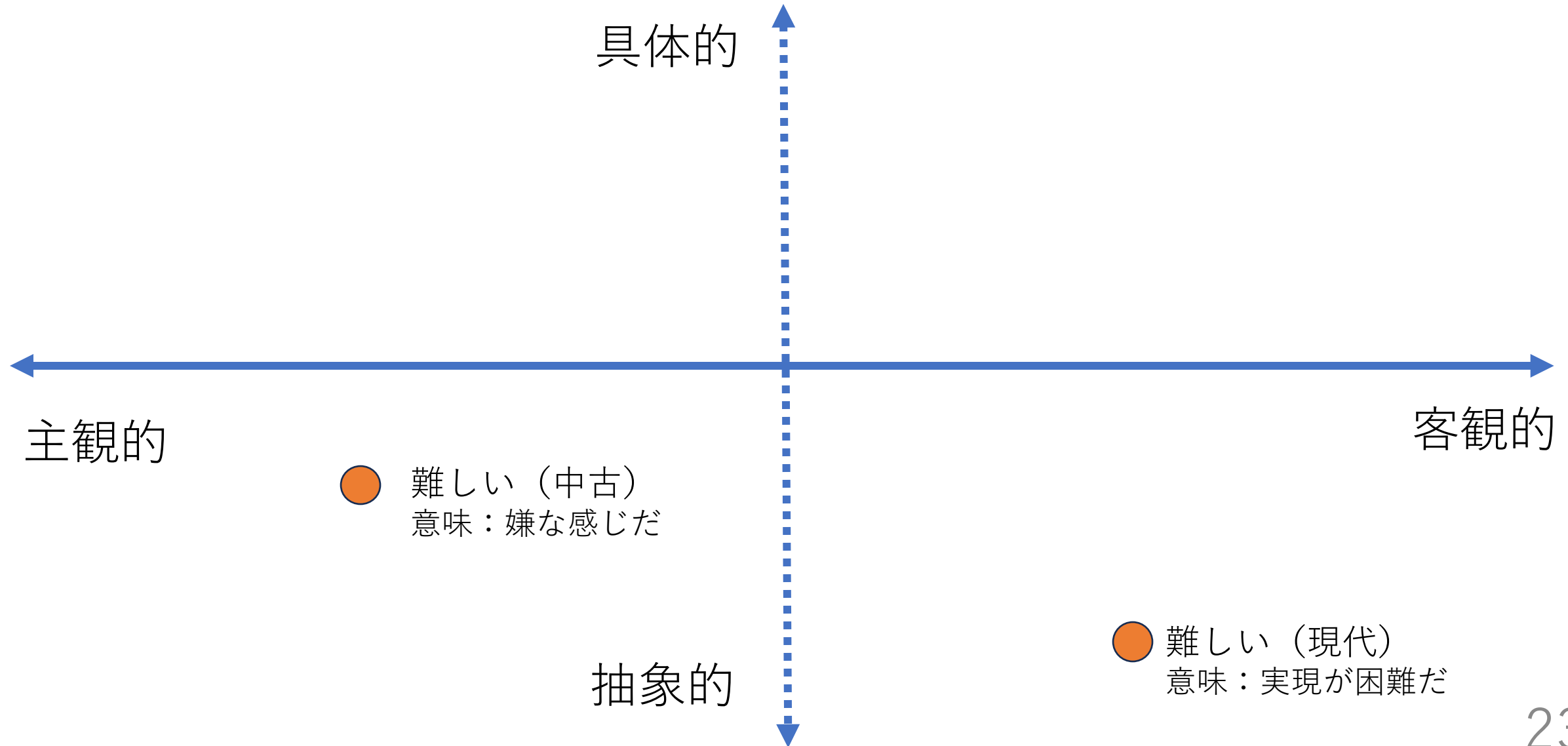
主観的
ex.感情

客観的
ex.状態

抽象的
形のないものを表す 例: 遅い



結果からの読み取り: グラフの横軸と縦軸





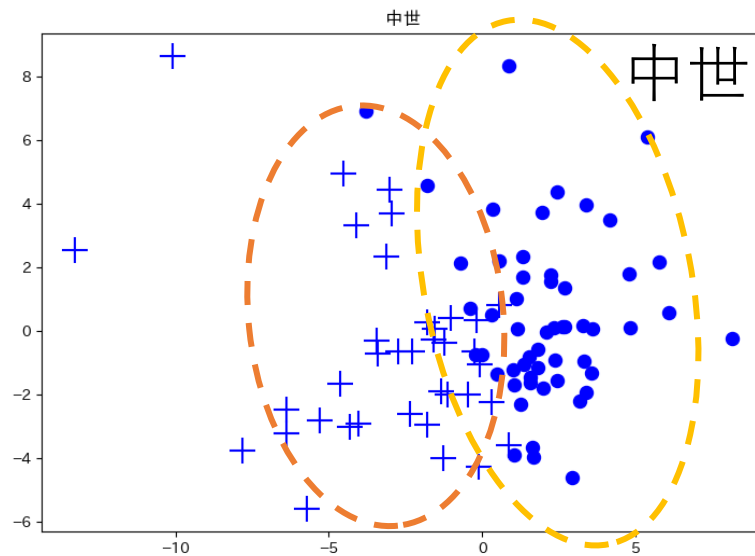
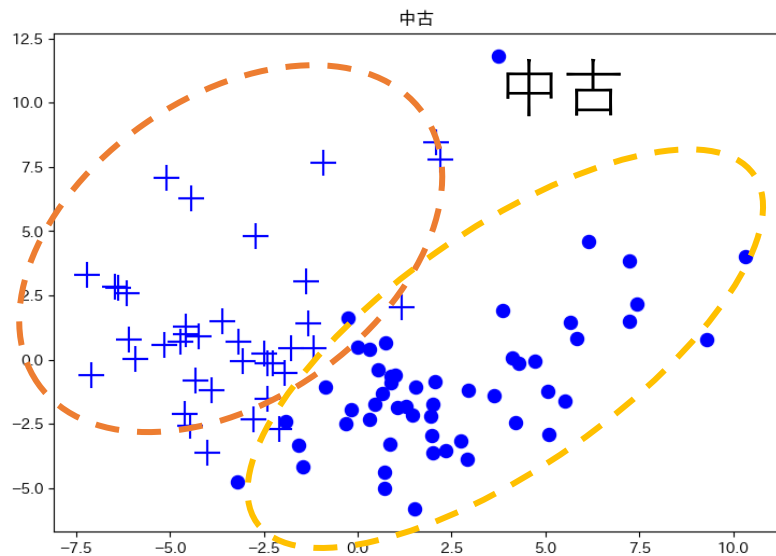
考察

質的变化

量的变化



考察①: 形容詞の意味の質的変化傾向



	現代
	60
	31

左(主観)→左(主観)

怪しい おかしい 辛い 悲しい 恐ろしい

左(主観)→右(客観)

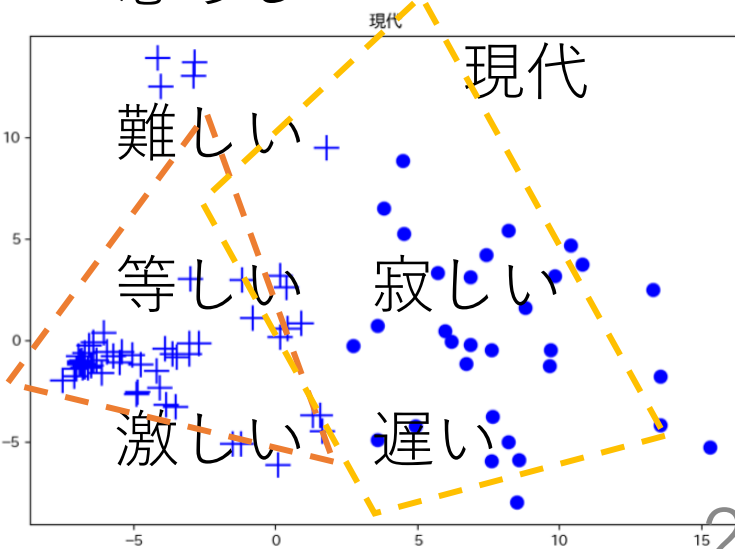
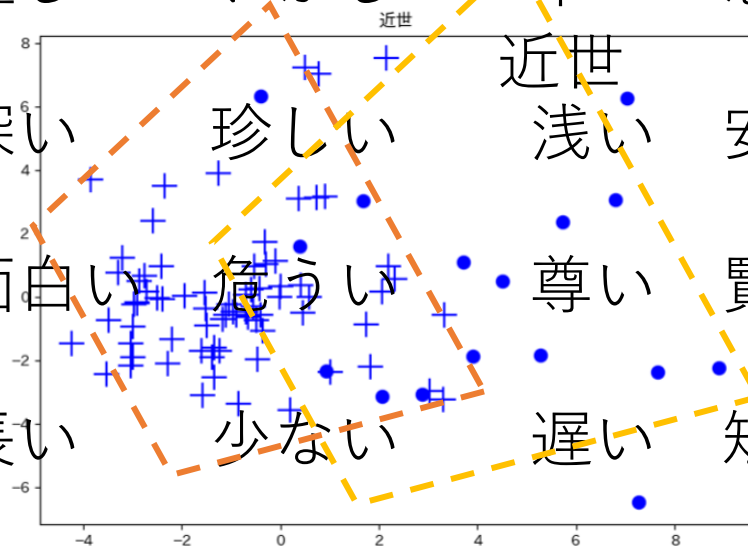
深い 珍しい 浅い 安い

右(客観)→左(主観)

面白い 危うい 尊い 賢い

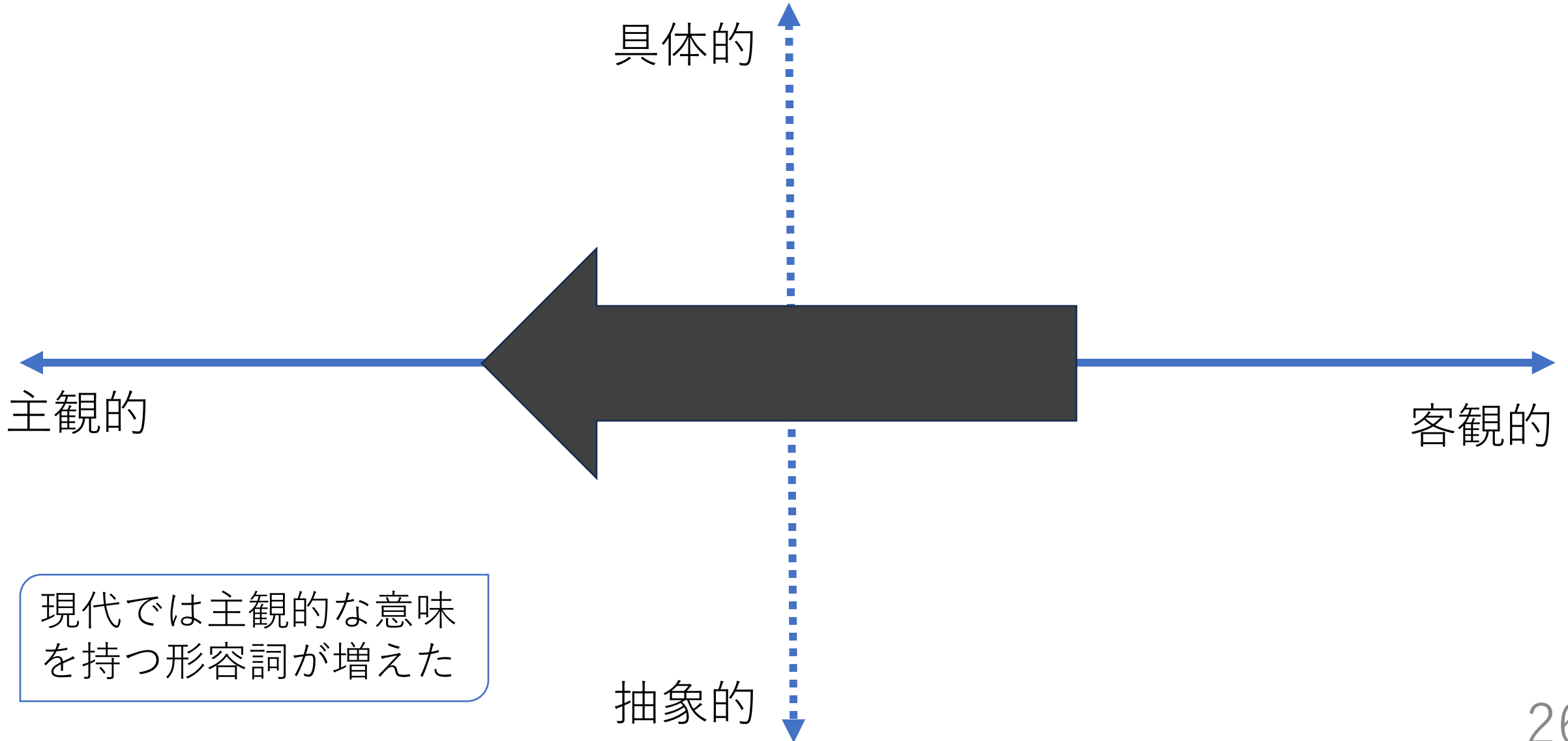
右(客観)→右(客観)

長い 少ない 遅い 短い



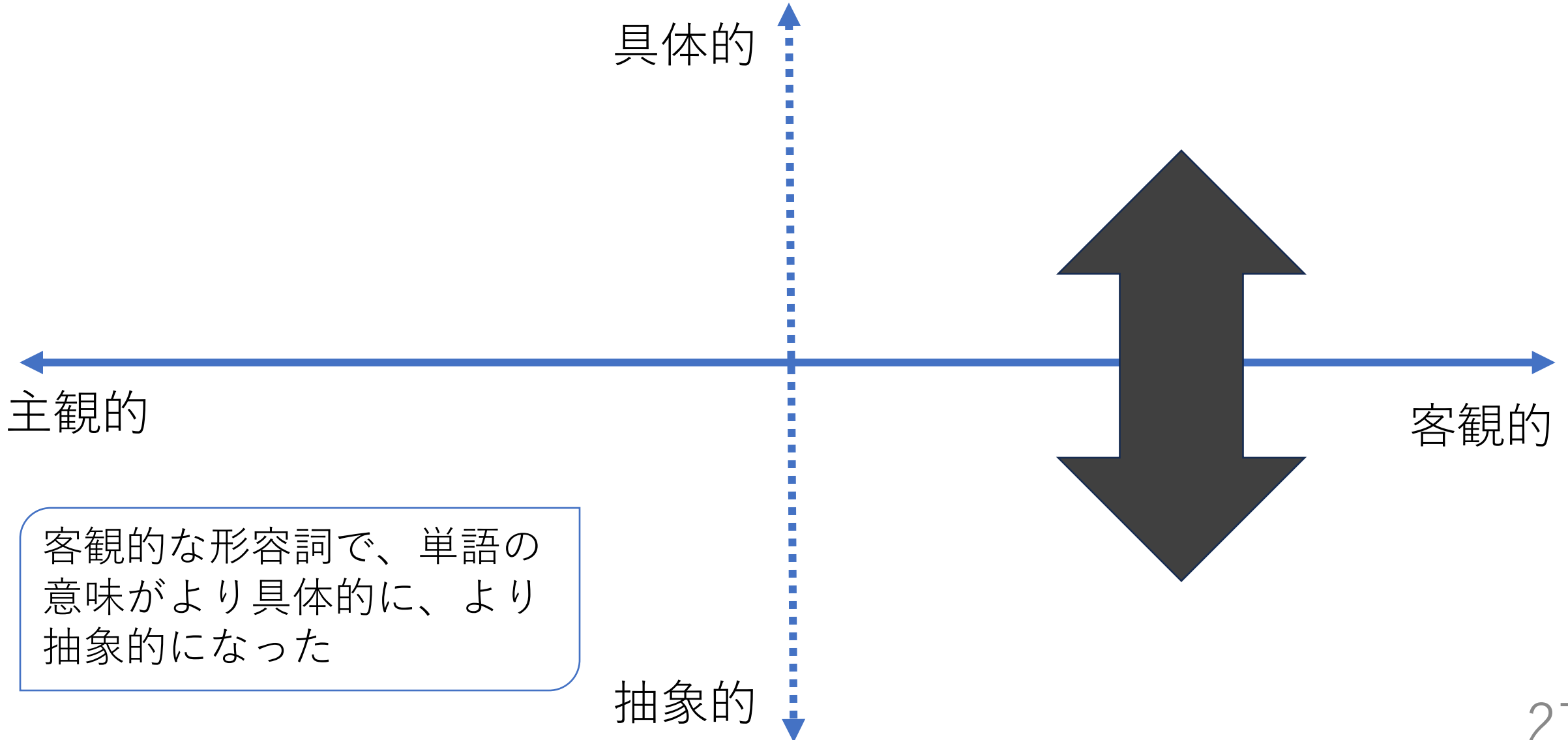


考察①: 形容詞の意味の質的変化傾向

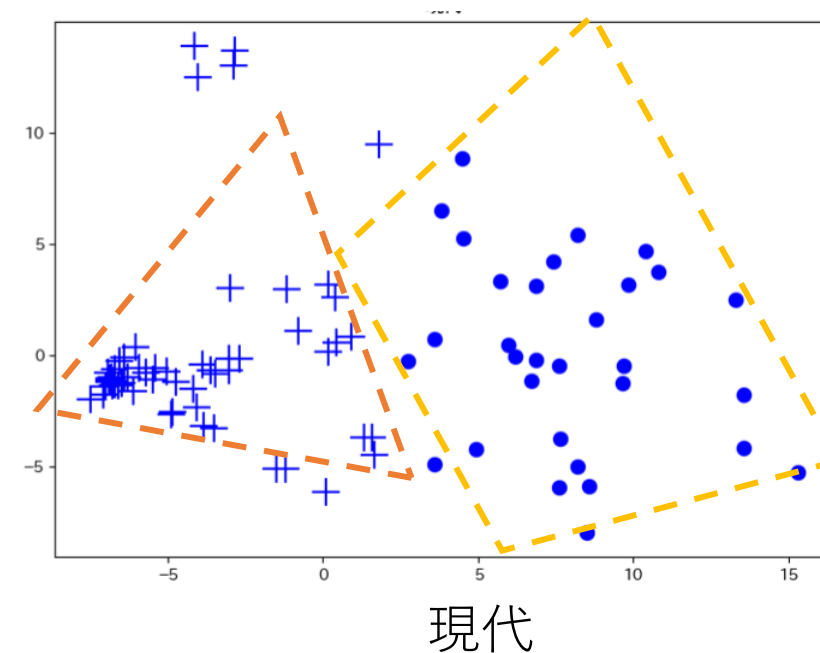
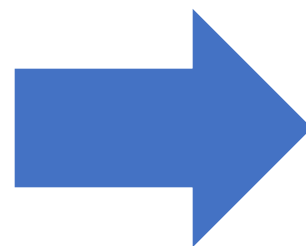
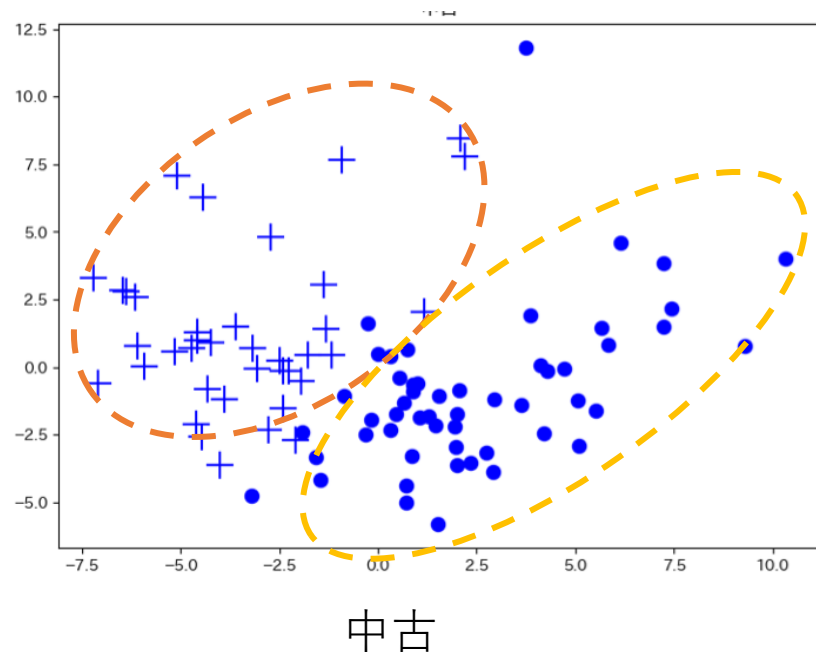




考察①: 形容詞の意味の質的変化傾向



考察②: 形容詞の意味の量的変化傾向



現代になるにつれ十字が集中している



単語の意味は
それが使われた**文脈**によって決まる

単語が似たような文脈で使われるようになった

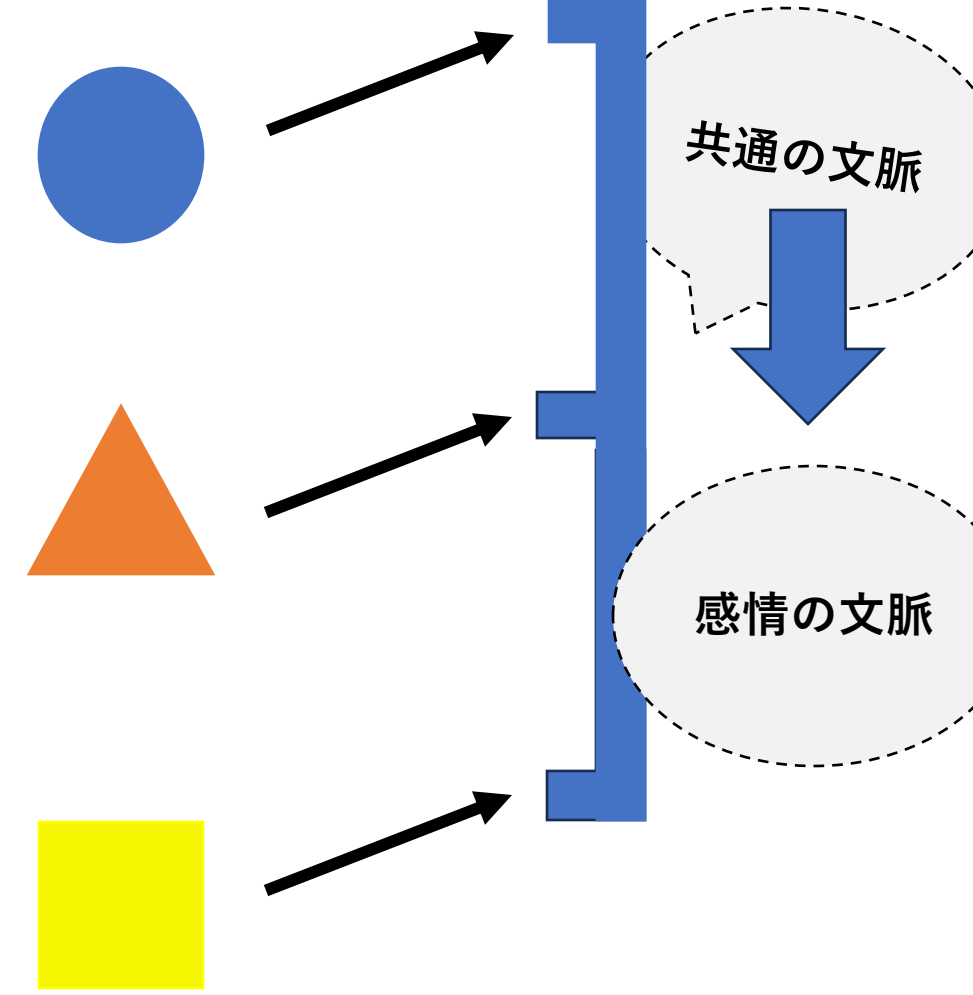
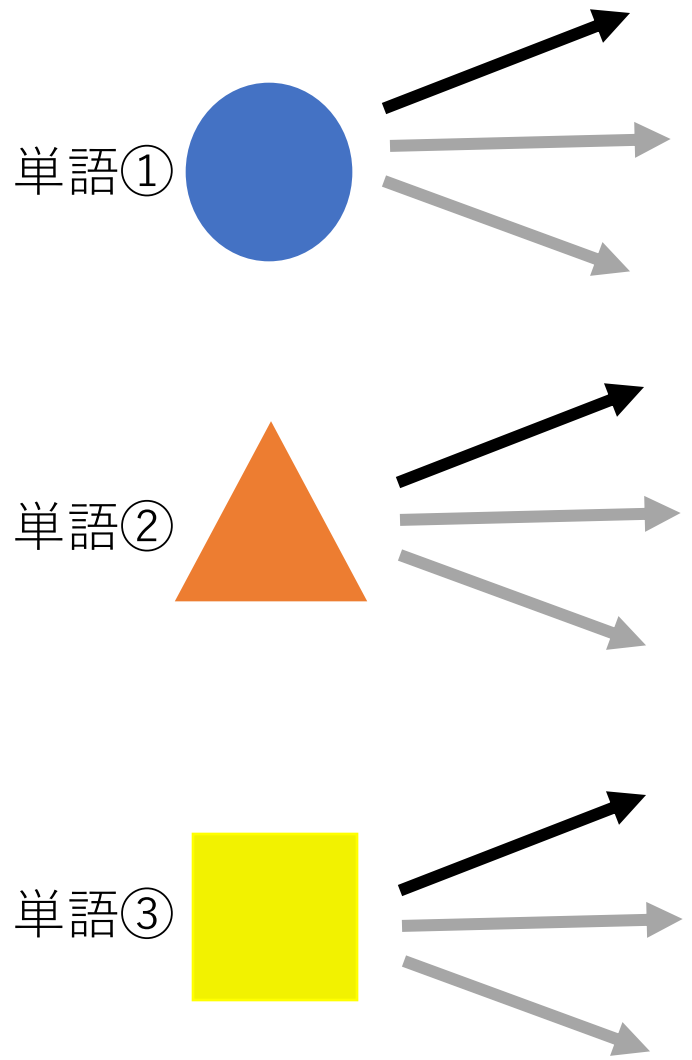
考察②: 形容詞の意味の量的変化傾向(主観的な形容詞)

昔…様々な文脈

多義的

今…ある一つの文脈に固まる

一義的





結論と今後の展望
Conclusions and Prospect



現時点での結論(仮説)

日本語形容詞の意味を考える上での2つのポイント

主観的 or 客観的

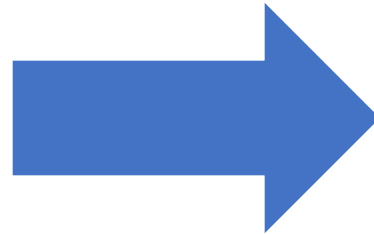
具体的 or 抽象的

現時点での結論(仮説)

●日本語形容詞の質的な意味変化の傾向

江戸時代頃を境に

客観的



主観的

●日本語形容詞の量的な意味変化の傾向

現代になるにつれて

主観的な形容詞は 多義性を失い
使われる文脈が似てきた



本研究の成果

- **日本語形容詞の意味変化**に関するいくつかの知見が得られた
- 言語の意味を**機械学習**によって解析する言語研究の可能性が改めて示された
- 本研究の手法は、通時的な変化だけでなく、**地域による差**などの研究にも応用可能



課題

- Word2vecの仕組み上、意味の多義性を反映できていない
- 自分たちで作成した単語ベクトルは、単語数が少ないので精度が低い
- 文献の性質が時代によって異なっている
- 10世紀以前の時代について調査できていない

今後の展望

- 様々な視点から結果を分析し、考察を深めていきたい
- 研究の精度をより上げていくとともに、古典学習等への活用の可能性を考えていきたい



参考文献

- 「Word2vecとは | モデルの種類やその仕組み、活用事例まで紹介！」.AINOW. <https://ainow.ai/2021/04/08/254071/>.(参照日2023年8月18日)
- 相田太一他「単語分散表現の結合学習による単語の意味の通時的変化の分析」(言語処理学会・26回年次大会・発表論文集・2020)
- 近藤康弘.「コラム:「意味の語形変化」をめぐって」.ことば研究館(国立国語研究所).<https://kotobaken.jp/digest/11/d-11-11/>.(参照日2023年8月18日)
- <https://ie.u-ryukyu.ac.jp/~tnal/2021/dm/static/4-nlp/nlp1.html#id9>
- 「自然言語処理における前処理の種類とその威力」.Qiita.
<https://qiita.com/Hironsan/items/2466fe0f344115aff177>.(参照日2023年8月18日)
- コーパス.goo辞書.
<https://dictionary.goo.ne.jp/word/%E3%82%B3%E3%83%BC%E3%83%91%E3%82%B9/#jn-75912>.(参照日2023年8月18日)

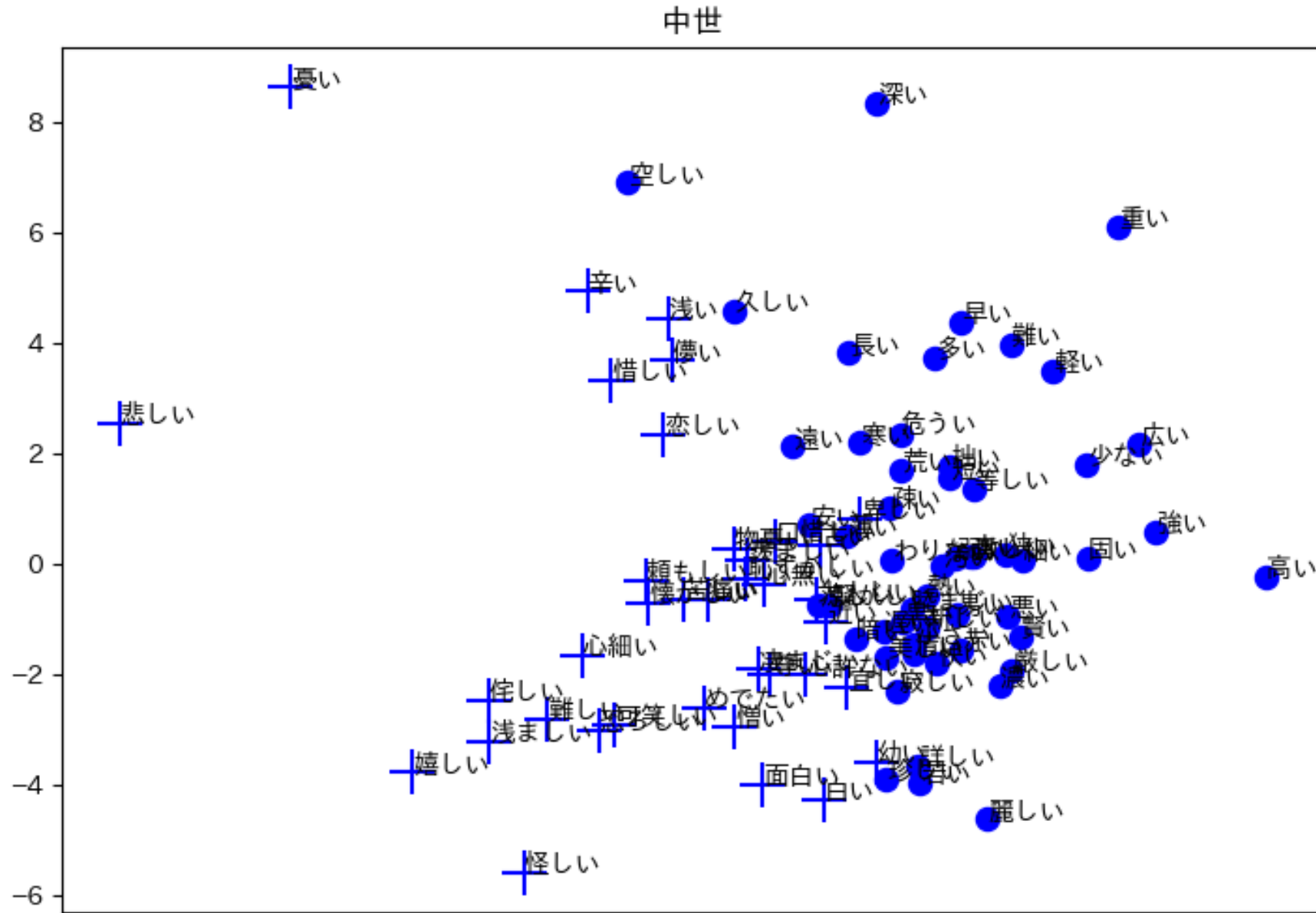


参考文献

- 鈴木正敏(2017)「日本語Wikipediaエンティティベクトル」
http://www.cl.ecei.tohoku.ac.jp/~m-suzuki/jawiki_vector/ (参照日2023年11月19日)
- 堤 智昭, 小木曾 智信 (2023) 「複数のUniDic辞書による形態素解析支援ツール『Web茶まめ』の実装と運用」, 情報処理学会, Vol.64, No.3, pp.749-757.<http://doi.org/10.20729/00225271>
- 小木曾 智信, 小町 守, 松本 裕治: 「歴史的日本語資料を対象とした形態素解析」, 自然言語処理, Vol.20, No.5, pp.727-748 (2013).
- Radim Řehůřek (2009) 「Gensim:Topic modelling for humans」
<https://radimrehurek.com/gensim/> (参照日2023年11月19日)
- 山本俊英「形容詞ク活用・シク活用の意味上の相違について」 (『国語学』23集・1955)

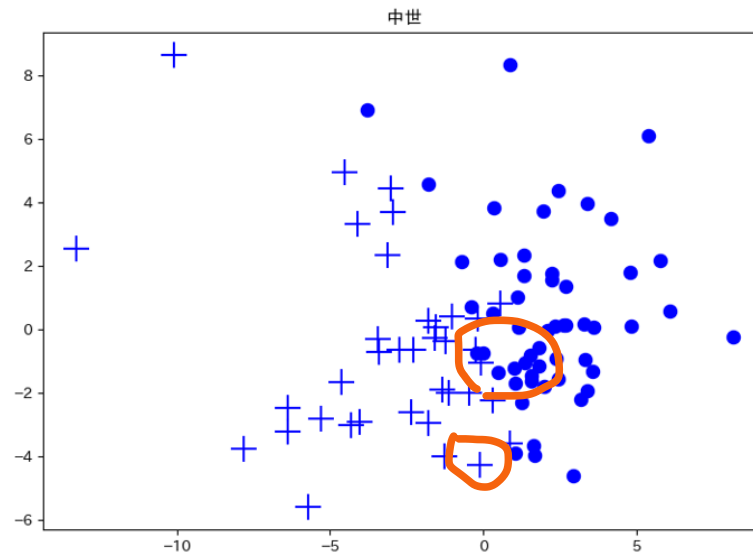
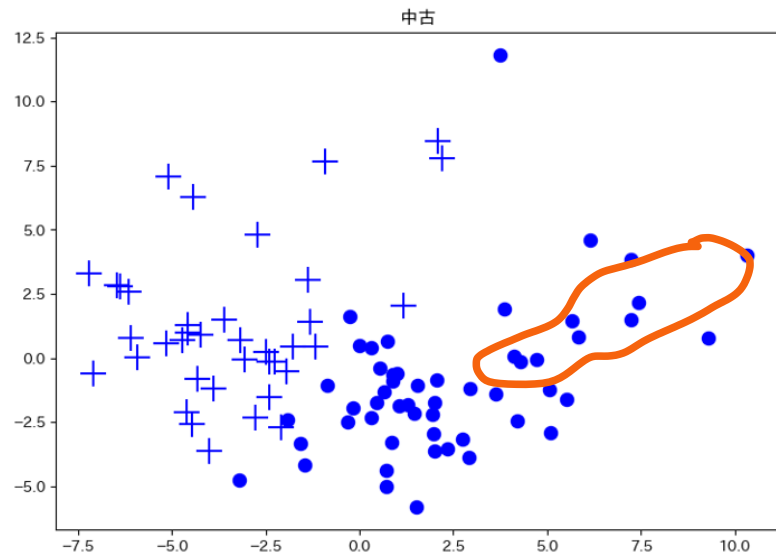


単語ラベルありのグラフ(中世)





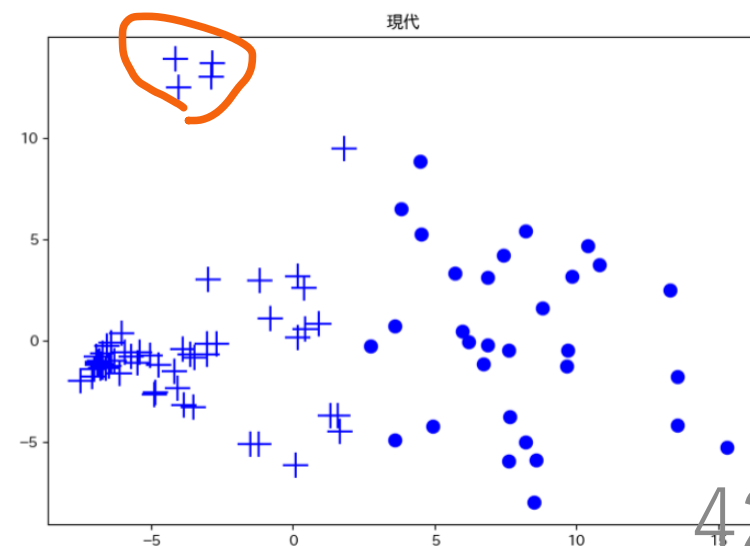
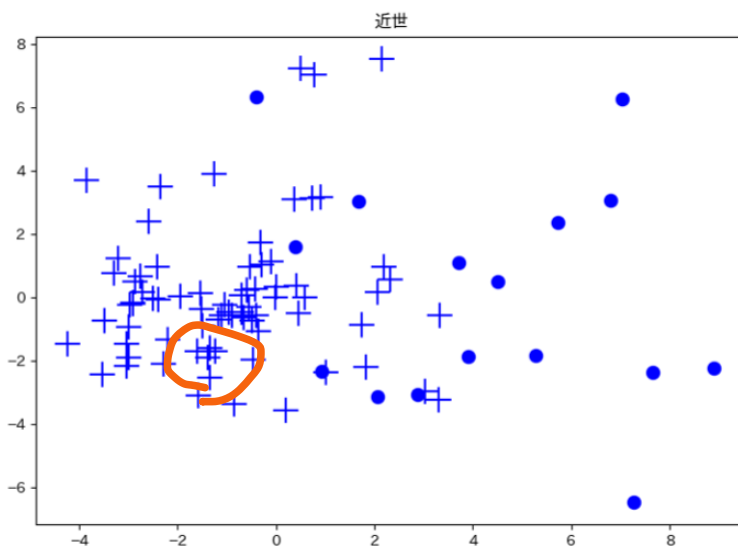
色を表す形容詞の変化



中古…客観的
貴族社会の物語

色が身分や季節、心情を示す

近世や現代…主観的
近世…町民が主役
現代…Wikipedia



色が何かを象徴する意味が
薄れ、感覚を示す意味が強
くなったのではないか