

歌ことば「橘」「梅」「桜」における関連対の抽出

ホドシチェック ボル
大阪大学
bor@lang.osaka-u.ac.jp

山元 啓史
東京工業大学
yamagen@ila.titech.ac.jp

材料

- 八代集データベース**
- ・新編国歌大観CD-ROM版の二十一代集に相当するデータ
 - ・国文学研究資料館編集二十一代集データベース (中村他 1999)
 - ・新日本古典文学大系本二十一代集に相当する書籍 (小山・新井 1989, 片瀬 1990, 小町谷 1990, 久保田・平田 1994, 川村他 1989, 片野・松野 1993)
 - ・新潮日本古典集成の新古今集 (久保田 1979)
 - ・ヴァージニア大学日本語テキストイニシアティブ監修
http://etext.lib.virginia.edu/japanese/
- 山元 (2007) で下記資料を収集
シソーラスに異なる表記を登録、正規化した
- (新編国歌大観編集委員会 1996)

方法

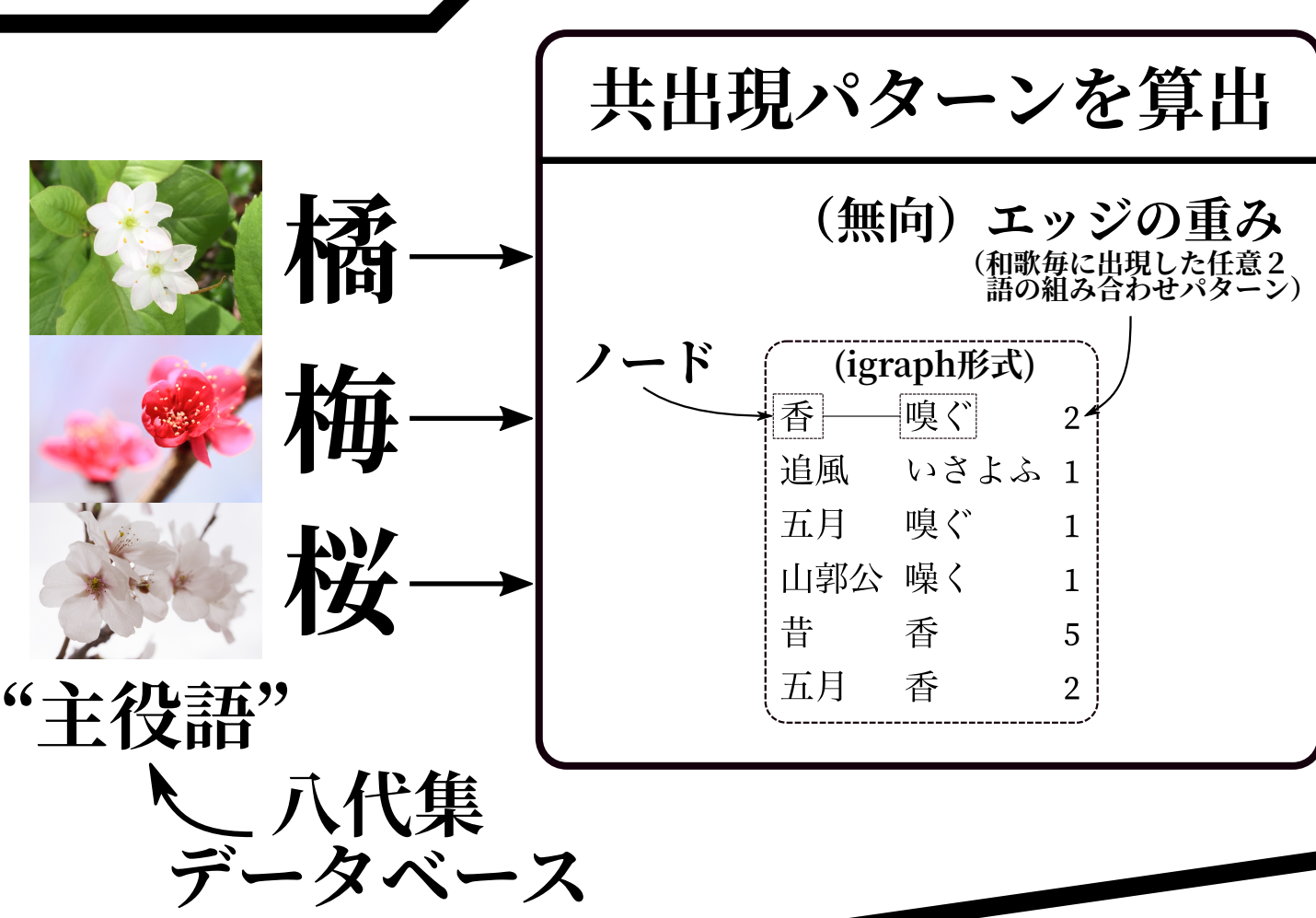


表1: 橘のサブクラスター: 1番から10番までを抽出。average法、mcquitty法、single法のクラスタリングを用いた。括弧内はそれぞれのMaximum partition densityを示す。

No.	average (.43)		mcquitty (.43)		single (.38)	
	node	edge	node	edge	node	edge
1	昔	7	昔	7	昔	5
2	匂ふ	6	匂ふ	6	匂ふ	4
3	風	5	今年	4	夢	4
4	夢	5	辺	4	香る	3
5	今年	4	待つ	4	今年	3
6	辺	4	風	4	初む	3
7	待つ	4	夢	4	五月雨	3
8	香る	3	初む	3	折	3
9	五月雨	3	香	3	枕	3
10	初む	3	闇	3	思ひ寝	3

表3: 「梅」のサブクラスターから抽出された「文無し」「花笠」「植う」に関わる歌

No.	歌番号	作者	歌
1.	10041	朝恒	春の上の/やみはあやなし/梅花/色こそ見えぬ/かやはかくる、
2.	20029	読人不知	心もて/おるかはあやな/梅花/かをとめてたに/とふ人のなき
3.	30016	朝恒	香をとめて/誰おらさむ/梅の花/あやなすみ/たちなかくしそ
4.	30031	能宣	匂ひをば/かせにそふとも/むめのはな/色さへあやな/あたにちらすな
5.	40051	能宣	梅花/にほふあたりの/夕暮は/あやなく人/にあやめられつ、

考察

関連対

直接的⇔間接的

- ・連想しやすい関連対よりも間接的な関連対が存在する
- ・和歌の短い文脈において主役語が同時に現れない
- ・関連対と類似対

可視化からみた問題点

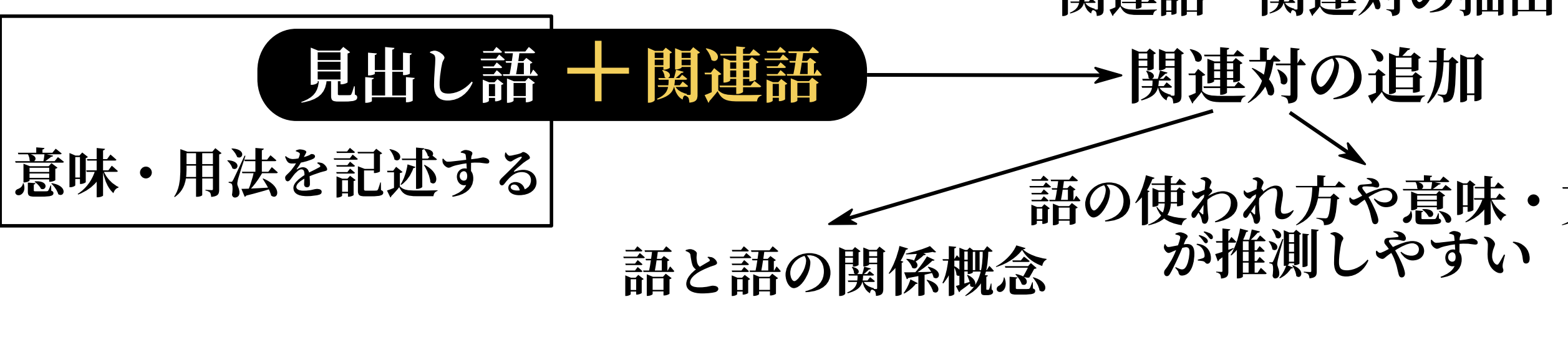
単語単位: 同じ文に使われた ≠ 連語が成立していた
重複したエッジ: 同一語形のため多義語が一義的に扱われてしまう

結論

歌ことば辞典の開発にあたり、従来の「見出し語とその解説」による記述に加え、「見出し語—関連語」の形式による記述を提案し、「橘」「梅」「桜」のコミュニティ分析によって、関連対の抽出を行い、それらは実際の和歌において関連した語であることが確認できた。

問題

過去のものとなってしまった古代語の意味記述
従来の見出し語とその解説に加え...



結果

