Development of the Dictionary of Poetic Japanese Description

Hilofumi Yamamoto
Tokyo Institute of Technology

Bor Hodošček
Osaka University

Introduction

- This paper proposes to further the development of a dictionary of classical Japanese poetry using pairwise term information (Yamamoto et al., 2014).
- Information on pairwise terms between an index and related term such as "flower–spring" is not included within traditional modern and classical Japanese dictionaries, even though this information connects terms with their contexts in a transparent way and thus offers an unbiased method for inferring the meaning of old Japanese terms.
- An R package for the analysis of linked communities in networks, linkcomm (Kalinka and Tomancak, 2011), is used to extract subordinate terms. Average, McQuitty, and single linkage methods are evaluated for the quality of their extraction of subordinate clauses of terms representing the 'cherry', 'plum', and 'orange' flowers. All methods extracted similar subordinate terms, which were quite natural in the context of classical Japanese poetry.

Problem

1. Many scholars of Japanese poetry have tried to explain poetic vocabulary based on their intuition and experience.
2. As scholars can only describe constructions that they can consciously point out, those that they are unconscious of will NEVER be uncovered. ⇒ In order to conduct more exact and unbiased descriptions:
   1) using computer-assisted descriptions;
   2) using co-occurrence weighting methods on corpora of Japanese poetry; and
   3) using linkcomm R package, extract the lists of words grouping sub communites.
   ⇒ allows one to BETTER GRASP the construction of poetic words.

a.


b.


c.


Methods

Calculation: ***Linkcomm*** (Kalinka and Tomancak, 2011) for sub communities of three flowers: *ume* (plum), *sakura* (cherry), and *tachibana* (mandarin orange).

Material: ***Hachidaishū (ca. 905–1205)*** from *Kokkataikan* (Shin-pen Kokkataikan Henshū Committee, 1996), Nijūichidaishū database published by NIJIL (Nakamura et al., 1999), Shin-Nihon Koten Bungaku Taikei (Kojima and Arai, 1989), and Shin-kokinshū (Kubota, 1979).

Result

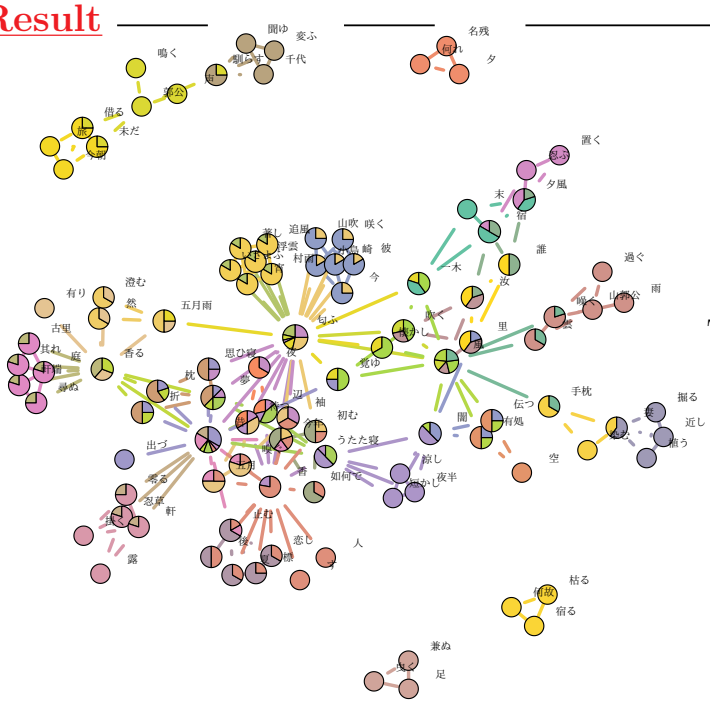

Fig. 1: Network of Words; mandarin orange.

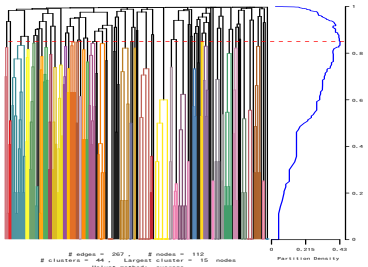

Fig. 2: Link Community Dendrogram.

Table 1: Sub-clusters of orange.

	average (.43)		mcquitty (.43)		single (.38)	
No.	node	edge	node	edge	node	edge
1	mukashi (old days)	7	mukashi	7	mukashi	5
2	nihofu (smell)	6	nihofu	6	nihofu	4
3	kaze (wind)	5	kotoshi	4	yume	4
4	yume (dream)	5	atari	4	kaoru	3
5	kotoshi (this year)	4	matsu	4	kotoshi	3
6	atari (aroud)	4	kaze	4	somu	3
7	matsu (to wait)	4	yume	4	samidare	3
8	kaoru (fragrance)	3	somu	3	ori	3
9	samidare (summer rain)	3	kaori	3	makura	3
10	somu (to dye)	3	yami	3	omohine	3

Conclusion

- Pairwise term information generated by the community centrality procedure works well.
- R package "linked communities" could extract proper sub cluster terms which contribute to the description of classical Japanese poetry.

Reference

- Kalinka, A. T. and Tomancak, P. 2011. linkcomm: an R package for the generation, visualization, and analysis of link communities in networks of arbitrary size and type. *Bioinformatics*. 2011–2. **27** (14).
- Yamamoto, H., Hajime Murai, Bor Hodošček. 2014. Development of an Asymptotic Word Correspondence System between Classical Japanese Poems and their Modern Translations. Symposium for Computer and Humanities, 2014. The information processing society of Japan. Vol. 2014, No. 3, 157–62.