Vol.2019-CH-120 No.9
2019/5/11

N-gram 統計を用いた近代短歌テキストの分析

村田祐菜†1

概要：近代短歌結社の一つである「アララギ」は,大正五年～昭和初期にかけて島木赤彦,斎藤茂吉,中村憲吉,古泉千樫,釈迢空らの同人を中心に,歌壇において大きな影響力を持った。彼らが歌壇において隆盛を極めた一因として結社意識の強さが指摘できるが,具体的な短歌表現に基づいて彼らの表現意識の共通点を明らかにした研究は少ない。本発表では短歌テキストに N-gram 統計を用いたテキスト分析の手法を用いて,同人間の短歌表現の分析を行った。

Text analysis of Japanese poems using N-gram statistics

YUNA MURATA†1

Abstract: “Araragi” is one of the most popular and authoritative literary group in Modern Japanese Poems from about 1916~1926. The main members of this group are Shimaki Akahiko , Saito Mokichi, Nakamura Kenkichi, Koizumi Chikashi. Shaku Cho-ku, and so on. One reason “Araragi” could become authority is their strong consciousness of association, but little research have explained about it in terms of the expression of their Tanka. In this report, I digitize their texts and compare with their expression by using N-gram statistics .

1. はじめに

近代短歌雑誌「アララギ」は正岡子規の周囲に集った根岸短歌会の歌人たちの雑誌「馬酔木」を源流とし,明治四十一年に創刊された。明治四十年代には伊藤左千夫,長塚節らを中心として活動していたが,歌壇の主潮流からは離れていた。

伊藤佐千夫が大正二年七月三十日に亡くなると,「アララギ」は同月,会員組織になるとともに大正四年二月からは島木赤彦が編集発行人となり結社内での指導力を増していった。また茂吉『赤光』,赤彦・憲吉『馬鈴薯の花』,憲吉『林泉集』等が刊行されると次第に読者を増やし,歌壇の中心となっていくが[1],その歌風はしばしば万葉尊重,写生主義という言葉で表される[2][3].大正中期以降の中心的な同人としては,島木赤彦,斎藤茂吉,中村憲吉,古泉千樫,釈迢空が挙げられるが,彼等は正岡子規の説いた「写生」の理論を拡充し,「実相に観入して,自然・自己一元の生を写す」という茂吉の「短歌における写生の説」や,赤彦「鍛錬道」などの写生論を共有しその理論は歌壇に大きな影響力を与えた。

反アララギ勢力の超結社誌「日光」の創刊（大正十三年四月）や,指導者赤彦の死（大正十五年三月）などにより,昭和期に入ると勢力にやや陰りが見えるようになるが,昭和三十年に編集を受け継いだ土屋文明を中心として,戦後も平成九年まで歌誌の発行は続いた。

「近代短歌とはまさしく諸結社を中心とする短歌運動によって進展し,深められていったのである。」[4]というように,結社という単位で近代短歌の運動は進められてきたと言える。「アララギ」はともすれば写生一辺倒に陥りかねない閉鎖性も批判の対象とされてきたが,別の視点から見ればそれは同人たちの強い結社意識に裏打ちされたものであり,大正期における「アララギ」の「歌壇制覇」はその強い連体意識によって成し遂げられたものだと言える。これまでの研究において彼等の結社意識は主に写生論の共有という理論面への注目が中心であり,具体的な短歌表現における影響関係の指摘は一部分に留まっている。本稿は大正期から昭和初期にかけて刊行された「アララギ」同人の歌集を対象にし,N-gram 統計を用いたテキスト分析によって短歌表現に現れる具体的表現に基づいた数量的な分析の可能性を提示したい.

2. 先行研究

古典文学作品のテキストに N-gram 統計処理を用いた文字列分析を行ったものとして,近藤みゆき氏の研究が挙げられる[5][6].氏は N-gram による文字列総比較によって,初期定数歌における「返し」の手法や,作者推定,年代推定,ジェンダー論の視点から和歌や『源氏物語』におけることばの性差の指摘などの成果をあげられており,本稿も氏の研究に大いに示唆を受けた。

N-gram 分析を用いる利点として,形態素解析では捉えきれない,複数の品詞が一体となった長い表現の連なりを取り出せるという部分が大きい。近藤氏は具体的な文章表現

†1 東京大学人文社会系研究科
Graduate school of Humanities and Sociology , The University of Tokyo

ⓒ 2019 Information Processing Society of Japan

に現れる文字列を「ことばの型」と表現している[7]が,近代短歌においても,単語の具体的なあらわれ方において各歌人の短歌表現の類似性や独自性を明らかにできると考える。

3. 近代短歌テキストのデジタル化と N-gram 分析

近代文学の電子化テキストが入手できる場として青空文庫[8]が挙げられるが,近代短歌テキストは北原白秋,斎藤茂吉ら著名な歌人のみに限られているため,今回は各人の全集を OCR 処理をし,人手による修正という作業を行った。今回分析対象として扱った歌集は以下の通りである。

- 斎藤茂吉

M1『赤光』a (大正 2 年：明治 38 年～大正 2 年)
M2『あたらま』(大正 10 年：大正 2 年～大正 6 年)
M3『つゆじも』(大正 11 年：大正 6 年～大正 11 年)
M4『遠遊』((昭和 22 年：大正 11 年～12 年)
M5『遍歴』(昭和 23 年：大正 12 年～大正 14 年)

- 島木赤彦

A1『馬鈴薯の花』b(大正 2 年：明治 42 年～大正 2 年)
A2『切火』(大正 4 年：大正 3 年)
A3『氷魚』(大正 9 年：大正 4 年～大正 9 年)
A4『太虚集』(大正 13 年：大正 9 年～大正 13 年)
A5『柿蔭集』(大正 15 年：大正 13 年～大正 15 年)

- 中村憲吉

K1『馬鈴薯の花』(大正 2 年：明治 41 年～大正 2 年)
K2『林泉集』(大正 5 年：大正 2 年～大正 5 年)
K3『しがらみ』(大正 13 年：大正 6 年～大正 10 年)
K4『軽雷集』(昭和 6 年：大正 10 年～昭和 3 年)

- 古泉千樫

C1『川のほとり』(大正 14 年：明治 37 年～大正 13 年)
C2『屋上の土』(昭和 3 年:明治 43 年～大正 6 年)
C3『青牛集』c (昭和 8:大正 7 年～昭和 2 年)

- 釈迢空

S1『海やまのあひだ』(大正 14 年：明治 37 年～大正 14 年)

※括弧内の年号は(刊行年：収録歌の製作年代)である。また,歌集の前には各歌人の頭文字と数字を刊行年代順に付した。

分析用テキストは目次,題,詞書,跋文,ルビを削除した短歌本文テキストのみを作成し,一首ごとに改行しているものを使用した。旧字,歴史的仮名遣いはそのまま残した。作成したテキストの N-gram 分析においては morogram[9]を Windows の実行形式に変換したものを使用[10]し,複数のテキストの処理結果の統合には ngmerge[11]を用いた。

4. 分析結果

4.1 複合語「にけるかも」

統計処理の結果,全歌人に渡っての使用がみられた複合語のひとつとして「にけるかも」を取り上げる。各歌集における出現頻度は以下の通りである。

歌集	出現頻度
M1　赤光	12
M2　あらたま	22
M3　つゆじも	8
M4　遠遊	0
M5　遍歴	0
A1　馬鈴薯の花（赤）	0
A2　切火	13
A3　氷魚	6
A4　太虚集	4
A5　柿蔭集	2
K1　馬鈴薯の花（憲）	0
K2　林泉集	11
K3　しがらみ	6
K4　軽雷集	3
C1　川のほとり	3
C2　屋上の土	2
C3　青牛集	5
S1　海やまのあひだ	11

表 1「にけるかも」の出現頻度

用いたテキスト中最も早い用例は明治三十八年作,斎藤茂吉

黒き實の圓らつぶらとひかる實の柿は一本たちにけるかも
けふの日は母の邊にゐてくろぐろと熟める桑の實食みにけるかも
數學のつもりになりて考へしに五目並べに勝ちにけるかも
（『赤光』「折にふれて」）

a 『赤光』は大正 10 年に茂吉による改訂削除を経た改選版,大正 14 年に改選『赤光』第三版が出版されているが,今回は初版のテキストを採用した。
b 赤彦,憲吉の共著。
c 『日本現代文學全集 52』(講談社,1965 年)を底本としたため,「青牛集抄」となった。

ⓒ 2019 Information Processing Society of Japan

の三首であり,この歌を含め『赤光』では 12 首の用例がみられる。一方,『赤光』の三か月前に刊行された『馬鈴薯の花』には全く用例が見られず,次の歌集『切火』及び『林泉集』において急激に数が増加している。他の歌人の用例も『赤光』刊行（大正 2 年 10 月）以前の用例は見られないため,この複合表現の広がりは『赤光』の影響による可能性が高いと考えられる。

大正後期になると『遠遊』『遍歴』では全く使用されず,他の歌人も以降減少傾向にあることから,『赤光』刊行以後の大正二,三年から大正十年前後において盛んに「アララギ」同人の中で使用されていた表現であると言える。

和歌の用例は万葉集に十八例見られるが,それ以降は勅撰集には全く見えない表現であり万葉集からの表現の摂取である可能性が高い。以下に数例をあげるd。

早来ても見てましものを山背の高の槻群ちりにけるかも（巻三,277）
いはばしる垂氷の上のさわらびの萌えいづる春になりにけるかも（巻八,1418）
萩の花咲けるを見れば君に逢はずまことも久になりにけるかも（巻十,2280）
秋萩の枝もとををに露霜置き寒くも時はなりにけるかも（巻十,2131）

一方で,この複合表現の使用ををすぐにアララギ同人たちのみに限定することには留保が必要である。例えば,『赤光』刊行以前の歌集である北原白秋『桐の花』（大正 2 年 1 月）にこの表現は全く見られないが,『雲母集』（大正 4 年）には十三例現れてくる。

大空に何も無ければ入道雲むくりむくりと湧きにけるかも
幅びろの光なだるるなだら坂動くばかりに見えにけるかも
（北原白秋『雲母集』）

これらの例を考えると,「アララギ」同人だけには限らない歌人にも広がっていた表現であると言え,さらに広い範囲での調査が必要である。

ここで,斎藤茂吉に「一種の『かも』」（「アララギ」,大正 4 年 9 月）という文章がある。

『停車場に旅びととして我が姿ほこりかにして立てりけるかも』（小泉千樫）此歌は大正元年十月発行「アララギ」所載のものであつて,結句の『立てりけるかも』といふ用法が特殊のものである。かかる用法はおほかた古泉千樫に始まつた様である。次いで赤彦も憲吉もかかる用法を為し,引いて一般歌壇のなかにも往々斯る用法が見当るやうになつた。

と,「たてりけるかも」というの表現の端緒を小泉千樫であると表明している。「りけるかも」という形での出現を以下の表に示す。

歌集	出現頻度
M1　赤光	8
M2　あらたま	6
M3　つゆじも	0
M4　遠遊	0
M5　遍歴	0
A1　馬鈴薯の花（赤）	0
A2　切火	5
A3　氷魚	0
A4　太虚集	0
A5　柿蔭集	0
K1　馬鈴薯の花（憲）	0
K2　林泉集	5
K3　しがらみ	0
K4　軽雷集	0
C1　川のほとり	0
C2　屋上の土	0
C3　青牛集	0
S1　海やまのあひだ	0

表 2「りけるかも」の出現頻度

「にけるかも」に比して,初期の歌集の数例にかぎられる。この二例の複合表現の比較を通して,近代短歌の中に「けるかも」という万葉的な表現を取り入れる過程で,短歌における出現形の取捨選択の様相が浮かびあがってくる。

これは茂吉の証言から遡って用例を調査する従来の研究方法では明らかにできない部分であり,N-gram 統計を用いる意義は大きいと言える。

4.2 独自表現

各歌人の表現を対照してゆくと,ある歌人にのみ多くみられる表現が抽出できる。特に特徴的な茂吉の例を挙げる。

あらはれ,あまつ日,いつしか,かがよふ,かたはら,かへりみ,くろぐろ,しんしん,はるばると,ひむがし,むらがり

M1～M5 の歌集に連続して出現する一方,他の歌人の歌集には数例,もしくは全く見られない表現であることから,

d 別傾向の用例としては,「秋の野の尾花が末の生ひなびき心は妹に寄りにけるかも」（巻十,2239）の下線部に類似した表現が 9 例見られる。

茂吉の短歌において特徴的な語であると言えるe。「かがよふ」を例にとると,

しろがねのかがよふ雪に見入りつつ何を求めむとする心ぞも（『赤光』）
まかがよふ淺茅が原のふかき晝むかうの土に豚はねむりぬ（『あらたま』）

など,計十五例見られ,他の歌人の用例は無いf。この単語は

灯火のかげにかがよふうつせみの妹が笑まひし面影に見ゆ（『万葉集』巻十一,2642）
見渡せば近きものから石隠りかがよふ珠を取らずは止まじ（巻六,951）

などにみられ,茂吉独自の万葉集からの表現摂取の様相がうかがわれる。またこれらの独自表現をそれぞれ取り上げ,各首の表現や古典の用例との比較を通してつぶさに検討していくことによって,各歌人の表現の独自性を明らかにできると考える。

4.3 表記の差異—「さびし」「さみし」「寂し」

	寂し	さびし	さみし
M1　赤光	15	10	2
M2　あらたま	26	32	0
M3　つゆじも	28	6	0
M4　遠遊	10	11	0
M5　遍歴	15	11	0
A1 馬鈴薯の花（赤）	10	10	0
A2　切火	11	3	0
A3　氷魚	32	8	0
A4　太虚集	28	0	0
A5　柿蔭集	11	0	0
K1　馬鈴薯の花（憲）	12	6	2
K2　林泉集	20	2	5
K3　しがらみ	17	0	19
K4　軽雷集	14	0	7
C1　川のほとり	0	0	0
C2　屋上の土	6	4	0
C3　青牛集	5	3	0
S1　海やまのあひだ	7	71	0

表 3「寂し」「さびし」「さみし」の出現頻度

の表現研究において大きな成果を挙げてきたが,近代短歌

「寂し」「さびし」「さみし」という単語をとりあげると,「さびし」と「さみし」において歌人の使用傾向が明確である。「さびし」は『海やまのあひだ』での使用が最も多く71 例で,迢空において「寂し」「さみし」の例は少ない。二番目に使用が多いのが茂吉であり,『あらたま』では著しく増加している。赤彦も大正前期に「さびし」を多用しているため,「寂し」「さみし」に比して「さびし」は特殊な位相を持った語であると考えられる。茂吉は「この『さびし』の語は,人間本来のある切実な心の状態うぃあらはすのに適当な語であるから,縦ひ色調上の細かい変化はあつても,根本に於ては依然として『伝統』を続けることが出来るであろう」[11]と,『万葉集』や『古今和歌集』に連なる,日本人の伝統的な感情表現を見ている。

一方,憲吉においては,『馬鈴薯の花』で使用していた「さびし」が見えなくなると同時に『しがらみ』以降,「さみし」が多用されている。この差異は複数の歌集に渡って「かなし」という表現の優位性が見られるため,憲吉の意識的な使用と判断できる.「さびし」と「さみし」は表記上の差異だけでは捉えきれない表現上の差異があると考えられる。

5. N-gram 統計を用いることの意義と今後の課題

近代短歌における表現は表記の問題も含め,全て作家個人の表現意識に基づくものだと言える.N-gram 統計によって単語単位にとどまらない,短歌表現における単語のあらわれかたを拾い,比較検討していくことによって,結社の同人間における表現の共有や,各歌人の表現の独自性を明らかにすることができるものである。一方,今回取り上げたものはごく一部の表現についてであり,膨大なデータの中から有用なものを取り出すためには常にテキストの表現に帰ってひとつひとつ検討していく作業が重要である。

また,形態素解析の手法と合わせた単語単位での分析や Word2vec など,N-gram 統計以外の分析手法の有効性を比較検討していくことが必要である。

6. おわりに

近代短歌テキストは従来デジタル化されている資料が少なく,小説や戯曲といった他の近代文学のジャンルと比較して,テキスト分析の対象となることはほとんど無かった。これは視点を変えるならば,デジタル化とテキストの検索,比較,分析によって可能になる新たな研究の余地が多く残されている分野であると言える。新編国歌大観 CD-ROM によってテキスト検索が可能になったことは古典和歌テキストも検索可能なテキストを整備していくことで,『万

f 「アララギ」以外では北原白秋『雀の卵』（大正 10 年）に 2 例見られる。

葉集』や『古今和歌集』をはじめとした古典との表現の連続において捉えることが可能になるのではないかと考える。

今後は他の近代短歌テキストのデジタル化を進めるとともに,「アララギ」同人にとどまらない広い範囲での表現を対象とすることで「アララギ」という結社単位での独自性や,結社内外における影響関係を考えていきたい。

謝辞

東京大学大学院人文情報学研究科人文情報学拠点における人文情報学の授業において,下田正弘先生をはじめとする諸先生方にご指導いただいたくとともに,一連の作業において永崎研宣先生にご助力をいただいたことを感謝とともに記しておく。

参考文献

[1]斎藤茂吉:『続明治大正短歌史』,中央公論社,P103(1951).
[2]本林勝夫氏筆:『日本近代文学大辞典第五巻』「アララギ」の項,講談社,(1977 年).
[3]永田和宏氏筆:『岩波現代短歌事典』「アララギ」の項,岩波書店(1999).
[4]藤岡武雄:近代歌壇の形成,和歌文学講座第九巻,勉誠社(1994).
[5]近藤みゆき:『古代後期和歌文学の研究』,風間書房(2005).
[6]近藤みゆき:『王朝和歌研究の方法』,笠間書院(2015)
[7]近藤みゆき:n グラム統計処理を用いた文字列分析による日本古典文学の研究―『古今和歌集』の「ことば」の型と性差」,千葉大学人文研究,vol29(2003).
[8]“青空文庫”https://www.aozora.gr.jp/,(参照 2019-04-11).
[9]“morogram: Unicode 対応 N グラムツール”http://morogram.osdn.jp/.師茂樹氏作成(参照 2019-04-11）.<入手先>.https://ja.osdn.net/projects/morogram/releases/p1796.(参照 2019-04-11）.
[10]<入手先>http://japanese.gr.jp/tools/ngmerge/.(参照 2019-04-11).
[11]斎藤茂吉:「さびし」の伝統,『斎藤茂吉全集第十四巻』,岩波書店(1952).

ⓒ 2019 Information Processing Society of Japan