

Ⅱ 和歌の文学的特徴分析

——Dice係数による類似和歌検索方法の導入を合わせて——

武 石 悠 霞、金 中

要 旨

ディープラーニングが進歩に伴い、人工智能(DI)による詩歌の創作研究が盛んに行われるようになった。著者の研究チームは最近深層学習による和歌生成モデル「和歌_{DI}」を提案した。「和歌_{DI}」は条件付き変分オートエンコーダーと自己注意メカニズムを組み合わせ、指定されたキーワードによって質の高い和歌を生成することができる。本論文ではDice係数による類似和歌の検索方法を提案し、「和歌_{DI}」モデルによって生成された3首の_{DI}和歌を例に挙げ、人間和歌との比較の観点から分析を行った。Dice係数を計算するこの検索方法で類似和歌を客観的に順位付けし、膨大な和歌データから最も目標和歌と近い和歌を選び出すことができる。この和歌検索方法は研究者が和歌を検索する上での参考の一つとして提示することが出来ればと考えている。同方法によって、和歌生成モデル「和歌_{DI}」が創作した_{DI}和歌と類似する人間和歌を選び出し、その文法や語彙の組み合わせを比較し、_{DI}和歌の特徴と獨創性を分析した。分析結果により、「和歌_{DI}」モデルは人間和歌の主題、文法、句の構造、及び用語を正確にマスターし、獨創的な単語の組み合わせ方により斬新的な和歌を生成できることが認められた。

一、人工知能による和歌生成の概述

近年、人工知能(AI)による詩歌の生成はコンピューターサイエンスの分野で注目を集め、ディープラーニングによって詩歌を生成するシステムにおいて大きな成果を上げている。先行研究ではリカレントニューラルネットワーク(RNN)と敵対的生成ネットワーク(GAN)の言語モデルを用いて、俳句の生成方法¹⁾が研究されてきた。俳句に比べ、和歌を生成することはより難易度が高く、関連する研究も少ない。要因としては、和歌は俳句よりも時代が古く、使われている言葉が古語に近く、ディープラーニングに用いることのできるデータが少ないためである。また、和歌の多くのデータは濁点がついておらず、データのプレ処理で濁点を付けなければならない。更に、俳句が十七音しかないのに対し、和歌は十七音の上の句と、十四音の下の句によって成り立ち、上の句と下の句の間で関連性を持つ必要がある、かつその主題も季節、生活、恋など多岐に渡る。Masadaなどの研究では近代作品を主とする和歌を主題とし、 α -生成序列に基づいて和歌を評価する方法を提案した。²⁾この方法ではLDA(Latent Dirichlet Allocation)によって生成した和歌を評価し、純粋な α -方法に比べ生成した和歌の質は高まったが、この方法では和歌データに対して濁点付けなどのプレ処理は行っていない。中辻真などはLSTM時間列のディープラーニング方法で開発した α -短歌生成器「恋する α -歌人」³⁾は、人工により第一句を入力し、第一句に基づいて第二から第五句までが順番に生成される仕組みである。使用されているデータは近代女性歌人、与謝野晶子、岡本かの子、柳原白蓮、九条武子、原阿佐緒ら五者の作品である。故にこの生成器で生成した作品は近現代の短歌であり、古典の和歌とは異なる。和歌の自動生成方法の研究によって、様々な言語や文体における自然言語生成システムの潜在的な創作力を発掘し、文学の創作に役立つことを期待したい。

以下においてまず本論文に関連する我々西安交通大学の詩歌生成研究チームの前期研究の結果⁽⁴⁾、和歌生成モデル、和歌データの処理、和歌及びその人工評価について紹介したい。

1-1、和歌生成モデル「和歌A」

ディープランニングによって和歌の生成モデルを構築するには、和歌の形式、言語の正確さ、獨創性、多様性などの問題をクリアする必要がある。形式においては、五七五七七の音数を満たし、かつ上の句と下の句の間の関連性が不可欠である。言語の正確さにおいては、文法的に正しく、かつ意味が通る句を創り出さなければならない。獨創性においては、独自の和歌を生成し、学習データと同じ系列を生成することを防ぎ、和歌データからの盗作にならないようにする。多様性においては、異なるキーワードを入力した際には生成される和歌内容も異なるべきであり、同じ内容の重複を避けるべきである。

以上の点をクリアするために、我々はシーケンシャル変分トランスフォーマーの和歌生成モデル「和歌A」を提案した。「和歌A」は条件付き変分オートエンコーダーと自己注意メカニズムを組み合わせ、指定されたキーワードによって質の高い和歌を生成することができる。和歌生成モデルの性能向上に大きく貢献したものは次の通りである。

和歌の獨創性と多様性を図るため、隠れた変数シーケンスによって、和歌データの単語の変動性に対してモデリングする。更に、和歌に使われる言語のマルチラレベルな特徴を表すため、マルチラレベル融合自己注意メカニズムを用いて、生成された和歌が優れた流暢さ、関連性、および意味を備えるようにする。我々は、このを平安時代の貴族歌人をイメージして「中太郎⁽⁵⁾」と名づけた。

一―二、和歌データのプレ処理とデータベースの構築

我々は大型の和歌データベース^⑥を構築し、八世紀から十六世紀までの和歌201,528首が収録され、『万葉集』二十一代集及び多くの私家集や歌合などの作品が含まれている。これらのデータから重複和歌や、少数のアルファベット表記の和歌を取り除いた。和歌データは歴史的仮名遣いの形式によって表記されていて、濁音の文字には濁点をつけておらず、一句の中では単語分けもされていないため、コンピューターは語句を認識できず、ディープラーニングの学習データとしてはそのまま使用できない状態であった。このために、我々は二万首の濁点付き和歌データを用いて、条件付き確率場に基づくシーケンスラベリングモデル^⑦をトレーニングし、トレーニングされたこのモデルを用いて日文研データベースのすべての和歌データに濁点を付け、更に日本国立国語研究所の形態素解析ツールウェブ茶^⑧まめを用いて日文研和歌データの単語分け処理を行い、171801首の濁点付きで単語分けされた和歌データベースを構築した。

一―三、AI和歌とその評価

モデルのトレーニング段階では、プレ処理された和歌データベースを使って和歌生成モデルをトレーニングする。和歌生成の段階では、トレーニングされたモデルにキーワードを入力し、キーワードを含む和歌が生成される。「和歌VLM」モデルによって生成された和歌を評価するため、我々は隠れた変数シーケンスRNN VAE^⑨、Transformer 言語モデル、及び単一の隠れた変数のTransformer VAEを用いて、六つのベースラインモデルを構築した。キーワードを「和

歌Ⅴ」^⑧と他の六つのベースラインモデルに入力し、各モデル百首、合計七百首の和歌を生成した（生成日時二〇二〇年十二月二十八日）。コンピュータによる評価と、和歌研究者による人工評価により、ベースラインモデルと比較して、隠れた変数シーケンスを搭載した「和歌Ⅵ」^⑨は、人間和歌の主題、文法、語句の意味関係を正確に学習し、より豊かな語彙、多様性、独創性を備える和歌を生成できることが確認された。

一四、「和歌Ⅶ」モデルによって生成されたAI和歌の例

本論文では「和歌Ⅴ」「モデルによって生成された三首のⅡ和歌を例に挙げ、人間和歌との比較の観点から分析を行う。我々はDice係数による類似和歌の検索方法を提案し、Ⅱ和歌と人間和歌の文法や語彙の組み合わせを比較し、Ⅱ和歌の特徴と独創性を分析することで、Ⅱによって生成された和歌のオリジナル性を評価することを試みる。Dice係数による類似和歌の検索方法により「和歌Ⅶ」モデルが生成したⅡ和歌と類似する人間和歌を探し出し、それらの和歌を類似度係数に従って客観的に順番付けすることが可能となった。

下記の三首の和歌は「和歌Ⅶ」モデルによって生成されたⅡ和歌である。

AI和歌① 明けてゆく 峰の木の葉の 梢より 遙かに続く 小牡鹿の声

AI和歌② 山里の 松の梢に 埋つもれて 霞を分けぬ 鶯の声

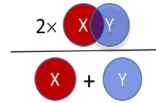
AI和歌③ 冬来ぬと 思ひ明かしの 柴の庵 枕に残る 霜の音

11、Dice係数による類似和歌の検索方法

和歌の研究においては、研究対象の和歌と類似する和歌を検索する必要がある。これは、類似和歌と比較し、研究対象の和歌の内容、語句、出典及び作者の創作背景などを考察するためである。和歌の研究では、類似する和歌を探し当てることで、和歌と人間の和歌の語彙の組み合わせの特徴を確認し、和歌が人間の和歌から学んだこと及びその独自性と創造性について論じること、和歌生成モデルを改善できるだけでなく、詩歌創作に対し参考となる情報を提示することができる。一般的な和歌の研究において、データベースから和歌を見ていく際に使われるのは『新編国歌大観』CD-ROMや国際日本文化研究センターの和歌検索サイトである。類似する和歌を検索するには、研究対象の和歌の中からいくつかのキーワードを選び出して入力するが、その際に、しばしば膨大な数の和歌がヒットする。これらの和歌から最も類似する和歌を捜し出すには多くの時間がかかる。本論文においてはDice係数⁽¹⁰⁾によって類似和歌を検索する方法を提案する。Dice係数を用いた類似和歌の検索方法は、目標和歌と類似する和歌を迅速に探し出し、かつ類似度係数に基づいて類似和歌を順番付けすることができる。必ずしも係数の高い和歌が一番近いとは限らないものの、数値は客観的であり今後の研究に活用されることが期待される。

Dice係数は二つのサンプルの類似度を計算するのに用いられるが、統計学において広く用いられる方法である。情報検索では、集合Xと集合Yが与えられると、類似度は、【図1】で表示したとおり、共通情報の二倍（つまり、XとYの重複部分）を集合XとYの合計で割ったものとして定義される。

【図1】 集合Xと集合Yにより定義される類似度



γ 和歌と人間和歌の類似度をDice係数によって計算する際、文字列の類似度測定式⁽¹¹⁾に従って、二つの和歌の類似度係数Dは次のように記述できる。

$$D = 2 \times n_i / (n_i + n_j) \quad (1)$$

ただし、 n_i と n_j はそれぞれAI和歌と人間和歌の総単語数を表し、 n_i は γ 和歌と人間和歌における同一単語の数を表す。類似度係数の計算において、Dice係数の規則に従い、同じ単語が和歌の中で数回繰り返し使われている場合も一度のみとして数える。類似度係数は γ 和歌と人間和歌の類似性を表し、その係数の値が大きいほど、 γ 和歌と人間和歌の中でより多くの同一単語が使用されていることを表す。Dice係数に基づく類似和歌の検索方法は下記の二つのステップによって行われる。①ステップの一、研究者が人工的に対象和歌を歴史的仮名遣いに変換し、和歌を単語分けする。②ステップの二、対象和歌を「類似指標を用いた和歌検索システム⁽¹²⁾」に入力し、コンピューターによってデータベースの和歌との類似度係数を計算し、和歌を順位付けまで成し遂げることができたのである。

表1の和歌は、Dice係数による検索方法でデータベースから選び出したY和歌①と類似する二十首の人間和歌であり、これらの和歌の類似度係数は0.600から0.413の間である。

表1 Y和歌①「明けてゆく 峰の木の葉の 梢より 遥かに続く 小牡鹿の声」と類似する二十首の人間和歌

番号	和歌(歌集・巻・歌番号・作者)	類似度
①	山里は 峰の木の葉に 競ひつつ 雲よりおろす さ牡鹿の声(風雅集・巻五秋上・524・式子内親王)	0.600
②	焦がれゆく 千汐の木の葉 踏み分けて なのみあさまの さ牡鹿の声(内裏歌合・1187・詠み人知らず)	0.533
③	尾の上より 通ふ嵐に 類来て 松の梢に さ牡鹿の声(千五百番歌合・秋一・1187・詠み人知らず)	0.519
④	吹きおくる 嵐の伝に 聞こゆなり 峰遥かなる さ牡鹿の声(為世十三回忌和歌・123・詠み人知らず)	0.482
⑤	竜田山 峰の紅葉葉 散り果てて 嵐に残る さ牡鹿の声(夫木抄・巻十二秋三・4626・慈円)	0.482
⑥	深山より 暁露に 立ち滞れて 野原に通ふ さ牡鹿の声(嘉元百首・1434・俊光)	0.482
⑦	初瀬山 木の葉色づく 秋風に まづいねがたの さ牡鹿の声(続千載集・巻四秋上・400・知家)	0.482
⑧	露にふす 野辺の千草の 曙に おき滞れてゆく さ牡鹿の声(続拾遺集・巻四秋上・258・後鳥羽院)	0.482
⑨	山里は 籬のはきに つまこめて 庭になれゆく さ牡鹿の声(沙玉集・全・47・貞成親王)	0.466
⑩	秋はなほ 心づくしの 木の間より 月にもりくる さ牡鹿の声(壬二集・上・後鳥羽院建仁百首・541・家隆)	0.466

⑪	三笠山 月射し昇る 空晴れて 峰より高き さ牡鹿の声 (続古今集・卷五・秋下・450・道家)	0.466
⑫	かねてより 心ぞいとど すみ昇る 月待つ峰の さ牡鹿の声 (新拾遺集・卷五秋下・456・西行)	0.466
⑬	三室山 夜半の谷風 吹くままに 峰よりおろす さ牡鹿の声 (夫木抄・卷十二秋三・4628・慈円)	0.466
⑭	静なる 深山の月の 影更けて 空になりゆく さ牡鹿の声 (内裏御会・87・詠み人知らず)	0.466
⑮	秋の色は 木の葉のみかは 竜田山 袖にしぐるる さ牡鹿の声 (夫木抄・卷十二秋三・4624・慈円)	0.451
⑯	ありあけの 月は葉山に 傾きて 野辺に送るる さ牡鹿の声 (元応二年八月十五夜月十首・49・詠み人知らず)	0.428
⑰	山里は 松のとぼその 明け方に 庭より帰る さ牡鹿の声 (永享百首・秋・廿二・詠み人知らず)	0.428
⑱	いつとても 月に袂は 濡れこしを 湧きて今宵の さ牡鹿の声 (後鳥羽院御集・歌会・歌合歌・1590・詠み人知らず)	0.424
⑲	深山辺の 松の梢を 渡るなり 嵐に宿す さ牡鹿の声 (新古今集・卷五秋下・442・惟明親王)	0.413
⑳	眺めつる ありあけの月は 傾きて 山より出づる さ牡鹿の声 (千五百番歌合・秋三・1364・詠み人知らず)	0.413

表1の和歌①を例に類似度係数を次のように計算してみる。まず、之和歌①と人間和歌①を単語分けした結果、それぞれ以下の様になる。

明けてゆく、峰の木の葉の、梢より、遥かに続く、さ牡鹿の声
 山里は、峰の木の葉に、競ひつつ、雲よりおろす、さ牡鹿の声

A1 和歌と人間和歌の単語数はそれぞれ十八と十七であり、Dice係数の規則によりそれぞれ四回、三回出現する「の」は一回として数え、故にその総単語数は Σ 和歌も人間和歌も十五となる。 Σ 和歌と人間和歌の同一単語は「峰」、「木」、「の」、「葉」、「より」、「に」、「さ」、「牡鹿」、「声」であり、同一単語数は九なので、類似度係数は $9 \times 2 / (15 + 15) = 0.600$ となる。

和歌の文法構造と内容からみると、 Σ 和歌①と最も近い人間和歌は表1の和歌⑩である。A1和歌の第二句、第三句「峰の木の葉の梢より」と和歌⑩の第二句、第三句「心づくしの木の梢より」は文法構造が同じで、句末の同じ箇所を助詞「の」と「より」が使われ、第二句全体が第三句の最初の単語「梢」と「木の梢」を修飾し、二つの単語は意味合い的にも類似する。 Σ 和歌の第四句、第五句「遙かに続くさ牡鹿の声」と和歌⑩の第四句、第五句「月にもりくるさ牡鹿の声」も文法構造が同じで、それぞれ第四句の最後の動詞「続く」と「もりくる」の連体形により第五句を修飾する。以上から分かるように、Dice係数による検索方法はより早く、便利に参考となる和歌の範囲を提供してくれるが、範囲内から目標和歌と類似する和歌を選択する際には人工的に和歌の内容を判断しなければならない。

Dice係数による検索方法の特徴を紹介するため、 Σ 和歌①を例に従来の類似和歌検索方法とDice係数による検索方法の違いについて説明する。従来の方法に従い、キーワードとなる単語を国際日本文化研究センターの和歌検索サイトで検索する。例えば、 Σ 和歌①で「鹿」と「声」の組み合わせを検索すると、1218首の和歌がヒットする。あまりにも数が膨大なため、「梢」、「鹿」、「声」とキーワードを増やして絞り込み検索を行うと、ヒットする和歌は三首となり、表1の和歌③と⑨、及び下記の和歌となる。和歌③と⑩の内容は Σ 和歌に近く、類似度係数も高い和歌である。係数が小さく表に入り得なかった下記の和歌の内容を見ると、季節や主題が Σ 和歌①と大きく異なることが分かる。

鹿の山 梢に通ふ 浦風は 水に残る 漣の声

(秋篠月清集・卷二夜百首・146・良経)

このように、従来の方法では検索キーワードを増やすことで和歌の範囲を狭めることが可能である。しかし三単語による絞り込み検索の場合、そのうちの二単語しか含まないものの、内容が目標和歌に非常に近い和歌を漏らしてしまう可能性がある。例えば、表1の人間和歌①は「和歌に内容の近い和歌であるが、「梢」、「鹿」、「声」の検索方法ではヒットせず漏れてしまう。従来の方法で「峰」、「鹿」、「声」で絞り込み検索すれば、六十二首ヒットし、表1の和歌①、⑥、⑦、⑩、⑫、⑬がその内に含まれる。「木の葉」、「鹿」、「声」で絞り込み検索すれば、二十首ヒットし、表1の和歌①、②、④、⑪、⑬がその内に含まれる。

故に、従来の方法で「鹿」、「声」と検索した場合はヒット数が多いにも膨大で、閲読に時間を費やす。キーワードを増やし絞り込み検索することで和歌のヒット数を減らすことはできるが、キーワードの選択に漏れがあった場合は、目標和歌に近い内容の和歌も漏れてしまう恐れがある。Dice係数を用いた類似和歌の検索方法は、類似度係数に基づいて類似和歌を順番付けすることができる。

三、Ⅱ和歌①「明けてゆく 峰の木の葉の 梢より 遙かに続く 小牡鹿の声」の分析

Ⅱ和歌①は早朝の山で鹿の鳴き声が梢の間を通って林に広がっていく様を描いている。赤や黄に染まった秋の木の葉が朝日を浴び、色彩豊かな風景が浮かぶ。近くの枝から遠くまで鹿の声広がり、キーワード「声」を際立たせている。

三二一、AI和歌と人間和歌の類似性の分析

まず、㊦和歌①と表1の人間和歌二十首の類似点について分析し、人間和歌の用語、文法、句の間の関連性を「和歌<L>」モデルが如何に学んだかについて検討する。

例を挙げると、㊦和歌の第三句「梢より」には副詞「より」が含まれており、「より」の前は名詞でなければならぬ。この組み合わせは「雲より」など、人間和歌①、③、⑥、⑩、⑪、⑫、⑬、⑰、⑳にあり、モデルが「名詞＋より」の文法を正確にマスターしたことが認められる。また、㊦和歌の第五句「さ牡鹿の声」は、表内のすべての人間和歌の第五句で使われている。秋を詠む和歌において、最後の句にさ牡鹿の声を描くことで、遠くまで響く鹿の声を読者に想像させ、和歌の余韻を残すことができるのである。和歌のこのような細やかな作風を「和歌<L>」モデルは学習し、㊦和歌の生成に応用することが出来たのである。国際日本文化研究センターの和歌検索システムによる検索で、㊦和歌①の「声」と「梢」を同時に含む和歌は百五十四首あることが分かった。これらの人間和歌は主に鶯、蝉を主題とするものが多い。Dice係数による検索方法を用いて、単語の類似性係数に基づいてこの百五十四首を順番づけると、一番と二番の和歌はそれぞれ表1の和歌③と和歌⑱となる。人間和歌③と㊦和歌①の類似点は鹿の声がと林という空間からではなく梢から響いて来るといふ描写である。和歌⑱も深山の鹿を主題とし、鹿の声は風の音とともに松の梢を通り、㊦和歌①の情景と類似する。

この㊦和歌の問題としては、第一に、「峰の木の葉の梢」という表現は、「峰」という大きな物から、「木の葉」という小さな物に視点が移ったのち、また「梢」という大きな物に移り、通常の大さきによる視点移動の順番と異なっ

ている。一般的なの順番に従えば、「峰の梢の木の葉」が正しく、物の大きさの認識とその順序表現については、モデルはまだ十分に学習できていないとが分かった。第二に、「明けてゆく」という連接は古典語にはほぼ存在せず、徐々に夜が明けるといふ表現には「明けゆく」(動詞の連用形+ゆく)という表現方法が古典語においては一般的である。モデルが「明け+て+ゆく」という三つの単語を組み合わせたのは、「徐々」にという意味の動作の進行や継続を表す補助動詞の「ゆく」を学習したのではなく、どこかの場所に「向かう、行く」という意味の動詞の「行く」を学習したためである。実際、データベースにおいて、「動詞の連用形+て+ゆく」の表現方法により動詞の「ゆく」を含む人間和歌は六百五十九首存在し、下記二首はその例である。

忘られて 年ふるさとの ほととぎす 何に一声 鳴きてゆくらむ

(後撰集・卷十四恋六・1006・詠み人知らず)

起きてゆく 人の心を 白露の 我こそまづは 思ひ消えぬれ

(後撰集・卷十二恋四・863・詠み人知らず)

この二首の和歌はいずれも、「鳴きながら飛び去って行く」、そして「起きて、行き去ってしまう」という意味であり、ここで「ゆく」は移動を表す動詞「行く」という意味で使用されている。このことから、進行や継続を表す補助動詞「ゆく」の使用方法について、モデルはまだ十分に学習できていないことが覗われる。

三―二、AI 和歌の独創性の分析

Ⅳ 和歌の用語の独創性を調べるため、Ⅳ 和歌①における四組の組み合わせを国際日本文化研究センターの和歌検

索システムを用いて検索する。「木の葉」と「梢」、「声」と「梢」と「続く」、「遙か」と「続く」と「続く」の四組の単語が人間和歌で使われている頻度を調べ、之と和歌の特徴を考察する。

「木の葉」と「梢」の組み合わせについて。人間の作品においてこの組み合わせは三十八首あり、これらの和歌の主題は主に梢の間から覗く「月」や梢の間を吹き抜ける「嵐」などであり、「鹿」など動物を主題とする和歌は含まれていない。下記の和歌を例に見ることにする。

小倉山 麓の里に 木の葉散れば 梢にはるる 月を見るかな

(新古今集・卷六冬・603・西行)

上記の和歌の焦点は木の上の梢にあり、木の葉が落ちて月の光が差し込む様子を描写している。「梢」や「木の葉」を主題とする和歌は視覚の焦点が近場であることが多い。対して、之と和歌①はまず焦点を近場の木の上に当てたのち、さらに「声」が「遙かに続く」の句によって描写の範囲を遙か遠くへ広げている。焦点を近場から遠くへ広げ、視覚と聴覚を組み合わせた独創的な和歌であると言える。

「声」と「梢」と「続く」の組み合わせについて。この組み合わせは人間和歌においては以下の三首のみ存在する。

柴下に 絶え間もなきは 夏山の 梢に続く 蝉の諸声

(続千載集・卷三夏・334・寂蓮)

ほととぎす 深き山江に 住む峽は 梢に続く 声を聞くかな

(山家集・卷下百首・1470・西行)

しみつせく 岩間の夏の 絶え間にて 梢は蟬の 声続くなり

(壬二集・卷上文治百首・933・家隆)

この組み合わせは人間和歌においては少なく、比較的新しい語句の使い方と言える。三首和歌の中で、第一首と第三首は蝉の鳴き声を、第二首はほととぎすの鳴き声を描き、対して二和歌は鹿の声を描く。更に、三首の人間和歌の下の句、「梢に続く蝉の諸声」、「梢に続く声を聞くかな」、「梢は蝉の声続くなり」から分かるように、動詞「続く」によって動物が鳴き続けるという時間の継続を表している。対して二和歌の「続く」は前に「遙か」があるため、時間の継続と同時に空間における声の広がりも表す。声が林という空間で遠くへ広がる様子は和歌に立体感を与え、斬新的である。

「遙か」と「続く」の組み合わせについて。この組み合わせは人間和歌においては三首しか存在せず、下記の三首である。

色々に ほむけの風邪を 吹きかへて 遙かに続く 秋の小山田

（玉葉集・卷四秋上・624・阿仏）

見渡せば 遙かに続く 小山田の 色濃き方は 仮初めにけり

（延文百首・143・進子内親王）

藤城や 山路遙かに 見渡せば 麓に続く 和歌の浦波

（熊野懷紙・53・後鳥羽院）

第三首の和歌は波が遠くへ広がる様子を描き、第一と第二首の和歌はともに「遙かに続く」によって修飾されているのは「小山田」である。故に、人間の和歌において「遙かに続く」は主に平面的な景色が遠くまで広がる様を描く時に使われる表現であり、対して二和歌は鹿の声が林という空間内で遠くまで広がるという立体的な描写で、その斬新さがうかがえる。

「遙か」と「続く」と「声」の組み合わせについて。前出の「遙か」と「続く」の組み合わせに「声」を加えて検索するとヒットする人間和歌はなく、ㄱ和歌における独自の組み合わせであることが分かった。鹿の声は秋の風物詩であり、その声が近くから遠くへ、色づく木の葉を渡って林の遙か先まで広がる情景は美しくも寂しく、儂いものである。その情景をㄱ和歌は独自のな語句の組み合わせによって表現したのである。

以上の分析から分かるように、「和歌<」モデルは人間和歌の主題や文法を正確にマスターし、単語の使い方を応用することができる。「遙かに続く」一句は人間和歌では主に平面的な広がりを描くものであるが、ㄱ和歌においては山に広がる鹿の声を描き、「和歌<」モデルが独自のな作品を生成することができることが分かった。更に、モデルは「声」、「梢」、「遙か」、「続く」など、人間和歌にない単語の組み合わせを使い、立体的で斬新な内容となっている。

四、ㄱ和歌②「山里の 松の梢に 埋つもれて 霞を分けぬ 鶯の声」の分析

四一、ㄱ和歌の分析

ㄱ和歌②の獨創性は主に以下三点である。第一に、下の句「霞を分けぬ 鶯の声」における「分く」は擬人的手法を用いている。自然界において霞が散るのは、日の光や暖気によるものであるが、ここでは鶯の声によって打ち破られたという趣のある表現となっている。鶯の声が霞を打ち破り、山里の霞を散らせるという、早春の景色を擬人的手法を通し、生き生きと描いているのである。第二に、第一句から第四句における「山里の」、「松の梢に」、「埋つも

れて、「霞を分けぬ」はすべて最後の句、「鶯の声」にかかっているのである。本題の「鶯の声」は和歌の第五句に隠されて、和歌の最後に現れる。「鶯の声」が山里深く、梢や霞の間に「埋つもれ」ていると描写されている通り、和歌の構造においても深く隠されているのである。第三に、冬を表す「松の梢」の間から春を表す「霞」が立ち、「鶯の声」が聞こえて、冬が去り春がやってくる情景から、この和歌は気候を表す単語を正確に使用したことによって季節感に富んだ内容を描いていると言える。

IV 和歌②の問題は、「埋つもれる」という一語は、宿や植物などが雪に埋もれる様子を描く際に使われる場合がほとんどで、鶯の声が霞に「埋つもれる」という表現は和歌の伝統からすれば、いささか大胆すぎるのである。

四―二、助動詞「ぬ」の解釈

第四句の「分く」は「開く、分ける」などの意味であるが、後ろに続く「ぬ」を打ち消しの助動詞「ず」の連体形「ぬ」として否定の意味ととらえる場合、下の句は「霞を打ち破ることなく、鶯の声が梢の中に埋もれている」という意味になる。「ぬ」を完了を表す助動詞「ぬ」の終止形「ぬ」として捉える場合、下の句は「鶯の声が霞を打ち破って出て来た」と言う意味になる。「分く」が下二段動詞のため、未然形と連用形は同じく「分け」であり、「分く」の活用形によって「ぬ」の性質を特定することはできず、文脈からの判断を行なうことにする。ただ、前後とのつながりから考えれば、完了を表す助動詞「ぬ」として捉えた場合、「ぬ」は終止形であるため、和歌は第四句で一度区切られてしまい、第五句の「鶯の声」とのつながりが悪くなってしまう。打ち消し助動詞「ず」の連体形として捉えた場合、「分けぬ」が修飾する体言はすぐ後ろの「鶯の声」と解釈できるため、打ち消しとして捉えた方がつながりが

良いのである。

また、文脈から判断する場合、この和歌は早春において鶯の声が春を告げる内容と考えられるため、「ぬ」を完了を表す助動詞と捉え、鶯の声が霞を打ち破って上空まで響いている様子を描いているものと解釈することもできるが、下記の二首人間和歌の内容を参照し、打ち消しの「ぬ」として捉えるものとする。「ぬ」を打ち消しの助動詞「ず」の連体形として捉えた場合、㊦和歌の下の句は、「鶯の声は霞を分けない」という意味になるので、「鶯の声が霞の内にある」、或いは「鶯の声は霞と共にある」という内容になると考えられる。類似する内容を描いた和歌の例は下記二首である。

山深み 霞こめたる 柴の戸に ともなふものは 谷の鶯

(玉葉集・1815・西行法師集)

たづねつる 宿は霞に うづもれて 谷の鶯 一声ぞする

(後拾遺和歌集・春上・23・藤原範永)

上記の第一首人間和歌においては、霞が山深くの宿を覆い、柴の戸も、谷の鶯も、その霞に覆われた様子を描いている。㊦和歌の「ぬ」を打ち消しとして捉えた場合、鶯が霞の内に隠されているという意味において、㊦和歌と人間和歌は非常に近いと言える。この人間和歌で描かれているのは、鶯そのものの姿が霞に隠された様子であり、対して㊦和歌は鶯の姿よりもその声の描写の対象となっているところに主な違いがある。

また、上記の第二首人間和歌においては、尋ねようと訪れた宿は霞に覆い隠されており、どこからか谷の鶯の一声が聞こえてきた様子が描かれている。この和歌も鶯の声が霞のうちに込められている様子を描いており、㊦和歌の「ぬ」を打ち消しとして捉えた場合の意味合いに似ている。

いずれにせよ、古人の和歌においては、鶯の声が霞と一体となつて描くのが一般的と思われ、「分けぬ」と「鶯の声」のつながりの良さも考慮し、ここでは「ぬ」を打ち消し助動詞「ず」の連体形と捉え、鶯の声は霞を分けて、内に込められている様子を描いているものとする。

ところで、上記の二首和歌は国際日本文化研究センターにおいて、「鶯」、「込める」、「埋つもれる」などのキーワード検索により得られたものである。この場合、D₁₀係数を使って単語の類似度から検索するよりも、和歌の一首全体の意味合いをキーワードによつて検索する方が「鶯が霞の中に隠されている」という具体的な内容の和歌にヒットすることができることが分かった。故に、我々はD₁₀係数による検索方法を和歌検索の唯一無二の方法として提案するのではなく、研究者が和歌を検索する上での参考の一つとして提示することが出来ればと考えている。

四一三、AI 和歌と人間和歌の類似性の分析

表2の和歌は、D₁₀係数による検索方法でデータベースから選び出した Σ 和歌と類似する十首の人間和歌である。

表2 Σ 和歌②「山里の 松の梢に 埋つもれて 霞を分けぬ 鶯の声」と類似する十首の人間和歌

番号	和歌(歌集・巻・歌番号・作者)	類似度
①	古巣立つ 都も春の 梢にて 霞を出でぬ 鶯の声(正治後度百首・210・雅経)	0.642
②	春來ぬと 霞む景色を しるべにて 梢に伝ふ 鶯の声(拾遺愚草・二見浦百首・107・藤原定家)	0.592

⑩	山の端は 分けよるままに 現はれて 霞を込む 松のむらだち (寂蓮法師集・163・寂蓮法師)	0.357
⑨	山高み 霞を分けて 眺むれば 遙かに見ゆる 住吉の松 (文治六年女御入内和歌・春・22・詠み人知らず)	0.370
⑧	鶯の 声をともにて 暮らすかな 一人眺むる 春の山里 (源広綱朝臣歌合・3・詠み人知らず)	0.461
⑦	山里の 草の閉ざしに ことよせて 心細きは 鶯の声 (堀河院御時百首和歌・春・54・源頭仲)	0.500
⑥	秋風に 雲も残らぬ 山の端の 梢を分けて 出づる月影 (新葉集・秋上・316・詠み人知らず)	0.500
⑤	竜田山 梢の嵐 秋暮れて 変はらぬ松に 声時雨るなり (仙洞十人歌合・67・詠み人知らず)	0.518
④	深山辺の 雪解の雲を 分け果てて 霞にうつる 鶯の声 (正治初度百首・808・慈円)	0.538
③	いたづらに 松の梢し 埋つもれぬ 誰踏み分けて 峰の白雪 (内裏百番歌合・1501・詠み人知らず)	0.638

表2の和歌は、Dice係数による検索方法でデータベースから選り出した Δ 和歌と類似する十首の人間和歌である。

表2の人間和歌①の第四、五句は「霞を出でぬ鶯の声」であり、 Δ 和歌の第四、五句「霞を分けぬ鶯の声」に非常に近い。モデルは、下二段活用動詞「出づ」に助動詞「ぬ」が続く場合、連用形「出で」に変化することを学習したものとと思われる。故に、 Δ 和歌の第四句の下二段動詞「分く」は、同じくこの活用方法によって助動詞「ぬ」を接続させ、文法が正確であり、内容も意味合いがはっきりしている。更に、鶯の声を「分く」ことは、単に霞から「出でる」よりもっと力強く、春を待ちわびる気持ちがかげえる。以上から、モデルは人間和歌から文法と文学的表現を学習し、単語を入れ替えて、正確な活用形を適用させることにより、独自性に富んだ和歌を創作できることが分かった。

表2の人間和歌②は Δ 和歌と類似し、春先の霞んだ景色と、梢に響く鶯の声を描いている。「霞む」、「梢」、「鶯」、「声」など同じ単語が使用されている。しかし、人間和歌②において、鶯の声はただ梢より聞こえて来るものであるが、 Δ 和歌において、鶯の声は自ずから霞を打ち破って、外に聞こえるような能動的な存在である。

同じく、表2の人間和歌③の内容も Δ 和歌と類似し、山里において鶯の声を聞きながら、春を待ちわびる様子を描いている。一首の単語の組み合わせを見ても、「鶯」、「声」、「山里」など、同じ単語が使用されている。しかし、人間和歌③はあくまでも人間の視点を中心とし、鶯を春を告げる景色の一つとしてとらえているに過ぎない。それに対し、 Δ 和歌は鶯の声を主役に置き、霞を分けて春を告げる力強い印象を与え、一首の内容の中心として描いている。

以上から、「和歌 Δ 」モデルは人間和歌から単語の組み合わせを学習し、独創的に単語の組み合わせることにより、和歌の描く情景に似た内容の Δ 和歌を創り出すことができることが分かった。Dice係数を用いれば、目標和歌と使用単語が類似している和歌を選り出すことができ、更にその内容の違いについて分析することができる。

五、Ⅱ 和歌③ 「冬来ぬと 思ひ明かしの 柴の庵 枕に残る 霜の音かな」 の分析

五―Ⅰ、ⅠⅠ 和歌の分析

ⅠⅠ 和歌③においては、第二句に掛詞「明かし」が使用されて、「夜を明かす」の意味と、地名「明石」と二重の意味を持つ。「明石」を詠んだ歌と言えば光源氏が思い出されるため、この一首の主人公を光源氏と想像することもできる。更に、下の句「枕に残る霜の音かな」は独特な表現である。現実世界において、霜は木の葉や草など室外の地面近くに降るものであり、視覚でとらえられるものである。しかしⅡ和歌③において、霜は室内、更には枕元に降っているという幻想的な描写をしている。加えて、本来視覚でとらえるべき白い霜には実は音があり、その音が枕に響き残っているのように詠んでいる。こうした多重の非現実的な描写手法により、初冬の海辺の庵の寒さが凜として伝わり、一晩中物思いに更け、夜を明かした主人公の寂しさが一層引き立てられる。⁽¹³⁾

このⅡ和歌の問題としては、第一に、「思い明かす」ことは、恋など心情にまつわる事柄故に、一晩中思い明かすという題材が伝統的である。冬が来たことを思い明かすということは通常では考えられず、おそらくは冬が来て寒くなった、あの人はどうしているのだろうか、という内容が省略されているものと考えられる。第二に、霜が室内や枕に降るということは、月の白い光を霜に喩えていると考えることができても、霜の音というのは何の喩えであるのかわかりづらく、比喩の範疇を超え、大胆すぎる手法となっている。

五―Ⅱ、ⅠⅡ 和歌と人間和歌の類似性の分析

表3-12 和歌③ 「冬来ぬと 思ひ明かしの 柴の庵 枕に残る 霜の音かな」と類似する十首の人間和歌

番号	和歌(歌集・巻・歌番号・作者)	類似度
①	秋来ぬと 枕につくる 鐘の音に やがてもかかる 袖の露かな(和歌所影供歌合・18・詠み人知らず)	0.533
②	風の音も 雲の景色に 冬来ぬと 思ひ果つれば 打ち時雨つつ(三百六十番歌合・冬・459・詠み人知らず)	0.516
③	冬来ぬと 今朝は岩田の ははそ原 音に立てても 降る時雨かな(続拾遺集・冬・380・知家)	0.516
④	冬来ぬと 思ふばかりに 山里の いとど寂しく なり勝るかな(新統古今集・冬・607・道濟)	0.500
⑤	秋来ぬと 今朝は知られて 色変へぬ 松に霜吹く 風の音かな(延文百首・2038・実俊)	0.484
⑥	冬来ぬと 思ふばかりの 朝ぼらけ ことのほかにも 変はる空かな(夫木和歌抄・冬・6339・嘉陽門院越前)	0.482
⑦	冬来ぬと つげの枕の したさえて まつ霜凍る うたた寝の袖(拾遺愚草員外・四十七首歌・232・藤原定家)	0.466
⑧	いつしかと 末枯れにけり 冬来ぬと いふばかりなる 霜の下草(弘長百首・冬・355・為氏)	0.466
⑨	冬来ぬと 時雨の音に 驚けば 目にも清かに 晴るる木のもと(拾遺愚草・千五百番歌合百首・1057・藤原定家)	0.451
⑩	冬来ぬと 夕霜寒き 浅茅生の 枯葉の風の 音ぞ寂しき(新統古今集・冬・610・定為)	0.444

表3の和歌は、Dice係数による検索方法でデータベースから選び出した2和歌と類似する十首の人間和歌である。和歌①の第二、三句においては、枕に響く鐘の音を描いている。袖が涙によって濡れるという情景の中、秋の鐘の音が枕元まで響いてくることによって、いっそう寂しさが引き立てられる様子を描いている。この人間和歌は主題と内

容においてㄥ和歌と類似している。現実的には枕につくことのない霜や、鐘の音を、枕に付いているかのように描くという比喩的手法を、モデルが人間和歌から学習したことがうかがえる一首である。ㄥ和歌と人間和歌①の一番大きな違いとしては、鐘には音があるが、霜は自然界では無声である。

表3の和歌②における「思い果つ」は「思いつめる」の意味であり、この和歌の第三、四句はㄥ和歌第一、二句の内容に近く、冬がやってきて物思いに耽る作者の様子を描いている。モデルは冬に物思いに耽るという主題を学習し、その単語の組み合わせも学習したことがうかがえる。更に、「思い明かす」は「思い果つ」より更に具体的で、一晚中考え込んでいた様子がありありと伝わり、モデルの和歌が独創性が描き出されている。

表3の和歌⑩は、「冬来ぬ」や「音」などの単語がㄥ和歌と一致し、冬の寒く寂しい情景を主題とする点において、ㄥ和歌と類似する。この人間和歌は枯れ葉を吹く風の音を描き、対するㄥ和歌は霜の音という幻想的な情景を描く。以上の分析から、「和歌ㄥ」モデルは人間和歌からその主題と単語などを学習したうえで、独創的に単語の組み合わせることにより、幻想的かつ大胆な内容を創り出すことができることが分かった。現在、「和歌ㄥ」の生成する和歌は人間の作品に完全に真似るレベルには至っていないが、人間が想像し得ない大胆な発想は、人間の創作に参考とアイデアを提供することを期待する。¹⁴⁾

六、結論

本論文はD₀係数に基づく和歌検索方法を提案し、この方法では類似度係数によって検索対象和歌を客観的に順位付けすることができるようになった。この和歌検索方法は研究者が和歌を検索する上での参考の一つとして提示す

ることが出来ればと考えている。「和歌 \times 」モデルが生成した三首の \times 和歌の用語を人間和歌と比較した結果、「和歌 \times 」モデルは人間和歌の主題、文法、句の構造、及び単語を正確にマスターし、独創的な単語の組み合わせ方により斬新的な和歌を生成できることが明らかになった。本研究は \times 和歌の研究のみならず、人間和歌の研究にも役立つことができる。

謝辞

本論文で使用された和歌生成モデル「和歌 \times 」は、西安交通大学コンピューター学院、楊新宇教授の指導のもとで、博士課程羅晶と修士課程牛明軒の協力を得て構築したものである。また、成蹊大学平野多恵教授（指摘と）意見を賜ることができた。この場を借りて感謝の意を表する。

〔注〕

- (1) Wu X, Iyen M, K, Ito K, Chen Z [Haiku generation using deep neural networks] (『言語処理学会第23回年次大会発表論文集』一一三三頁、二〇一七年三月)・廣田敦士、岡夏樹、荒木雅弘、田中一晶「学習データセットを分けたseqGANによる俳句生成」(『言語処理学会第24回年次大会発表論文集』一二九二頁、二〇一八年三月)・米田航紀、横山想一郎、山下倫央、川村英憲「深層学習を用いたモチーフ画像に基づく俳句生成」(『人工知能学会社会におけるAI研究会研究資料』二〇一八年三月一日)を参照。

- (2) T.Masada, A.Takasu [LDA-based scoring of sequences generated by RNN for automatic Tanka composition]

- 〔『International Conference on Computational Science』三九五頁、二〇一八年六月〕。
- (3) 中辻真、奥井颯平、野口あや子、加古陽「歌人A(人工智能歌人)の歌力ーついに歌壇デビューか！筆名は『恋するA歌人！』」〔『短歌研究』八号、九一頁、二〇一九年〕。
- (4) Takeishi Y, Niu M, Luo J, Jin Z, Yang X「WakaVT: a sequential variational transformer model for Waka generation」〔『Neural Processing Letters』二〇二一年〕 (<https://doi.org/10.1007/s11063-021-10654-z>)。
- (5) 金中「和歌Aで中日詩歌交流を促進」〔『人民中国』六号、二〇二二年六月〕。
- (6) データベースの構築方法に関しては、注(4)を参照。
- (7) Grzegorz K, Daniel M, Kevin K「Cognates Can Improve Statistical Translation Models」〔『Proceedings of HLTA-NAACL 2003: Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics』四六頁、二〇〇三年五月〕。
- (8) 下記ウェブサイトWeb茶まめには国立国語研究所の形態素解析ツールが含まれており、このオンライン解析ツールを用いて和歌などの古文を単語分けすることができむ。(<https://chamame.ninjal.ac.jp>)。
- (9) 本論文において、モデルが生成したA和歌に対し、人によって詠まれた和歌を「人間和歌」として定義づけする。
- (10) Dice係数は研究者Lee Raymond Diceによって命名され、類似度を測定する関数である。Dice係数の簡単な紹介については下記ウェブサイトを参照いただきたい。(<https://mathwords.net/jaccardkeisu>)。
- (11) Sutton G, McCallum A「An introduction to conditional random fields」〔『Foundations and Trends in Machine Learning』四号、二六七頁、二〇一二年〕。

(12) 類似指標を用いた和歌検索システムは作者の研究チームが構築した人間和歌データベース、和歌検索とDice係数計算のソフトを含む。

(13) 金中「中国西安交通大学における Σ 和歌創作の研究——村尾誠一先生のご指導と関連して」『樹間爽風』創刊号 六一―二頁、二〇二二年十二月。

(14) 馬場公彦「 Σ が人間の創造能力を超越する日は来るか?——西安交通大学の「和歌」自動生成プログラム」(<https://hon.jp/news/1.0/0/31262>)。AI和歌がいずれの日か、人間の作品を超えることができるかについて、馬場公彦氏の記事を(参照いただきたい)。

〔付記〕 本稿は、第45回国際日本文学研究集会（国文学研究資料館主催、令和四年五月）における口頭発表の原稿を基に改稿したものである。