

俳句における品詞の遷移の分析-俳句生成での利用に向けて-

Analysis of Transition of Part of Speech in *Haikus*: Toward Using in *Haiku* Generation

伊藤 拓哉[†], 小方 孝[†]
Takuya Itou, Takashi Ogata

[†]岩手県立大学ソフトウェア情報学部
Faculty of Software and Information, Iwate Prefectural University
g031n019@s.iwate-pu.ac.jp

Abstract

The authors studied narrative generation from *haikus*. This paper describes the analysis of transition of parts of speech in *haikus*, *haiku* generation using the result of this analysis, and questionnaire results regarding the comprehension of the generated *haikus*. Eventually, the authors aim to generate narratives using automatically generated *haikus*, utilizing it in an integrated narrative generation system they have developed.

Keywords — *Haiku*, Narrative Generation

1. はじめに

新田は文章を入力として、俳句を生成する方法を提案した[1]. 図1にその過程を載せる.

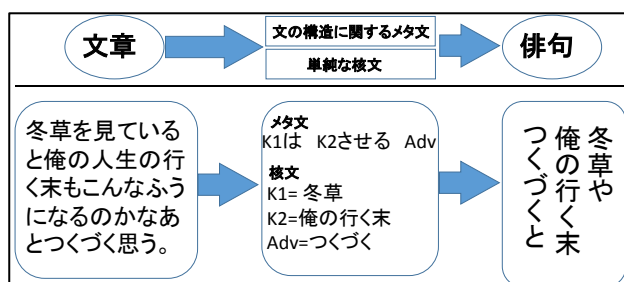


図1 文章からの俳句の生成過程

入力として与えた文の構造に関するメタ文を設定し、そのメタ文に当てはまるように、意味が単純で自明な核文を設定している. 図の例で言えば、文に関するメタ文が「K1は K2 させる Adv」という形で設定され、K1, K2, Advに当てはまるように文から核文が作られ、最終的に俳句「冬草や 俺の行く末 つくづく」とが生成されている.

筆者は俳句から物語を生成する手法の研究を行っている[2]. 俳句を入力としたとき、俳句を形態素解析し、獲得した名詞や動詞から関連する他の単語を選び、俳句に関する新しい解釈文の生成方法を研究している.

今回は、入力として利用する俳句を、新田のアプローチとは異なる方法で生成することを目指し、俳句中での個々の単語の品詞に注目し、品詞の遷移パターンを分析、そのパターンに基づいた俳句生成を行った. この研究の大きな目標は、筆者らが開発中の統合物語生成システムの入力として俳句のような定型詩を扱えるようにすることである.

2節では、名詞と動詞に限定した品詞の遷移の分析について、3節では全品詞の遷移の分析について述べる. 4節で実際に生成した俳句の評価の実験とその結果に関する考察を述べる.

2. 名詞と動詞の遷移の分析

品詞の遷移を分析するにあたり、まず俳句において最も出現する頻度が高いと考えられる名詞と動詞のみに絞って分析を行った. 2.1節ではその分析の方法について説明する. また2.2節では、分析の結果を踏まえて俳句の型を作成し、試験的に生成した俳句について説明する.

2.1. 俳句に含まれる名詞と動詞の遷移の分析

今回の分析では、Yahoo! テキスト解析 Web API に基づく日本語形態素解析を行えるウェブサイト「みんなの知識ちょっと便利帳 日本語形態素解析¹」を使用した. このサイトでは、形態素解析の結果の出力を任意で調節することができ、今回の品詞の情報の獲得に有効であると考えたためである.

今回の分析では、松尾芭蕉らの「おくのほそ道」に掲載されている63句を対象とした. 俳句は掲載されている書籍などによって表記が異なるが、今回は「日本の古典をよむ 20 おくのほそ道 芭蕉・蕪村・一茶名句集」の表記を利用した[3].

この分析では、俳句の中で名詞と動詞がどのような

¹ みんなの知識ちょっと便利帳 日本語形態素解析(最終確認日 2017 年 6 月 30 日)
http://www.benricho.org/moji_conv/japanese-analysis.php

順番で出現しているか、また俳句の中に、いくつ品詞(ここでは名詞と動詞)が含まれているかを調べた。解析結果の一例を表1に示す。

表1 名詞、動詞のみの品詞遷移解析結果の一例

俳句	草の戸も 住替る代ぞ 雛の家
品詞遷移	名詞→名詞→動詞→名詞→名詞→名詞
名詞	草, 戸, 代, 雛, 家
動詞	住替わる
品詞数	6

表2 名詞、動詞の出現数

	1	2	3	4	5	6	7	合計
名詞	59	51	32	42	33	14	1	232
動詞	4	12	30	17	6	0	0	69
合計	63	63	62	59	39	14	1	

表2にこの解析の結果をまとめた。この表では、獲得した品詞の一番目から七番目で名詞と動詞がそれぞれいくつ出現したかを表している。ただし、この分析では名詞と動詞の出現の順番にのみ着目しており、獲得する品詞を名詞と動詞に限定したため、「あかあかと日は難面も 秋の風」のように俳句の冒頭が副詞などで始まる場合は名詞の「日」が一番目となっている。

一番目、二番目では名詞の出現する割合が動詞に比べて圧倒的に高く、特に一番目が動詞である俳句は4句なのに対して、名詞は59句である。また名詞、動詞を六個以上含む俳句は63句のうち15句であり、七個含む俳句は1句であった。なお一つの俳句に含まれる品詞(名詞と動詞)の平均はおおよそ4.76個であり、動詞を二個以上含む俳句は63句のうち13句だった。

2. 2. 名詞と動詞の遷移を利用した俳句生成

俳句における名詞と動詞の遷移に関する情報を利用して、試験的に俳句を生成する。筆者らが開発中の統合物語生成システムが持つ、言語の表記に関する情報が記述された言語表記辞書を利用する。

前節の分析で、俳句の中に含まれる名詞、動詞の数は平均で4.76個、一番目、二番目の品詞は高い確率で名詞であり、動詞は二つ以上含まれる確率はあまり高くないといった結果を得た。これらの情報をもとに、俳句生成のための型を作成した。今回作成した型は名詞を4個、動詞を1個含み、動詞は三番目に固定した。三番

目に固定したのは、動詞の出現する割合が最も高かったためである。型に使用する名詞、動詞の音の数もあらかじめ指定しておき、最終的に十七音になるように「や」や「かな」といった、切れ字とよばれる一文字から三文字のものを組み合わせ調整した。以下の表3に生成のための型とそれによって生成された俳句を示す。

表3 名詞と動詞のパターンを利用した例

型	名詞(2)名詞(3)*動詞(3)名詞(3)切れ字(1)* 名詞(3)切れ字(2)
生成例	俗モデル 化けるおやまや 討っ手かな 嫁攻め手 見える飲み手へ 間抜けかな だれどっち 落ちるはとこか 野武士けり

表3で示した型は、括弧の中の数字が音の数を表しており、アスタリスクで五・七・五を区切っている。生成例に示した俳句は、俳句に必要な季語を特に意識していない。そのため川柳のような句になっている。

今回は、この名詞と動詞のみのパターンに加えて、俳句の冒頭に形容詞か形容動詞を含む俳句も生成を試みた。以下の表4にその生成のための型と生成例を示す。

表4 形容詞と名詞、動詞のパターンを利用した例

型	形容詞(5)または形容動詞(5)*名詞(3)動詞(3)切れ字(1)*名詞(2)の名詞(2)
生成例	広大だ マリア選ぶか 梵論の家 巨大だ ニンフ避けるか おじの舎 物凄いい やり手決るじ うばのモボ

3. 全品詞の遷移の分析

名詞と動詞のみの品詞の遷移の分析から、さらに詳細な品詞の遷移の分析を行った。3.1節ではその分析に関して説明し、3.2節では、この分析で得られた品詞の遷移パターンを利用した俳句生成について説明する。

3. 1. 俳句に含まれる全品詞の遷移の分析

形態素解析には2節と同様の形態素解析器を利用し、この形態素解析で得ることのできるすべての品詞(名詞、動詞、形容詞、形容動詞、副詞、連体詞、感動詞、接頭辞、接尾辞、接続詞、助詞、助動詞)に関して、その遷移を分析した。またこの分析では、俳句生成をより詳細に行うために、俳句を構成する五・七・五それぞれの品詞の構成と各品詞の音の数を分析した。以下の表5

と表 6 に分析結果の例を示す。

表 5 「草の戸も住替る代ぞ雛の家」の分析結果	
俳句	草の戸も 住替る代ぞ 雛の家
品詞遷移	名詞→助詞→名詞→助詞→動詞→名詞 →助詞→名詞→助詞→名詞
名詞	草, 戸, 代, 雛, 家
動詞	住替る
助詞	の, も, ぞ, の
音の数と構成	2 1 1 1 * 5 1 1 * 2 1 2

表 6 「あかあかと日は難面も秋の風」の分析結果	
俳句	あかあかと 日は難面も 秋の風
品詞遷移	副詞→助詞→名詞→助詞→名詞→ 接尾辞→助詞→名詞→助詞→名詞
名詞	日, 難, 秋, 風
副詞	あかあか
助詞	と, は, も, の
接尾辞	面
音の数と構成	4 1 * 1 1 2 2 1 * 2 1 2

音の数と構成の項目では、五・七・五をアスタリスクで区切って、それぞれの品詞の音の数と構成を表している。表 5 に関して言えば、最初の五音は四つの品詞から成り立っており、二音、一音、一音、一音という構成になっている。

表 6 では、名詞、副詞、助詞のほかに、接尾辞「面」が形態素解析の結果として得られている。しかし、この俳句に含まれている「難面」という語は「つれなく」と読み、「つれなく」は形態素解析では形容詞と解析される。「おくのほそ道」に掲載されているような俳句には、現代ではほとんど使用されないような表記があり、これらが予想とは異なる形態素解析の結果となる。今回の解析では俳句の表記に基づいた結果を利用し、特に修正は行っていない。この分析によって得た品詞の遷移のパターンを論文の最後に掲載した図 3 に示す。

3. 2. 全品詞の遷移パターンを利用した俳句生成

前節で獲得した品詞の遷移のパターンを利用して俳句の生成を行う。現在、統合物語生成システムの言語表記辞書には名詞、動詞、副詞、形容詞、形容動詞、助詞の情報がある。しかし今回の分析の結果では接尾辞や助動詞など言語表記辞書にはない品詞も獲得された。

そのため言語表記辞書を利用して俳句生成を試みると、生成は辞書がある六品詞のみの利用に限られてしまう。この問題を解決するために、今回は「おくのほそ道」の 63 句の形態素解析の結果から得られた単語を利用して俳句を生成した。この方法では、言語表記辞書を利用する場合に比べ、利用できる単語の数ははるかに少ないが、獲得しているパターンの中に出現する品詞はすべてカバーすることができる。以下の表 7 に品詞遷移のパターンと、そのパターンを利用して生成した俳句の例を示す。

表 7 パターンを利用した俳句生成の例	
パターン	名詞助詞*名詞名詞助詞*名詞助詞名詞
音の数と構成	4 1 * 2 4 1 * 1 1 3
生成例	松島を 鳥撫子に 緒の幟
	木啄の 岩兵ぞ 夏の初
	一枚の 関啼魚ぞ 緒の白毛

4. 生成した俳句の予備的な評価

名詞と動詞のみのパターンから生成した俳句と、全品詞の遷移パターンを利用し生成した俳句に対して試験的な評価を行った。4.1 節では生成した俳句の面白さの調査について述べ、4.2 節でその考察を行う。また、4.3 節では生成した俳句の内容の理解度の調査について述べ、4.4 節でその考察を行う。

4. 1. 生成した俳句の面白さに関する調査

2.2 節で示したような、名詞と動詞の組み合わせのパターンから生成した俳句と、3.2 節で示したような全品詞の遷移パターンを利用して生成した俳句をそれぞれ 10 句ずつ選び、計 20 句を対象としてアンケート調査を行った。アンケート調査では、これら 20 句の俳句を被験者に読んでもらい、どの程度おもしろいと感じたかを五段階で評価してもらった。今回の実験では、生成された俳句の意味などを深く考えさせるようなことはせず、あくまで一読しただけでどの程度おもしろいと感じたかを調査するために、アンケートの回答時間は 2 分程度とした。

今回のアンケート調査の被験者は 16 名で、十代から二十代前半の男女の学生であった。

表 8 にアンケートで使用した俳句と、アンケートで得られた評価の平均値と評価の順位を載せる。番号 1 から 10 が名詞と動詞のみのパターンから生成した俳句、

番号 11 から 20 までは全品詞の遷移パターンを利用して生成した俳句である。なお、平均値は少数第三位で四捨五入した値である。

表 8 面白さに関するアンケート結果

番号	俳句	平均値	順位
1	勇ましい お化け溶けるぬ 試補の兵	3.19	12
2	あどけない 名無し跳ねるへ 星の牛頭	3.38	2
3	冴えるらん 空き巢小揚げよ 理化の我	3.31	5
4	活発だ シンパ逃がすや 僕の宅	3.07	17
5	興すかな おどけせがれへ 保父の油屋	2.75	19
6	眠るかな ダーメのろまぞ 鼻祖の白	3.13	16
7	遺児シャー 換えるカルトよ 姪のパル	3.19	12
8	嫁攻め手 見える飲み手へ 間抜けかな	3.63	1
9	ママコンパ 飽きるすべたつ キブツかな	3.31	5
10	驚ファン 冷えるねんねぞ ボウズけり	3.31	5
11	一枚を 足撫子の 三の幟	2.73	20
12	風流に 松撫子の 目は桜	3.25	9
13	笠嶋に 下兼房の 三の田植	3.00	18
14	一枚の 関啼魚ぞ 緒の白毛	3.19	12
15	卯の花を こさ啼魚に 夏の草鞋	3.38	2
16	卯の花の 霧岩霧に めかり声	3.25	9
17	たけくまの 馬鳥花を しのぶ声	3.25	9
18	さみだれに 馬花尿や めかり紅粉	3.31	5
19	行春は 下いづ松の しのぶ蟬	3.38	2
20	夏山も 馬下足に しのぶ也	3.19	12

4.2. 面白さの評価に関する考察

今回のアンケートでは、名詞と動詞のみのパターンから生成した俳句と、すべての品詞の遷移パターンから生成した俳句の評価の差は大きく現れず、二つの生成方法の俳句それぞれすべての評価の平均値は、前者が 3.23 で後者が 3.19 であった。

今回最も高い評価を得たのは、平均値 3.63 で「嫁攻め手 見える飲み手へ 間抜けかな」という句であった。次いで平均値 3.38 で「行春は 下いづ松の しのぶ蟬」、「卯の花を こさ啼魚に 夏の草鞋」、「あどけない 名無し跳ねるへ 星の牛頭」が並んでいる。これらの句を参考に高い評価を得た要因を考察する。

- ・俳句の読みやすさ：上位の句には表記に平仮名が多く含まれていた。最上位の「嫁攻め手 見える飲み

手へ間抜けかな」は十六文字中八文字、2位の「あどけない 名無し跳ねるへ 星の牛頭」には平仮名は十六文字のうち十文字、「卯の花を こさ啼魚に 夏の草鞋」は十三文字のうち六文字、「行春は 下いづ松の しのぶ蟬」は十二文字のうち七文字と、上位4句はおおよそ半数以上が平仮名で表記されている。さらに5位タイで並んでいる4句も表記の半数以上が平仮名、または片仮名であった。このことから考察すると、平仮名や片仮名が多く、読みやすい俳句が高い評価を受けやすいのではないかと予想される。

- ・個々の表現のイメージしやすさ：高い評価を受けている句に含まれている表現のわかりやすさ、イメージしやすさも評価に繋がっているのではないかと予想する。上位の句の表現を抜き出してみると、「卯の花」、「草鞋」、「あどけない」、「跳ねる」、「行春」など単語一つからイメージができるようなものがあり、加えて「しのぶ蟬」、「ママコンパ」、「驚ファン」のように単語の組み合わせから場面をイメージできるようなものがある。今回1位になった句を見てみると、「嫁攻め手」、「飲み手」、「間抜け」と個々の表現がイメージしやすいだけでなく、一連の表現から一つのストーリーすら連想できそうである。このような点がこの句が最も高い評価を得た要因ではないかと筆者は予想する。

ここからは低い評価につながった要因を考察する。最も低い順位となったのは、平均値 2.73 で「一枚を 足撫子の 三の幟」という句であった。次いで、平均値 2.75 で「興すかな おどけせがれへ 保父の油屋」、平均値 3 で「笠嶋に 下兼房の 三の田植」という句が並んでいる。

- ・単語の読みにくさ：低い順位となった句を構成する個々の単語を見ると、どのように読めば良いのかわかりにくい単語がいくつか見られた。例えば「幟」や「保父」、「油屋」のような単語である。今回のアンケートの対象が学生であったこともあり、正しい読みができなかったのではないかと考えられる。16位の句でも「鼻祖」、12位タイで4つ並んだ句の中の1句にも「試補」のような、どのように読めば良いのか分かりにくい単語が含まれている。
- ・誤読による字余り：今回生成した俳句は音の数が五・七・五から逸脱しないように単語を選択しているが、

「三の田植」のように、「みのたうえ」と読ませたい部分が、読み手のほうでは「さんのたうえ」のように字余りだと思われてしまった可能性も考えられる。「三の幟」,「保父の油屋」でも同様の現象が起きたのではないかと予想される。

ここまで、アンケートで得られた評価を参考に、高い評価と低い評価のそれぞれにつながった要因の考察を行った。ここで考えられた要因はあくまで仮説であり、根拠に乏しい。実際「あどけない 名無し跳ねるへ 星の牛頭」という句は「牛頭」という読みにくい単語を含み、字余りと思われた可能性があるが、全体で二番目に高い評価を受けており、これらの要因を含む、含まないが評価に直結しているとは限らない。

4. 3. 生成した俳句の理解度に関する調査

俳句の面白さに関するアンケートの結果、その評価の要因として、読みやすさが考えられた。一方で、内容がどの程度理解されているのかは不明であった。表記以外にも内容の理解がおもしろさにつながるのではないかと考え、生成した俳句の内容の理解に関するアンケート調査を試験的に行った

面白さに関するアンケートと同じ俳句を使用し、その俳句の内容がどの程度理解できたかを五段階で評価してもらった。理解度について評価してもらうため、このアンケートには5分程度で回答してもらった。

実験の被験者は、面白さに関するアンケートの被験者とは異なり、今回は二十代から三十代の男女 10 名であった。表 9 にそのアンケートの評価の平均値と評価の順位、また面白さのアンケートでの順位を載せる。順位 1 が理解度に関するアンケートの順位、順位 2 が面白さに関するアンケートの順位である。番号は表 8 の面白さのアンケートの番号と対応している。

4. 4. 理解度の評価に関する考察

理解のしやすさで見ると、名詞と動詞のみのパターンから作成した俳句と、全品詞の遷移パターンから作成した俳句では、順位の平均は前者が 11.1 位、後者が 8.5 位と、全品詞のパターンを利用したほうがやや理解されやすかったようである。この要因として、全品詞の遷移パターンからの俳句生成では、「おくのほそ道」で実際に俳句に利用されている単語を使用しているため、単語レベルでの理解がしやすかったと予想される。逆に名詞、動詞のみのパターンからの俳句生成

表 9 理解度に関するアンケート結果

番号	平均値	順位 1	順位 2
1	2.2	15	12
2	2.4	12	2
3	2.2	15	5
4	2.8	8	17
5	2.9	5	19
6	1.7	19	16
7	1.5	20	12
8	3.1	3	1
9	2.4	9	5
10	2.9	5	5
11	1.9	17	20
12	3.4	1	9
13	2.5	10	18
14	1.9	17	12
15	2.8	8	2
16	2.4	12	9
17	3.1	3	9
18	2.5	10	5
19	2.9	5	2
20	3.3	2	12

では、統合物語生成システムの言語表記辞書中の単語を文字数のみの条件でランダムに選択していたため、日常的には利用しないような単語が選択され、理解を妨げたのではないかと考える。

面白さと理解度の関係を考えると、面白さと理解度のそれぞれの評価の平均値から相関係数を算出した結果、0.25 と低く、関係があると断言はできない。実際、面白さで 5 位だった句が理解度では 15 位であるなど、面白さと理解度に比例の関係があると言える結果ではない。しかし、面白さで 1 位であった句が理解度でも 3 位であることなどから、前節の考察で述べたように、単語レベルから情景やストーリーがイメージできることが面白さにつながる可能性はあると考えられる。

5. 今後の課題

アンケートの結果を踏まえ、俳句生成に関して、今後の課題を挙げる。まず、今回のアンケートで低い評価に繋がったと考えられる、読みに関する点である。次回以降アンケートを行う際には、俳句にルビをふるなどし

て、正しい読みをすぐにわかるようにしたいと考えているが、それ以外にも、俳句を生成するときに、比較的読みやすい単語を利用するように、生成の仕組みを改善していきたいとも考えている。また、アンケートの数がまだまだ足りないため、もっと規模を拡大して評価の実験を行いたい。他にも今回の分析で獲得した他の品詞の遷移パターンを使って俳句を生成できるように、生成の仕組みや辞書の作成にも取り組みたい。

今後、筆者らが開発中の統合物語生成システムとの連動も考えている。図2に統合物語生成システムのアーキテクチャを載せる[4][5]。

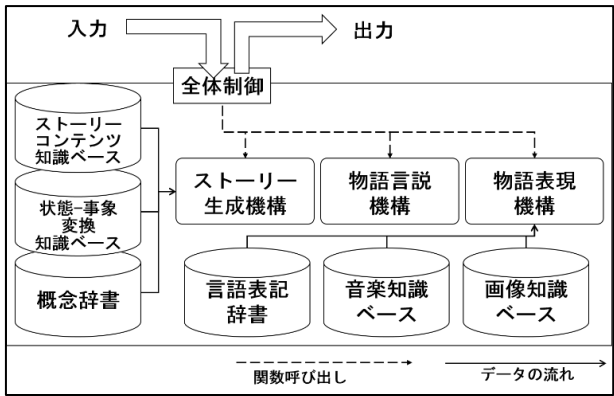


図2 統合物語生成システムのアーキテクチャ

統合物語生成システムは、ストーリー生成機構、物語言説機構、物語表現機構の3つの機構と、それらの機構が利用する知識機構によって構成される。今回、名詞と動詞のみのパターンで俳句を生成する際には、この知識機構の一つである言語表記辞書を利用した。全品詞の遷移パターンによる俳句生成では、この言語表記辞書がすべての品詞に対応していなかったため利用できなかった。そのため、まず言語表記辞書の拡張が課題として挙げられる。

統合物語生成システムは、これまでは入力として大局構造や物語の長さなど、物語の構成に関わる情報を扱っていた。筆者は俳句のような定型詩を入力として利用することを研究しており、今回の分析から得たデータに基づいて俳句を生成し、その俳句を入力として扱えるようにすることを考えている。

6. まとめ

今回は、俳句を自動生成するために「おくのほそ道」の63句の品詞の遷移のパターンを分析し、獲得した遷移パターンを用いて俳句の生成を行った。加えて生成

した俳句に関して面白さと理解度のアンケート調査を行った。このアンケート調査の結果、面白さに関しては単語の読みやすさ、表現のイメージしやすさが関係しているという仮説を立てた。また理解度に関しては、実際に俳句で利用されている語句を利用するほうがより理解されるという仮説を立てた。

今後、今回獲得した品詞の遷移パターンを利用し、さらに多様な俳句を生成できるようにし、面白さ、理解度を向上できるように生成方法を工夫していきたい。そして将来的には統合物語生成システムでの俳句の利用を目指す。

参考文献

[1] 新田義彦, (2015) “正規表現による俳句の解釈と生成”, 電子情報通信学会, 信学技報, Vol. 115, No. 69, pp. 31-36.
[2] 伊藤拓哉, 荒井達也, 小方孝, (2017) “俳句の意味ネットワークからの物語生成”, 2017年度人工知能学会全国大会(第31回)論文集, 1D3-OS-29b-1in2.
[3] 井本農一, 久富哲雄, 堀信夫, 山下一海, 丸山一彦, (2008) “日本の古典をよむ 20 おくのほそ道 芭蕉・蕪村・一茶名句集”, 小学館.
[4] Ogata, T. (2016). Computational and cognitive approaches to narratology from the perspective of narrative generation. In Ogata, T. & Akimoto, T. (Eds.). Computational and Cognitive Approaches to Narratology. 1-74. Hershey, PA: IGI-global.
[5] 小野淳平, 小方孝, (2017) “「ギャップ技法」を利用して「驚き」を作り出すストーリー生成の方法-テーブルトークロールプレイングゲームに基づく物語自動生成ゲームへのアプローチ-”, 認知科学, vol.24, No.3. (印刷中)

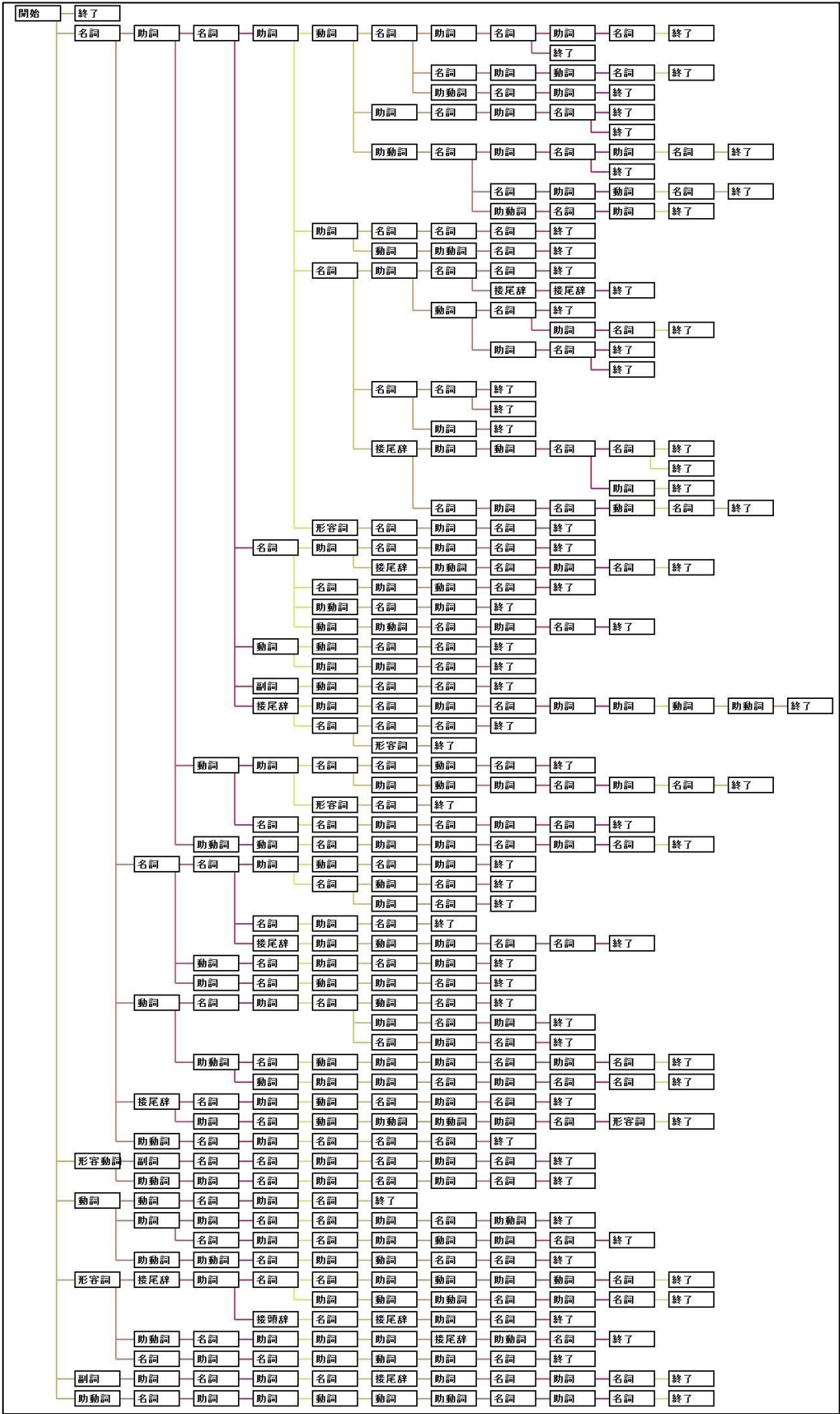


図3 分析した「おくのほそ道」の俳句の品詞遷移図