英語俳句投句支援システム構築に向けた構文解析

和田　武*1,　　墨岡　学*2
Email: wada@cite.ehime-u.ac.jp,　　sumioka@cc.matsuyama-u.ac.jp

*1：愛媛大学総合情報メディアセンター
*2：松山大学経営学部

◎Key Words 英語俳句，構文解析，対応分析，支援システム

1. はじめに

1994 年に我々のグループ（Shiki チーム）が立ち上げた正岡子規に関する英語俳句サーバ Shiki は，世界各国の俳句愛好家達に広く利用されている．今後，学生を始めとする初心者にも英語俳句に馴染んでもらうために，英語俳句の投句支援システムが必要と考えた．まず，英語俳句サーバのメーリングリスト Shiki Monthly Kukai のデータベースに形態素解析を加え，英語俳句でよく利用される語彙を月別に抽出集計した．次に，英語俳句の構造解析をはじめとする構文解析を試みたのでここに報告する．

2. 方法と結果

2.1 使用データと解析方法

2010 年に The Shiki Monthly Kukai に投句された月別投句数を表１に示す．左側は Kigo の部，右側は Free Format の部を示し，年間 2557 句，月平均 213 句が Kukai に投句されたことを示す．

表１．月別投句数

Year	Month	Kigo	投句数	Free Format	投句数	合計
2010	Jan	First Things	119	－	108	227
	Feb	Groundhog Day	97	－	119	216
	Mar	Planting/Sowing	118	－	111	229
	Apr	Emerging Animals	110	－	113	223
	May	Fishing	99	－	103	202
	Jun	Nakedness	86	Children	96	182
	Jul	Any Summer Grass	128	Anything Quirky	96	224
	Aug	August Moon	113	Moving	100	213
	Sep	Leaves Falling	147	Beach/Shore	117	264
	Oct	－	40	－	41	81
	Nov	Geese	127	Weaving	115	242
	Dec	Winter Sky	137	Ring	117	254
		合計	1321	合計	1236	2557

図１に，2010 年４月の Kukai で評価が高かった２句を示す．それぞれの句は３行詩で，切れ字は記号（―）や体言止めが用いられていることがわかる．なお，４行目は作者を示している．

dusk －　　railroad crossing
the geese　　their goodnight kiss
just darker than the sky　　one hundred boxcars long
aom (tim)　　Edward

図１．投句例

今回の研究では，まずこの 2557 句に Tree Tagger [1] による形態素解析[2]（文章を意味ある単語に区切り品詞や原形を求める）を加えた．例えば，”She like a cake or something like that.” に対して形態素解析を加えると，She / 代名詞　like / 他動詞　a / 冠詞　cake / 普通名詞　or / 接続詞　something / 代名詞　like / 前置詞　that / 代名詞　./ 記号
といった情報が得られる．この情報に対して，英語俳句によく用いられる語彙を月別に求めるために対応分析（コレスポンデンス分析）を行った．

次に，2010 年４月のデータを用いて，３行詩の１行ごとの語彙数のバランス，句ごとの名詞・動詞・形容詞などの数を調べた．解析には，IBM SPSS Statistics 21 とエクセル統計 2012 を用いた．

2.2 結果

図２は，対応分析（コレスポンデンス分析）を行った結果を示す．対応分析は，クロス表（χ^2 統計量で独立性の検定を行う）を基に，行と列の要素の相関係数が最大になるように数値化して次元縮約する方法で，χ^2距離法を適用する手法である．χ^2距離は，同じ要素の２点間の距離の２乗であり，２要素が似ているほどこの距離が小さく，離れているほどこの値が大きくなり，これらの関係が散布図として示される．

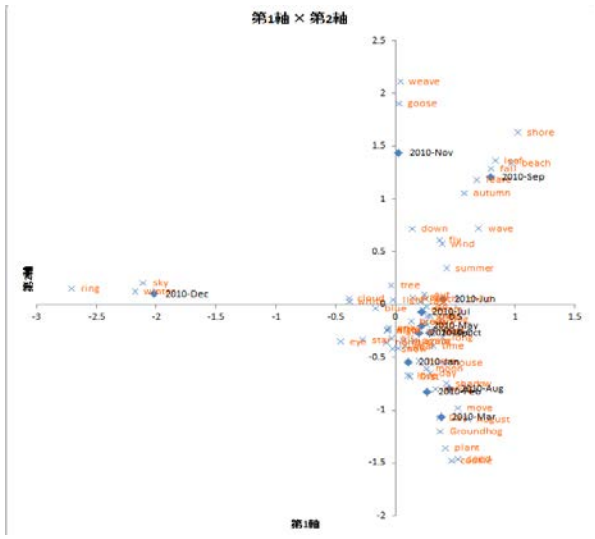

図２．対応分析（コレスポンデンス分析）

図２から，Kukai で指定された Kigo 他以外にも各月でよく使用される語句，例えば 9 月に autumn, 7 月に go, dream，４月に windows，５月に sun, rain, 1 月に home, 2 月では spring, 8 月では shadow などがよく利用されていることがわかる．なお，この分析では固有値表に示される累積寄与率が第２軸までで 31.6%であった．

次に，2010 年４月のデータを用いて，３行詩の１行ごとの語彙数のバランスを調べた結果を表２に示す．

表２．句別行別語彙数

句番号	1行目	2行目	3行目	計
1	2	2	5	9
2	2	3	3	8
3	3	3	3	9
4	2	5	3	10
5	3	5	2	10
6	4	3	3	10
7	2	2	3	7
8	2	4	3	9
9	3	2	1	6
10	3	4	3	10
11	3	2	2	7
12	4	4	0	8
(略)				
109	2	4	6	12
110	4	3	4	11
1	2	3	4	9
2	4	4	4	12
(略)				
113	3	6	4	13
合計	653	874	642	2169
平均	2.9	3.9	2.9	9.7
標準偏差	1.03	1.33	1.14	2.17
最大値	7	8	7	18
最小値	1	1	0	4

縦軸が句の番号で，2010 年４月は Kigo の部が 110 句，Free Format の部が 113 句の計 227 句が対象である．表２から，１行目は 2.9±1. 03，２行目は 3.9±1.33，３行目が 2.9±1.14 の語で３行全体で 9.7±2.17 語で構成されていることがわかる．

次に，句ごとの名詞・動詞・形容詞などの品詞のバランスを調査（表３）した．品詞の記号は次の通りである．接続詞 (CC,)，前置詞 (IN,)，冠詞 (DT,)，形容詞 (JJ, JJR,)，名詞 (NN, NNS, NP)，代名詞 (PP, PP$)，副詞 (RB, RBR)，記号 (SENT, -, ;,)，動詞 (VB, VBZ, VVZ,,) などである．

表３からわかるように，2010 年４月の Kukai に投句された句では，名詞 36.1%，冠詞 12.2%，動詞 12.0%，前置詞 9.9%に続き，記号が 9.4%使用されている．これらにより，英語は３行詩で，切れ字には記号が用いられていることがわかる．

３．　まとめ

英語俳句投句支援システムの構築を目的に，2010 年１月から 12 月に Shiki Monthly Kukai に投句された英語俳句をデータベース化し，形態素解析を加えたあと対応分析を行って，月別に英語俳句に用いられる語彙を抽出した．その結果，①１句当たり３行約 10 語で構成し，---や ; などの記号や体言止めを用いて切れ字を表す．②Kigo 以外にも月々でよく使用される語句，1 月に home, 2 月に spring, 4 月に window, 5 月に sun, rain, 8 月に shadow などがよく利用されていることがわかった．また，2010 年４月のデータを基に，３行詩の１行ごとの語彙数のバランスを調べた結果，１行目は 2.9±1.03，２行目は 3.9±1.33，３行目が 2.9±1.14 の語，３行全体で 9.7±2.17 語で構成されていることがわかった．

さらに，句ごとの英語俳句の名詞・動詞・形容詞の数を調べた結果，名詞 36.1%，冠詞 12.2%，動詞 12.0%，前置詞 9.9%に続き，記号が 9.4%使用されており，記号は切れ字等に用いられていることがわかった．

表３．句別品詞数（2010_Apr）

	A	B	C	D	E	F	G	H	I	J	K	HN	HO	HP	HQ
1	品詞	1	2	3	4	5	6	7	8	9	10	111	112	113	合計
4	CC						1								31
5	CD														18
6	DT	2	1	1	3	1	2		2		1	1	1	2	275
7	IN	1		1	2	2	1				1	2	2	1	223
8	JJ			2	1			2	1			2		1	171
9	JJR	1								1					8
10	NN	2	4	2	4	5	3	1	3	2	5	2	2	3	639
11	NNS	1	1	1				2	1	1				1	139
12	NP														35
13	PDT														2
14	POS							1							15
15	PP											1	1		34
16	PP$			1										1	46
17	RB	1					1	1							86
18	RBR														5
19	RP														14
20	-	1		1		1				1					71
21	SENT													3	42
22	;	1		1		1				1					73
23	,														26
24	TO		1								1				18
25	VBZ														12
26	VV										1				31
27	VVG		1					1	1	1					80
28	VVN						2								33
29	VVZ					1					1	2	2	1	65
30	VHP														5
31	VVP														25
32	VHZ														3
33	VVD														12
34	VBP														2
35	WP														1
36	WDT														1
37	WBD														1
38	MD														2
39	VH														1
40	WRB														2
41	UH														3
42	計	10	8	10	10	11	10	8	8	7	10	10	8	13	2250

今後，英語俳句投句支援システムの構築時には，英語俳句作成時に月ごとによく使用される語句が表示されるように画面を設計し，入力補助機能を用いて入力した文字列で始まる語の先読みや，表示された語句候補群から選択入力できる機能などを組み込み，ユーザフレンドリーなシステムを目指したい．

さらに，３行詩の動詞・名詞の位置関係などの分析をはじめ係り受け解析や人称（I, my, me, his など）の分析も行って，初心者でも英語俳句が簡単に作成できるような英語俳句初心者支援システムを構築したい．なお本稿の執筆にあたっては shiki チームの井上博民氏らに草稿段階から有益な助言をいただいた．ここに記して感謝する．

参考文献

1. http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/
2. 田中省作，形態素解析ツール-英語と TreeTagger を中心に-，九州大学情報基盤センター．