

シリトリシステムの開発

村田 真樹* 金丸 敏幸* 白土 保* 馬 青**,* 井佐原 均*

* 独立行政法人 情報通信研究機構 ({murata,kanamaru,shirado,qma,isahara}@nict.go.jp)

** 龍谷大学 理工学部 (qma@math.ryukoku.ac.jp)

1 はじめに

本稿では、われわれが開発したシリトリシステムについて紹介する。このシステムは、Web 上で動作するシステムとして構築した。Web 経由でつながるところならどこでも、このシリトリシステムを利用することができる。このシリトリシステムは、人間と計算機のシリトリの対戦が可能である。計算機の強さを調節することが可能である。また、連想的なシリトリが可能である。例えば、直前の回答と意味的に関連のある単語を利用して答える。言語処理分析技術を駆使した強力な戦法も利用可能である。また、ヒントを表示する機能や、単語の意味を表示する機能もある。Windows, UNIX のいずれでも動作する。

2 シリトリシステム

シリトリシステムの初期画面を図 1 に示す。

図のように Web ブラウザにおいて駆動するシステムである。人間と計算機の間でシリトリを行うことができる。図のようにニックネームをまず入力する。また、人間側が先攻か後攻を決めることができる。人間側が答える時間を制限するかどうかを決めることができる。また、シリトリの対象とする単語のジャンルを決めることができる。また、シリトリシステムのレベル(強さ)を変えることができる。ニックネームを入力して「ゲーム開始」のボタンを押すとゲームがスタートする。

図 2 のようにゲームが進む。人間側の先攻を選択したので、まず人間側が単語を入力する。単語はひらがなで入力する。ここでは「あり」(蟻)と入力した様子を示している。シリトリの最初の単語はどのような単語であってもよい。

図 3 は、人間の「あり」(蟻)に対して、計算機側が「りょうほう」(両方 / 療法 / 良法)と回答し、次の人

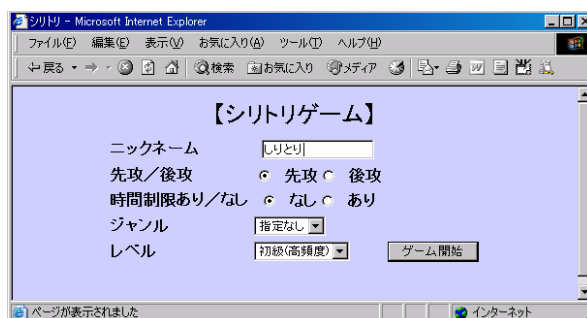


図 1: シリトリシステムの初期画面

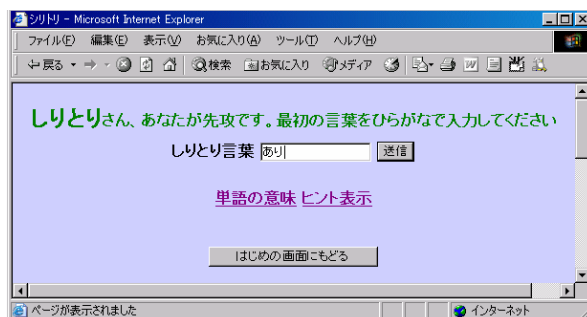


図 2: 「あり(蟻)」の人の入力

間の入力を待っている様子を示す。シリトリでは、相手の言った単語の末尾の文字から始まる単語を回答する。このため、「あり」の最後の文字の「り」から始まる「りょうほう」を回答している。このシリトリシステムは現在は同じひらがな表記の単語はすべてまとめて一つとして扱っている。その単語のうちいずれかが使われれば他の同じひらがな表記の単語も使えないというルールにしている。このため、「りょうほう」の入力で「両方 / 療法 / 良法」だけの単語を使ったものとみなし、以降のシリトリでは利用してはいけない単語となる。

ところで、このシリトリシステムでは、単語の意味を表示する機能がある。「りょうほう」としては「両方 / 療法 / 良法」という漢字表記があるがそれぞれの単語の意味が気になったとする。図 3 の「単語の意味」

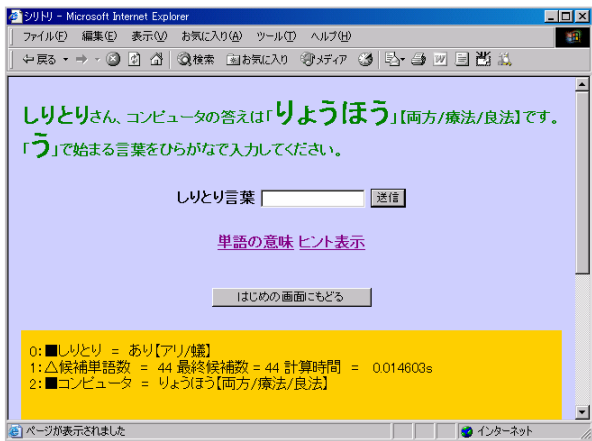


図 3: 計算機の回答: 「りょうほう」

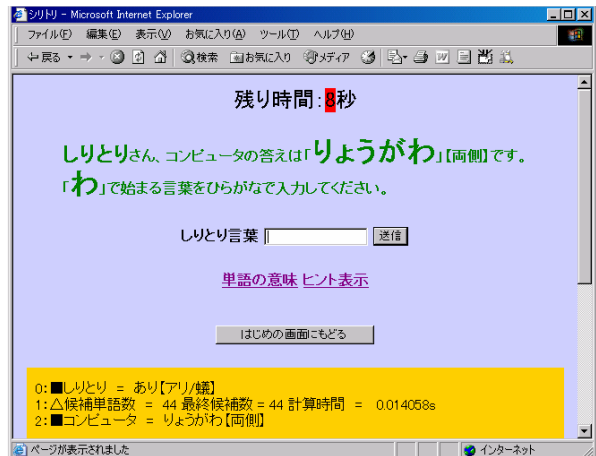


図 6: 時間制限ありの場合の画面

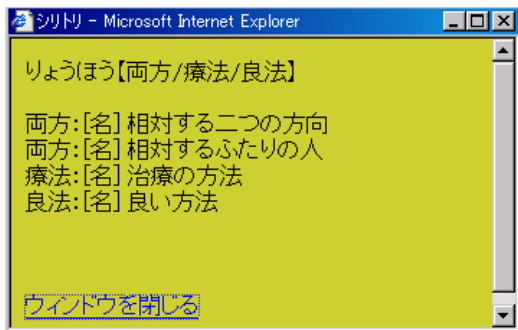


図 4: 意味表示の画面



図 5: ヒントの画面

をクリックすると、図 4 のように「両方 / 療法 / 良法」の単語の意味が表示される。

また、このシリトリシステムでは、ヒントを表示する機能がある。図 3 の状態で「う」から始まる単語を思いつかず、ヒントが欲しくなったとする。その場合、図 3 の「ヒント表示」をクリックすると、図 5 のように「う」から始まる単語のヒントを示す。「うらが」と表示して、「うらが」から始まりあと一文字足すと単語が一つできますよという表示をする。これをみる

と「うらがわ」という単語があると気づきそれを入力するという感じでヒントを利用した回答ができる。

シリトリの勝敗がつくように、計算機側はある程度回答できる単語の数が少なくなった場合にその単語の数に応じて確率的に負けるようにしている。この確率の値を調整することでシリトリのレベル(計算機の強さ)の調整をしている。

人間側は答えるのが疲れて答えなければ負けとなるが、それだけでは面白くないので時間制限の機能も設けている。図 1 で時間制限ありを選択すると人間側は制限時間内に回答をしないといけなくなる。図 6 に時間制限ありの場合の様子を示している。焦って単語を思いつきにくなり人間が負ける場合もこれで生じる。時間制限の時間を調整することでシリトリのレベルの調整をしている。

このようにして計算機が勝ったり、人間が勝ったりできるようにして、人間はシリトリを楽しむことができるようになっている。

また、計算機側の回答速度は、回答できる単語の数が少なくなり計算機側の負けが近づくと計算機側の回答速度が遅くなるように臨場感を持った設定にしている。

3 強力な戦法の利用

本システムでは言語調査 [1] に基づくシリトリのための強力な戦法を利用している。

表 1: 数量的調査の結果

文字	選好 1	選好 2	選好 3	選好 4	選好 5	選好 6
る	21.50	2	0.00	0	22	6
う	12.20	61	0.10	5	41	11
く	6.46	54	0.08	2	39	20
ず	4.50	4	0.33	1	11	5
い	4.10	126	0.12	19	43	18
つ	3.41	56	0.07	0	37	15
ら	3.21	14	0.18	1	21	10
り	2.68	50	0.23	3	38	19
や	2.18	33	0.15	4	26	8
ぶ	2.10	10	0.09	0	13	0
の	1.85	20	0.13	2	20	3
れ	1.53	30	0.09	0	21	12
ろ	1.50	26	0.21	2	21	6
み	1.45	51	0.15	7	26	11
び	1.42	19	0.21	2	16	7
ぎ	1.35	23	0.21	4	17	6
き	1.25	162	0.14	13	39	23
ば	1.22	37	0.14	3	22	8
ば	1.20	5	0.00	1	4	0
ば	1.11	9	0.25	2	5	0
ぐ	1.00	13	0.28	0	9	3
ち	1.00	95	0.17	4	33	19
ゆ	0.97	37	0.10	1	21	4
む	0.97	29	0.06	1	14	0
と	0.96	101	0.11	3	30	11
ば	0.94	35	0.24	2	20	7
め	0.92	36	0.08	4	23	10
え	0.91	43	0.17	3	22	9
わ	0.85	39	0.05	3	20	7
す	0.74	61	0.18	3	23	7
よ	0.70	74	0.14	3	17	3
ま	0.66	61	0.06	2	23	12
ど	0.63	57	0.10	1	22	7
ざ	0.62	13	0.13	1	5	2
ぶ	0.60	48	0.16	3	13	3
ね	0.58	31	0.16	2	12	10
け	0.56	82	0.22	2	30	9
て	0.56	66	0.14	2	21	6
じ	0.47	158	0.16	9	27	13
で	0.46	24	0.11	1	6	4
べ	0.45	11	0.21	0	5	2
し	0.44	347	0.17	17	38	18
た	0.42	119	0.10	2	23	7
も	0.42	31	0.16	1	7	5
な	0.42	55	0.11	2	16	10
げ	0.39	46	0.15	0	10	4
か	0.34	267	0.14	10	30	10
だ	0.33	42	0.14	1	10	5
が	0.33	36	0.18	1	9	1
ふ	0.31	72	0.16	5	12	5
さ	0.29	85	0.18	1	16	5
ば	0.27	26	0.16	1	5	3
に	0.26	46	0.18	2	8	5
ぬ	0.25	8	0.00	0	2	2
こ	0.25	182	0.16	7	17	8
ひ	0.21	92	0.07	7	12	2
び	0.20	5	0.00	0	0	0
そ	0.19	75	0.11	1	11	6
せ	0.12	128	0.19	2	10	4
お	0.12	100	0.05	1	9	4
ぜ	0.12	26	0.30	0	2	1
ぞ	0.07	15	0.21	0	1	1
あ	0.06	108	0.04	2	4	0
ほ	0.04	68	0.23	0	3	0
は	0.04	119	0.17	4	2	2
へ	0.00	22	0.27	0	0	0
べ	0.00	2	0.50	0	0	0

(表の上から「ち」の行までの文字が制約 1 を満足している)

下記の制約と選好を調査したその結果を表 1 に示す。調査には分類語彙表で“*”の印がついている単語を用いた。これは、現代雑誌九十種の語彙調査において標本使用度数が 7 以上であった語である。この調査結果で制約 1 を満足し各選好の値が大きい文字で始まる単語を利用して計算機側は回答する。

- 制約 1 ($END > BEGIN$)
その文字から始まる単語の数よりもその文字で終わる単語の数の方が多ければ攻めが続きやすい。(END はその文字で終わる単語の個数。BEGIN はその文字で始まる単語の個数。)
- 選好 1 ($\max END/BEGIN (END \gg BEGIN)$)
その文字から始まる単語の数に比べてその文字で終わる単語の数の方が比率的に多い文字である方が攻めが続きやすい。
- 選好 2 ($\min BEGIN$)
その文字から始まる単語の数が小さい文字で攻めるほど、相手は回答に困り勝ちやすい。(文献 [2] のシステムでは、この数に従って言う単語を決定している。)
- 選好 3 ($\max NN/(BEGIN + NN)$)
その文字から始まる単語群の総数に対して、「ん」で終わる単語の数が大きい文字で攻めるほど、相手は「ん」で終わる単語をいいやすく、相手は反則負けしやすく、勝ちやすい。(NN はその文字から始まり「ん」で終わる単語の個数。)
- 選好 4 ($\min SAME$)
はじめりとおわりが同じ文字である単語が多いときのそのはじめりとおわりの文字で攻める場合、この種の単語により回答され、切り返しを受けやすく攻めが続きにくくなる。(SAME はその文字から始まりその文字で終わる単語の個数。)
- 選好 5 ($\max BKIND$)
いろいろな文字ではじまる単語で攻めることができる文字ほど攻めが続きやすい。(BKIND はその文字で終わる単語の集合において始まりになりうる文字の種類の数。)

- 選好6 ($\max BKINDC2$)

特に2文字の単語であると想起がしやすいので、いろいろな文字ではじまる2文字の単語で攻めることができる文字ほど攻めやすい。(BKINDC2はその文字で終わる2文字の単語の集合において始まりになりうる文字の種類の数)

4 連想的な受け答え

このシリトリシステムには、単語と単語の連想度を、大規模テキストでのそれら単語の共起頻度などを利用して求めて、相手の単語と連想度が高い単語をなるべく使う機能がある[3]。この機能により、シリトリシステムが人間に似た連想的な受け答えを行えるようになる。

数式では以下のように表される。

$$S = kQ \times (1 - k)R \quad (1)$$

ただし、 k は0から1の係数、 Q は前節のシリトリの戦法としての良さ、 R は単語の連想度、 S は単語の選択のための全体としての良さを意味する。計算機側は S の高い単語を選択して出力する。

5 計算機側の強さの調節

シリトリの勝敗がつくように、計算機側はある程度回答できる単語の数が少なくなった割合に応じて確率的に負けるようにしている。確率 p を定義し、計算機側は優先度の高い回答候補から順に p の確率でその回答候補を回答し、 $(1-p)$ の確率で次の回答候補を p の確率でその回答候補を回答し、 $(1-p)$ の確率でまたその次の回答候補の回答を考慮するといった処理を繰り返して出力する。 n 個回答できる単語がある場合、それらすべてを回答しない場合は $(1 - p)^n$ であり、この確率で計算機は負けることになる。

6 おわりに

本稿では、われわれが開発したシリトリシステムについて紹介した。このシステムは、Web上で動作するシステムとして構築した。Web経由でつながるとこ

ろならどこでも、このシリトリシステムを利用することができる。このシリトリシステムは、人間と計算機のシリトリの対戦が可能である。計算機の強さを調節することが可能である。また、連想的なシリトリが可能である。例えば、直前の回答に意味的に関連のある単語を利用して答える。言語処理分析技術を駆使した強力な戦法も利用可能や、ヒントを表示する機能を持つ。また、単語の意味を表示する機能もある。Windows, UNIXのいずれでも動作する。

シリトリシステムは、日本語単語の教育において役立つと思われる。幼少期の日本人が日本語の単語を覚えるためにシリトリをする場面と、外国人が日本語の単語を覚えるためにシリトリをする場面の二つがある。実際にこれらの補助を目指したシリトリゲームの開発もなされている[2]。また、幼少期の日本人や外国人以外の人でも、専門用語の勉強をシリトリを通じて行うことで専門用語を楽しく学ぶことができる可能性がある。本システムはシリトリに使う単語のジャンルを指定する機能もあり、このような専門用語の学習のためにも利用できるものである。

本稿で示した技術は特許登録、また、出願中であり、権利を保有している[3]。共同開発、事業を進める企業を募集している。<http://www2.nict.go.jp/x/x161/member/murata/technique/technique.html>を参照のこと。

参考文献

- [1] 村田真樹, 井佐原均, シリトリに関する数量的調査, 言語処理学会 第8回年次大会, (2002).
- [2] Tiemi Chiristtine Sakata and Masahide Sugiyama, Internet shiritori using java, 情報処理学会第59回全国大会予稿集, 5N-10, (1999).
- [3] 村田真樹, 井佐原均, 言語獲得支援装置、言語獲得支援プログラム、及び言語獲得支援方法, 特許3733425, (2005).