

## shWiiFit Reduce Dependency Parsing

浅原 正幸 (奈良先端大) swfrdp+NLP2010@gmail.com

## 1 Introduction

本稿では日本語係り受け解析を題材としたゲームとその応用について紹介する。2004年に Sassano [7] により提案された係り受け解析手法は、Stack と Queue を準備し、Shift と Reduce の 2 つの操作を逐次的・決定的に選択することにより解析を行う。この操作をゲームソフトウェア『Nintendo Wii Fit』に同梱されている『バランス Wii ボード』により人間に行わせ、精度・解析速度を競わせるゲーム『shWiiFit Reduce Dependency Parsing』を開発した。応用として本ゲームを用いた心理言語実験による文の係り受け解析難易度の評価手法を提案する。Tokimoto [8] の実験で用いられたガーデンパス文を題材として、人間にとっての難易度の評価を行い、各種係り受け解析器の解析結果と比較した。その結果を報告する。

## 2 Japanese Dependency Parsing

本稿で取り上げるゲームは日本語の係り受け解析を題材とする。日本語においては文節単位で係り受け解析することが多い。日本語文中において、文節単位にはほとんどの場合主辞が修飾語より右にあり、係り受け関係は基本的に交差ししないという強い制約を持つ。また、格要素の省略が多く、かき混ぜが多くおきるという特性を持つ。

係り受け解析の手法として、大きく分けて遷移ベースの手法 [6] とグラフベースの手法 [5] が提案されている。日本語には上記のような係り受け構造の特性があり、単純な (1st-order) グラフベースの手法はよい精度を達成できず [1]、 $O(n^2)$  の Cascaded Chunking 法 [3]、 $O(n)$  の Shift Reduce (SR) 法 [7]、 $O(n^2)$  の トーナメント法 [1] など、遷移ベースの手法が広く用いられている (ここで  $n$  は文中の文節数)。他の枠組の手法として Relative Preference 法 [9] がある。

作成したゲームは文節数に対して線形回数 of アクションを決定的に逐次的に実行することで日本語文の係り受け解析を行う SR 法を用いる (Algorithm 1)。SR 法では、Stack  $S$  と Queue  $Q$  と係り受け関係の集合  $A$  の 3 つ組を用いこの 3 つ組の状態を遷移させながら解析を行う。まず、 $S$  を空列、 $Q$  を入力文節列  $W$ 、 $A$  を空集合により初期化する。 $S$  が空列の場合、 $Q$  の先頭要素  $q$  を  $S$  に移動させる default shift 操作を自動的に行う。 $Q$  が 1 要素  $q$  のみからなる場合、 $S$  の先頭要素  $s$  を  $q$  に係ける default reduce 操作を自動的に行う。 $S$  の先頭要素  $s$  と  $Q$  の先頭要素  $q$  が定義されている場合、 $\langle s, q \rangle$  間の係り受け関係の有無を判別し、係り受け関係が

## Algorithm 1 SR 法による日本語係り受け解析

```

% Initialization
 $\langle S, Q, A \rangle = \langle nil, W, \phi \rangle$ 
repeat
  if  $S == nil$  then
    % "default shift"
     $\langle nil, q|Q, A \rangle \Rightarrow \langle q, Q, A \rangle$ 
  else if  $|Q| == 1$  then
    % "default reduce"
     $\langle s|S, q, A \rangle \Rightarrow \langle S, q, A \cup (s, q) \rangle$ 
  else
    % Judge on  $\langle s|S, q|Q, A \rangle$ 
    if  $s$  and  $q$  has a dependency relation then
      % "REDUCE"
       $\langle s|S, q|Q, A \rangle \Rightarrow \langle S, q|Q, A \cup (s, q) \rangle$ 
    else
      % "SHIFT"
       $\langle s|S, q|Q, A \rangle \Rightarrow \langle q|s|S, Q, A \rangle$ 
    end if
  end if
until  $S == nil$  and  $|Q| == 1$ 
return  $A$ 

```

ある場合には REDUCE 操作を、係り受け関係がない場合には SHIFT 操作を行う。係り受け関係の有無の判別を、一般の係り受け解析器では機械学習器により行うが、本ゲームでは人間により行う。

## 3 Garden Path Sentences

本ゲームの用途として、文の係り受け解析難易度の定量評価を考える。評価対象として心理言語学でよく調査研究されているガーデンパス文を用いる。

ガーデンパス文とは、部分的に複数に解釈可能な構造を持ち、人間により解釈される際に、誤った構造にとられやすく、かつ一度誤った構造にとられると正しい構造の解釈が困難な文のことをいう。関係節を含む文・埋め込み文・同音異義語を含む文など様々なガーデンパス文が心理言語学の研究者によって提示され、自己ペースリーディング法・QA法・注視点計測法により検証されている [4, 8]。

本稿では Tokimoto [8] が用いた埋め込み文に基づく 3 種類の例文を用い、相対的難易度を係り受け正解率および反応速度により評価する。例文は 6 文節からなる以下のようなものである：

NP-が/NP-を/V-た/NP-に/\*V-た

NP が名詞句、V が動詞を表す。1-4,6 文節目の文節

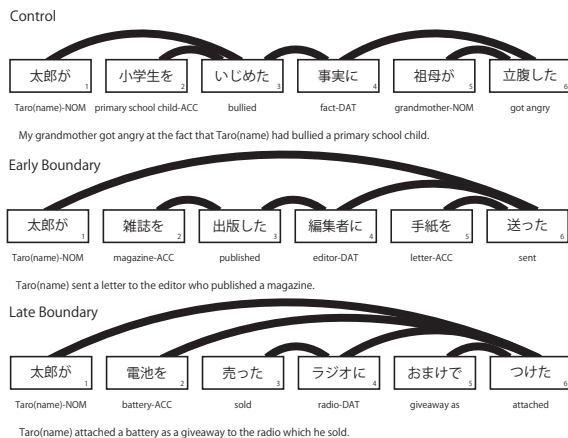


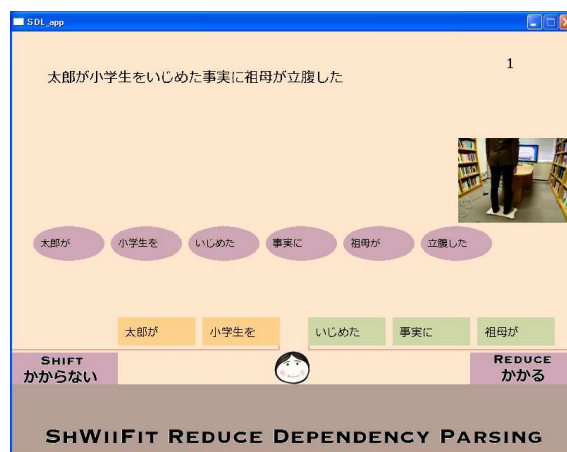
図 1: Example Sentences

は同じ形であるが、5 文節目に異なる形のものが入ることによりそれぞれ異なる係り受け構造を持つ。5 文節目が「NP-が」のものを Control(CTRL)、「NP-を」のものを Early Boundary(EB)、その他のものを Late Boundary(LB) と呼び、それぞれ図 1 に示す係り受け構造を持つ。Tokimoto [8] はこの例文を用い、インクリメンタルに文節を表示する自己ペースリーディング法と QA 法を用いて日本語文の再解析に要するコストを定量評価した。彼の実験において、各文の相対再解析コストが CTRL 文 < EB 文 < LB 文であることとともに、人によって再解析コストが異なることが示されている。

#### 4 shWiiFit Reduce Dependency Parsing

本ゲームは SR 法に基づき WiiFit に同梱されているバランス Wii ボードにより操作することにより日本語係り受け解析を行う。今回はバランス Wii ボードを操作に用いたが、ソフトウェア自体はジョイスティック・ゲームパッド (Wii リモコンを含む)・カーソルキーでも操作が可能である。

ゲームの画面を図 2 に示す。下部に記載されている URL をブラウザで表示するか画像をクリックすることにより動画を見ることができる。画面中段にある左トレイが Stack *S* を表し、右トレイが Queue *Q* を表す。それぞれ先頭要素が画面中央にあるとし、画面中央の 2 つの文節間に係り受け関係があるかを人間が判定する。画面下段にあるアイコンをバランス Wii ボード上での左右への体重加重により操作する。アイコンを左に移動すると [SHIFT] と書かれている壁にタッチすると「かからない」をアイコンを右に移動すると [REDUCE] と書かれている壁にタッチすると「かかる」を意味する。操作者に対する参考情報として文全体と途中まで決定した係り受け木を表示する。default shift・default right の最中および SHIFT・REDUCE 判定直後の画面遷移には



<http://bit.ly/swfrdpNLP2010Fig2>

図 2: shWiiFit Reduce Dependency Parsing

820-860ms のアニメーションが提示され、この間操作者が何もしなければ(重心を中央に保てば)アイコンは自動的に中央に戻る仕様になっている。1 文解析後、正しければ「OK」を、誤っていれば「NG」表示する。操作者はジャンプをすることで次の文へ移ることができる。バックエンドでは文の正解・不正解とともに各アクションの反応速度をはかることができる。

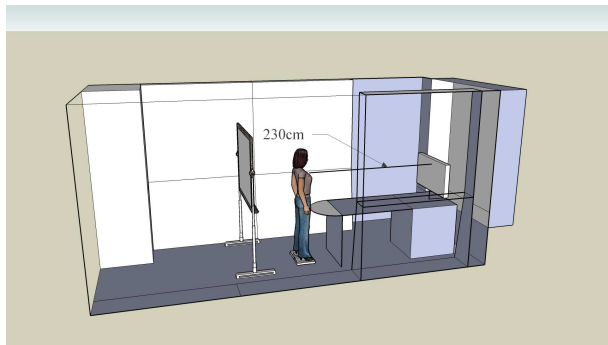
### 5 Experiments

#### 5.1 Experimental Settings

本ゲームを用いて係り受け解析難易度が定量的に評価できるかを検証する。検証のために、3 節に示した CTRL 文・EB 文・LB 文の 3 種類の例文を用いる。本試験用のデータとして Tokimoto [8] が実験に用いたものと同じ 3 種類の例文 10 文ずつと、今回我々で作成したフィラー (Filler) 文 30 文、あわせて 60 文を準備した。この 60 文のうち 40 文を、(1) フィラー文 5 文 → (2) フィラー文 15 文 + CTRL 文 5 文 + EB 文 5 文 + LB 文 5 文 = 混成 30 文 → (3) フィラー文 5 文の順に提示する。1 被験者が同じ文を 2 度解析することはなく、文の提示順も被験者毎に異なる。

各被験者は操作になれるために、約 5 分間の教示を行うとともに、本試験用のデータに基づき実験を行う前に、練習用データ 10 文 1 セットを 1-3 セット行う。教示では、小学 4 年の国語の教科書に掲載されている文節係り受けの解説 (2 pages) と、ゲームの操作方法および練習用データに含まれる係り受け構造の例を示した解説 (6 pages) を提示する。練習用データのセット数は被験者の意志で自由にセット数が設定する。平均練習文数は 15.7 文であった。

図 3 に実験環境を示す。42inch のプラズマモニターに 1024x768 解像度で表示した画面中、中央 800x600 ピクセルの領域をゲーム画面として提示した。バラ



<http://bit.ly/swfrdpNLP2010Fig3>

図 3: 実験環境

表 1: 人の文正解率と文反応時間

		Filler	CTRL	EB	LB
acc.	ave.	73	69	85	49
	stdev.	20.6	35.0	18.0	30.1
s.r.r.t.	ave.	-0.13	0.05	0.06	0.80
	stdev.	0.12	0.70	0.48	0.40

ンス Wii ボードとモニターの距離は 230cm、それ以外の部屋の状況は図 3 の 3D オブジェクトを計測することによって得られる。

## 5.2 Results and Discussions

以下実験結果を示す。被験者は大学院学生 16 人中、比較対象の 30 文を全問誤答とした 2 人と視力不足のために距離を変更した 3 人のデータを排除した 11 人分を用いる。尚、5 人分のデータを排除しなくても結果の傾向に変化はない。

表 1 に人の文正解率と文反応速度を示す。文正解率 (表中 “acc. ave.”) は各文タイプ毎平均を用いる。文反応時間の計測には、人が正解したデータのみに対して、いくつかの評価すべきでない要素を線形回帰により排除した内的にスチューデント化した残差の平均 (表中 “s.r.r.t. ave.”) を用いる。排除した要素はモーラ数・文字数・文節数・提示順・default shift 数・default right 数・SHIFT (左体重加重) 数・REDUCE (右体重加重) 数・SHIFT-REDUCE 間の遷移 (左右の体重移動) 数の 9 つ。

人の文正解率において LB 文が突出して正解率が悪かった。また EB 文は CTRL 文およびフィルター文と比較して有意に正解率がよかった。これは最初に全文 (および文の長さ) が確認可能なために、EB 文における 1 文節目の「NP-が」を最後 (6 文節目) の動詞に係けるというバイアスが強く働いたと考える。実際、CTRL 文の誤りにおいても、このバイアスにより誤った解析を行った例が多かった。これに

対し、LB 文の 2 文節目の「NP-を」は、最後 (6 文節目) の動詞ではなく隣の 3 文節目の動詞に係けるというバイアスが働いたと考える。

人の文反応時間において LB 文は他の文に比べて正解するために多くの時間を要することがわかった。これに対し CTRL 文と EB 文との間には文反応時間において統計的に有意な差はなかった。

上記は全体的な傾向を分析したものであり、人によっては EB 文/LB 文を全て正解するのに対し、CTRL 文を誤答する者もいた。

表 2 に人のアクション毎の反応時間を示す。表中文タイプ毎に 3 行表示しているうち、1 行目は当該文を解析に必要なアクションを意味し、s: default shift, r: default right, S: SHIFT, R: REDUCE を意味する。2 行目は正解したデータのみに対して、評価すべきでない要素を線形回帰により排除した残差 (表中 std. residual ave.) を文タイプ毎に集計し、人単位でマクロ平均をとったものである。排除した要素はモーラ数・文字数・提示順数の 3 つ。前の表で排除したその他の要素 6 つは文タイプ毎に全く同じ値を持つため。3 行目は同残差の標準偏差。4 行目は正解したデータのみについての実時間の人単位でのマクロ平均。

各文とも 1 アクション目、2 アクション目に時間を要している。これは Tokimoto の実験のように文をインクリメンタルに提示しているわけではなく、最初に文全体を提示しているために、被験者は一度文全体を読んで確認する傾向があるということがわかる。

表 3 に人と各種解析器の文正解率の対比を示す。解析器の実験設定は以下の通りである。形態素解析器として、JUMAN-5.1 (J5 と略記)・JUMAN-6.0 (J6 と略記)・MeCab-0.98 with IPADIC-2.7.0 (M と略記) を用いる。“KNP-2.0” は KNP-2.0、“KNP-3.0” は KNP-3.0、“CaboCha” は CaboCha-0.60 によるものでそれぞれデフォルトの設定を利用した。“Sassano nK” および “Iwatate nK” は JUMAN-5.1 (J5) の出力に、正しい文節区切りを与えたものに対して、それぞれ SVM に基づく SR 法 [7] およびトーナメント法 [1] による解析器で解析したものである。8K, 34K はそれぞれ京大コーパス 8,000 文, 34,000 文で訓練したことを意味する。なお “Sassano nK” の解析器の構成には、文献 [1] の実装および素性集合を用いたために、文献 [7] の解析器とは厳密には出力が異なるだろう。

全ての解析器は CTRL 文の正答率が高く、LB 文の誤答率が高かった。このことから解析器は名詞句を近い動詞に係けるという強い制約を持っていることがわかる。

EB 文に関しては、KNP とトーナメント法がいくつかの文について正解した。トーナメント法は少ない学習データによるもの (8K) のほうが多い学習

表 2: 人のアクション反応時間

アクション順	0	1	2	3	4	5	6	7	8	9
CTRL	s	S	R	R	s	R	s	S	r	r
std. residual ave.		1.36	-0.1	-0.39		-0.41		-0.45		
std. residual stdev.		0.92	0.72	0.78		0.59		0.45		
real time ave.	0.83	8.85	4.75	3.60	0.85	3.49	0.85	3.43	0.85	0.85
EB	s	S	R	S	R	S	S	r	r	r
std. residual ave.		1.32	0.06	-0.55	-0.40	-0.35	-0.09			
std. residual stdev.		0.84	0.89	0.66	0.36	0.83	0.88			
real time ave.	0.83	8.75	4.82	3.25	3.45	4.12	4.35	0.85	0.85	0.85
LB	s	S	S	R	S	S	r	r	r	r
std. residual ave.		1.28	0.11	-0.71	-0.31	-0.36				
std. residual stdev.		0.93	0.97	0.62	0.64	0.58				
real time ave.	0.83	10.79	6.13	3.01	4.28	4.18	0.85	0.85	0.85	0.85

表 3: 人・解析器の文正解率

	CTRL	EB	LB
人	69	85	49
KNP-2.0 (J5/J6)	100	30	0
KNP-3.0 (J5)	90	100	0
KNP-3.0 (J6)	90	60	40
CaboCha (M)	100	0	0
Sassano 8K (J5)	100	0	0
Sassano 34K (J5)	100	0	0
Iwatate 8K (J5)	100	70	0
Iwatate 34K (J5)	100	10	0

データによるもの(34K)よりよかった。少ない学習データの場合には、1文節目の名詞句と3文節目の動詞の係り受け関係がコーパス中出现しなかったが、1文節目の名詞句と6文節目の動詞の係り受け関係が出現した場合に正解したと考えられる。これに対し、多い学習データの場合には、両方の係り受け関係がともにコーパス中出现し、近い3文節目に係っていると考えられる。

LB文に関しては、格解析モデルを含むKNP-3.0(J6) [2] がいくつか正解した。形態素解析器の出力する意味情報に基づき解析しているために、形態素解析器のバージョンにより解析精度が異なる。

人がCTRL文に比べてEB文をよく正答しているのに対し、SR法(“Sassano nK”)がEB文を全く正答していないことを考えると、人にSR法を強いることによる解析結果の影響は少ないことが伺える。

## 6 Conclusions

本稿ではSR法に基づく日本語係り受け解析を題材としたゲームを紹介した。応用として心理言語実験に適用し、係り受け解析難易度の定量的評価を試験的に行った。心理言語実験において、デバイスとしてバランスWiiボードを用いることにより次のような利点があると考えられる。一般に心理言語実験における作業は単調であり、被験者が途中で作業を怠ることが多々ありこれを統制することは困難である。本ゲームを用いることにより、被験者は実験中重心を中心に保つことを強いられたために怠ることが困難になるとともに、その意識がある程度作動記憶

を妨害する効果もあると考える。

本研究の応用の1つとして、係り受け解析ゲームの国語教育における利用を検討している。我々は2009年7月22-24日の3日間本ゲームを用いた高校生向け模擬授業を実施した。機材の関係で少人数(参加者2名)であったが、今後ゲームパッドなどによる手法に転換し、多人数にも対応可能な模擬授業の実施手法を検討し、継続していきたい。

最後に自然言語処理研究者としての観点から、問題提起をしたい。どのようにして今回提示した文を計算機に解析させればよいだろうか。手法としてCoNLL-2009で実施された格解析と係り受け解析の結合推論が必須である。今回利用した例文は語彙親密度と新聞10年分の頻度情報で統制されており、頻度情報のみで全問正解することは困難であろう。この問題を解決することが我々の次のチャレンジだと考える。

## References

- [1] M. Iwatate, M. Asahara, and Y. Matsumoto. Japanese Dependency Parsing Using a Tournament Model. In *Proc. of COLING-2008*, 2008.
- [2] D. Kawahara and S. Kurohashi. A Fully-Lexicalized Probabilistic Model for Japanese Syntactic and Case Structure Analysis. In *Proc. of HLT-NAACL2006*, 2006.
- [3] T. Kudo and Y. Matsumoto. Japanese Dependency Analysis using Cascaded Chunking. In *Proc. of CoNLL-2002*, 2002.
- [4] R. Mazuka, K. Itoh, and T. Kondo. Processing Down the Garden Path in Japanese: Processing of Sentences with Lexical Homonyms. *Journal of Psycholinguistic Research*, Vol. 26, No. 2, pp. 207-228, 1997.
- [5] R. McDonald, F. Pereira, K. Ribarov, and J. Hajic. Non-Projective Dependency Parsing using Spanning Tree Algorithms. In *Proc. of HLT-EMNLP-2005*, 2005.
- [6] J. Nivre. Deterministic Dependency Parsing of English Text. In *Proc. of COLING-2004*, 2004.
- [7] M. Sassano. Linear-Time Dependency Analysis for Japanese. In *Proc. of COLING-2004*, 2004.
- [8] S. Tokimoto. Reanalysis costs in processing Japanese sentences with complex np structures and homonyms: Individual differences and verbal working memory constraints. Technical Report JCSS-TR-53, Japanese Cognitive Science Society, 2004.
- [9] 工藤拓, 松本裕治. 相対的な係りやすさを考慮した日本語係り受け解析モデル. 情報処理学会論文誌, Vol. 46, No. 4, 2005.