テキスト結束性を考慮した entity grid に基づく 局所的一貫性モデル

横野 光[†]・奥村 学[†]

本論文では entity grid を用いたテキストの局所的な一貫性モデルに対する改善について述べる。entity grid ベースの既存モデルに対して、テキスト結束性に寄与する要素である接続関係、参照表現、語彙的結束性、また、より詳細な構文役割の分類を組み込んだモデルを提案し、その性能を検証する。語彙的結束性に関しては、語彙的連鎖を用いたクラスタリングを行う。テキスト中の文の並びに対して、より一貫性のある文の順番の判定と、人手による評価に基づいた要約テキストの比較の2種類の実験を行い、その結果、本論文で提案する要素が entity grid モデルの性能の改善に寄与することが明らかになった。

キーワード：テキスト一貫性、テキスト結束性、テキストの評価

Local Coherence Model Based on Entity Grid Augmented with Text Cohesion

HIKARU YOKONO[†] and MANABU OKUMURA[†]

This paper describes improvements made to the entity grid local coherence model for Japanese text. We investigate the effectiveness of taking into account cohesive devices, such as conjunction, explicit reference relation, lexical cohesion, and refining syntactic roles for a topic marker in Japanese. To take into account lexical cohesion, we consider lexical chaining. Through the experiments on discrimination where the system has to select the more coherent sentence ordering, and comparison of the system's ranking of automatically created summaries against human judgment based on quality questions, we show that these factors contribute to improve the performance of the entity grid model.

Key Words: *text coherence, text cohesion, text evaluation*

1 はじめに

テキストの評価は、自動要約や機械翻訳などのようなテキストを生成するタスクにおいて手法の評価として用いられるだけでなく、例えば人によって書かれた小論文の自動評価 (Mitsakaki and Kukich 2004) といったように、それ自体を目的とすることもある。言語処理の分野においては前者のような手法評価の観点からテキスト評価に着目することが多く、例えば自動要約の

[†] 東京工業大学精密工学研究所, Precision and Intelligence Laboratory, Tokyo Institute of Technology

評価で広く用いられている ROUGE (Lin and Hovy 2003; Lin 2004) や機械翻訳で用いられている BLEU (Papineni, Roukos, Ward, and Zhu 2002) のような評価尺度が存在している。これらの評価手法は特に内容についての評価に重点が置かれている。つまり、評価対象のテキストが含んでいなければならない情報をどの程度含んでいるかということに焦点が当てられている。

しかし、実際にはテキストは単に必要な情報を含んでいれば良いというわけではない。テキストには読み手が存在し、その読み手がテキストに書かれた内容を正しく理解できなければ、そのテキストは意味をなさない。読み手の理解を阻害する原因には、難解な語彙の使用、不適切な論理展開や文章の構成などが挙げられる。これらはテキストの内容に関する問題ではなく、テキストそのものに関する問題である。従って、テキストの内容が正しく読み手に伝わるかどうかを考慮するならば、その評価においては内容に関する評価だけでなく、テキストそのものについての評価も重要となる。

テキストそのものについての性質のうち、テキスト一貫性(田窪, 西山, 三藤, 亀山, 片桐 2004)とは文章の意味的なまとまりの良さであり、例えば因果関係や文章構造などによって示される文同士の繋がりである。意味的なまとまりが悪ければ、テキストの内容を読み手が正確に理解することが困難になると考えられる。このことから一貫性の評価はテキストの内容が正しく伝わることを保証するために必要であると言える。また、テキスト一貫性が評価できるようになると、テキストを生成するシステムにおいて、例えば、一貫性が良くなるように文章を構成したり、一貫性の観点からの複数の出力候補のランク付けが可能となり、出力するテキストの質を高めることができる。

テキスト一貫性は局所的な一貫性と大域的な一貫性という2種類のレベルに分類できる。局所的な一貫性とは相前後する2文間における一貫性であり、大域的な一貫性とは文章における話題の遷移の一貫性のことである。一貫性の評価に関しては、この局所的な一貫性と大域的な一貫性の両方についてそれぞれ考えることができるが、局所的な一貫性は大域的な一貫性にとって重要な要素であり、局所的な一貫性の評価の精度の向上が大域的な一貫性の評価に影響すると考えられる。

以上のことから、本論文では、テキスト一貫性、特に局所的な一貫性に焦点を当て、この観点からのテキストの評価について述べる。

テキストの性質について、テキスト一貫性と並べて論じられるものにテキスト結束性(Halliday and Hasan 1976)がある。これは意味的なつながりである一貫性とは異なり、文法的なつながりである。一貫性が文脈に依存しているのに対し、結束性は脱文脈的で規則的な性質である(庵 2007)。テキスト結束性に寄与する要素は大きく参照¹、接続、語彙的結束性²に分けられる。こ

¹ 代名詞の使用や省略は参照に含まれる。

² 同じ語の繰り返しは語彙的結束性に含まれる。

れらはテキストの表層において現れる要素である。

一貫性は先に述べたように意味のまとまりの良さであり、これに寄与する要素は明示的な形で現れない。一貫性と結束性はどちらもテキストのまとまりに関する性質であり、それぞれが独立ではなく互いに関係している。従って、テキストの表層に現れる、結束性に関係する要素である接続表現や語彙的結束性を一貫性モデルにおいても考慮することで性能の向上が期待できる。

2章で述べるように、局所的な一貫性に関する研究はテキスト中の隣接する文間の関係を単語の遷移という観点から捉えているものが多い。その中でも Barzilay ら (Barzilay and Lapata 2005, 2008) の研究は、この領域における他の研究において多く採用されている *entity grid* という表現を提案しており、先駆的な研究として注目に値する。しかし、3章で詳述するように、このモデルでは要素の遷移の傾向のみ考慮しており、テキストのまとまりに関係している明示的な特徴はほとんど利用されていない。そこで本論文では4章で詳述するように、一貫性モデルに結束性に関わる要素を組み込むことによって、結束性を考慮に入れた局所的な一貫性モデルを提案する。

2 関連研究

Barzilay ら (Barzilay and Lapata 2005, 2008) は局所的な一貫性のモデルとして *entity grid* を提案している。このモデルはテキスト中で述べられている要素の遷移に着目している。これは、センタリング理論 (Grosz, Weinstein, and Joshi 1995) で示されているように、一貫性のあるテキストではその文中の要素の出現に規則性があるという考えに基づいている。

Elsner ら (Elsner and Charniak 2008) は、*entity grid* モデルがテキストの一貫性において要素の遷移にのみ着目しているということに言及し、例えば参照表現や対象の要素がこれまでに既に述べられている要素かどうかなどといった他の要素をモデルに組み込んでいる。Filippova ら (Filippova and Strube 2007) は *entity grid* モデルに要素間の関係を考慮したモデルを提案し、このモデルをドイツ語の新聞記事に対して適用した結果を報告している。

大域的な一貫性について、Barzilay ら (Barzilay and Lee 2004) は隠れマルコフモデル (HMM) を採用したモデルを提案している。このモデルでは文章中の話題を HMM における隠れ状態と見なし、話題の一貫性を隠れ状態の遷移確率によって表現している。Soricut ら (Soricut and Marcu 2006) や Elsner ら (Elsner, Austerweil, and Charniak 2007) は局所的な一貫性と大域的な一貫性を同時に考慮するモデルをそれぞれ提案している。これらのモデルは *entity grid* モデルと HMM を組み合わせたものである。

日本語の文章に対する一貫性の評価手法には、板倉ら (板倉, 白井, 黒岩, 小高, 小倉 2008) の提案する段落の一貫性指標がある。これは段落内で用いられている単語間の意味的な関係に

基づいている.

これらの手法では文書中の単語の出現に着目してテキスト一貫性を評価しており, 1 章で述べたように一貫性に影響すると考えられる, テキスト結束性に関する表層的な特徴は考慮されていない.

3 Entity Grid に基づく局所的な一貫性モデル

Barzilay ら (Barzilay and Lapata 2008) は, 一貫性のあるテキストではその中で述べられる要素の出現の分布には規則性があるという仮説に基づいた一貫性のモデルを提案している. また, このテキスト中の要素の分布パターンを捉えるために, entity grid と呼ばれる表現を導入している. これは, テキストを, 行に文を, 列に文章中の要素をそれぞれ対応させた行列として表したものである. その各項には文における要素の構文役割が入る. 用いられる構文役割は主語 (S), 目的語 (O), その他 (X), 出現せず (-) の 4 種類である. 図 1 に示すテキストに対応する entity grid を表 1 に示す. この一貫性モデルでは entity grid の作成の際に共参照解析を行い, 異なる表現であっても同じ要素を指すものをまとめている. 例えば, 図 1 において s_1 の “BELL INDUSTRIES Inc.” と s_4 の “Bell” は同じ要素を指すと判定されると, それらの構文役割は表 1 の同じ列 (“BELL”) に記述される.

局所的な一貫性の評価には entity grid を基に作られた文書ベクトルを用いる. ベクトルの要素は文 n-gram($n \geq 2$) における構文役割の遷移確率と, 構文役割の出現確率からなる. 構文役

s_1 [BELL INDUSTRIES Inc.]_S increased [its quarterly]_O to [10 cents]_X from [seven cents]_X [a share]_X.
 s_2 [The new rate]_S will be payable [Feb. 15]_X.
 s_3 [A record date]_S hasn't been set.
 s_4 [Bell]_S, based in [Los Angeles]_X, makes and distributes [electronic, computer and building products]_O.

図 1 テキスト例 (Wall Street Journal から引用)

表 1 図 1 のテキストに対する entity grid

	BELL	quarterly	10 cents	seven cents	share	rate	Feb. 15	date	Los Angeles	products
s_1	S	O	X	X	X	-	-	-	-	-
s_2	-	-	-	-	-	S	X	-	-	-
s_3	-	-	-	-	-	-	-	S	-	-
s_4	S	-	-	-	-	-	-	-	X	O

割の遷移を計算する際には、文書始めと文書終わりを含んだ遷移も考慮する．この確率は entity grid 中に存在する長さ n の全ての構文役割の遷移の数に対する対象の長さ n の遷移列の割合である．例えば、表 1 に示す entity grid において、[S-] という遷移の確率は、長さ 2 の全ての遷移の数（即ち、50）に対する対象の遷移列（即ち、3）の割合から求められる（即ち、0.06）．

また、テキストにおいて頻出する要素は、そのテキストの話題に関連すると考えられ、そのような要素はそうでない要素とは異なる遷移の傾向を持つという仮定から、出現頻度によってテキスト中の要素をグループに分け、それぞれのグループにおいて entity grid を作成している．連続する 2 文における遷移確率で構成された文書ベクトル d_1 , d_2 を図 2 に示す．図中の “0”, “1” はそれぞれ文書始め、文書終わりを表す．

Barzilay らは構文役割を dependency parser を使って推定しているが、これに対して本論文では構文役割を格助詞によって決定する（表 2）．また、文書ベクトルの作成において、考慮する構文役割の遷移は Barzilay らの手法と同様に文 3-gram までとした．

テキスト一貫性は、対象のテキストについて一貫性がある、一貫性がないといったように絶対的に評価することが困難であるため、Barzilay らはテキストの文の順番を局所的一貫性に基づいて順位付けすることで相対的に評価するモデルを提案している．テキストの順位付けにはスコア関数を導入し、その値を利用する．このスコア関数は、あるテキスト d_i の文の順番を並べ替えて生成したテキストを x_{ij} , x_{ik} とし、 x_{ij} の方が x_{ik} よりも一貫性があるとしたとき、

$$\mathbf{w} \cdot \Phi(x_{ij}) > \mathbf{w} \cdot \Phi(x_{ik})$$

という条件を満たすような関数である．この $\mathbf{w}$ の値は学習によって推定する．このパラメータの学習には ranking SVM が用いられる． $\Phi(x)$ はテキスト x の一貫性に関する性質を表す素性ベクトルであり、具体的には上述のテキスト中の要素の遷移確率で構成された文書ベクトルで

	SS	SO	SX	S-	OS	OO	OX	O-	XS	XO	XX	X-	-S	-O	-X	OS	OO	OX	O-	S1	O1	X1	-1	--
d_1	0	0	0	.06	0	0	0	.02	0	0	0	.08	.06	.02	.04	.02	.02	.06	.10	.02	.01	.02	.14	.32
d_2	.02	0	0	.04	0	0	0	.02	0	0	0	.08	.04	.04	.10	.02	0	0	.18	.02	.02	.02	.14	.26

図 2 長さ 2 の遷移列に基づく文書ベクトル

表 2 構文役割

構文役割	対応する助詞
主語 (S)	“が”, “は”
目的語 (O)	“を”, “に”
その他 (X)	その他の格助詞
出現せず (-)	文中に出現しない

ある。

モデルの評価は、テキストの文の順番を決定するというタスクを順位付け問題として定式化して行う。即ち、テキストを文の集合と見なし、それから生成できるテキストの順位付けを行う。元のテキストの順番で構成されたテキストの順位が最も高ければ、そのテキストに対する局所的な一貫性の良し悪しを正しく評価できたと見なす。しかし、実際には、同じテキストから生成された異なる文の順番を持つテキストの一貫性を比較することは困難であるため、元テキストとそのテキストから生成された異なる順番を持つテキストのペアに対して比較を行い、どの程度元のテキストの方に高い順位を割り当てることができているかで評価する。

この一貫性モデルは、テキスト中の要素の遷移列のみに着目している。参照表現に関しては共参照解析を行い、異なる表現であっても同じ要素を指すものは同一の要素として扱っているが、接続表現や同義語、類義語といったテキストのまとまりに関係しているその他の明示的な特徴は利用されていない。

4 テキスト結束性に関わる要素と構文役割の拡張

本論文では、Barzilay ら (Barzilay and Lapata 2008) の局所的な一貫性モデルに対して、テキスト結束性に寄与する、テキストに表層的に現れる文法的要素を考慮することで、その性能の向上を図る。これにより、既存手法では“正しい一貫性を持つテキストの性質”に着目したモデルを構築していたのに対し、本手法では“正しい結束性を持ち、且つ、一貫性を保っているテキストの性質”に着目したモデルを構築できると考えられる。具体的には、1章で述べたテキスト結束性に寄与する要素を、それぞれ、文書ベクトルへの素性の追加、接続関係毎の遷移確率の計算、意味的な類似性に基づく文中の要素のクラスタリングという形で、局所的な一貫性モデルに組み込む。

また、構文役割について、日本語の主題表現を考慮に入れた拡張を行う。

4.1 接続関係毎の遷移確率の計算

本論文では文の展開において接続関係の種類毎に文中の要素の構文役割の遷移の傾向が異なるという仮説を立てる。例えば、ある時点までで主題として述べられていた要素は、話題転換後には全く出現しない、あるいは別の主題の補助的な役割として出現するということが考えられる。この仮説による特徴を捉えるために、文の接続関係毎に遷移確率を計算する。

文間の関係の推定には接続表現を利用する。テキスト中の隣接する文に対して、後ろの文の接続詞の種類によって文間の関係を決定する。

接続関係には市川のカテゴリ (市川 1978) を採用し、これに基づいて接続詞を表3に示すグループに分類する。表中の括弧内の数字はそのグループに属する接続詞の数である。各グループへの

接続詞の対応付けは人手で行った．接続詞が存在しない場合はその2文間には連鎖型の接続関係があると見なす．

文間の関係の種類毎に遷移確率を計算するため，ベクトルの素性の数は接続関係の数に比例して増加する．このことはデータのスパース性を導くと考えられるので，さらにこの8種類の分類を表4に示す文脈形成の関係に基づく3種類のグループにまとめる．

図3に示すテキストとそのentity grid (表5) から生成される長さ2の構文役割の遷移確率のベクトルの例を示す．ここで $s_1, \dots, s_4$ は文であり， $e_1, \dots, e_3$ は文中の要素，図中の下付き文字はその要素の文中での構文役割である．

文 s_2 の文頭に接続詞“そして”があり，この接続詞は添加型に属するため， s_1 と s_2 間の関係は Group 2 となる． s_2 と s_3 間， s_3 と s_4 間に関しては s_3 と s_4 の文頭に接続詞が存在しないの

表 3 接続関係の分類

関係	説明	例
順接型 (35)	前文の内容を条件とするその帰結を後文に述べる型	従って
逆接型 (33)	前文の内容に反する内容を後文に述べる型	しかし
添加型 (41)	前文の内容に付け加わる内容を後文に述べる型	そして
対比型 (17)	前文の内容に対して対比的な内容を後文に述べる型	または
転換型 (19)	前文の内容から転じて，別個の内容を後文に述べる型	さて
同列型 (11)	前文の内容と同等と見なされる内容を後文に重ねて述べる型	つまり
補足型 (1)	前文の内容を補足する内容を後文に述べる型	なぜなら
連鎖型 (1)	前文の内容に直接結びつく内容を後文に述べる型	そうして

表 4 文脈形成の観点に基づく接続関係の分類

関係	対応する接続関係	説明
Group 1 (論理的結合関係)	順接型，逆接型	二つの事柄を論理的に結びつけて述べる関係
Group 2 (多角的連続関係)	添加型，対比型，転換型	二つ（以上）の事柄を別々に述べる関係
Group 3 (拡充的合成関係)	同列型，補足型，連鎖型	一つの事柄に対して拡充して述べる関係

$s_1.$ $[e_1]_S \dots [e_2]_O \dots$
 $s_2.$ **そして**, $[e_2]_S \dots [e_3]_O \dots$
 $s_3.$ $[e_3]_S \dots [e_1]_X \dots$
 $s_4.$ $[e_4]_S \dots$

図 3 テキスト例 2

表 5 図3に対する entity grid

	e_1	e_2	e_3	e_4
s_1	S	O	-	-
s_2	-	S	O	-
s_3	X	-	S	-
s_4	-	-	-	S

$$\begin{array}{cccccccccccccccc} SS_{G_1} & \cdots & --G_1 & SS_{G_2} & \cdots & S-G_2 & \cdots & --G_2 & SS_{G_3} & \cdots & S-G_3 & \cdots & --G_3 \\ 0 & \cdots & 0 & 0 & \cdots & .25 & \cdots & .25 & 0 & \cdots & .06 & \cdots & .13 \end{array}$$

図 4 接続関係を考慮した文書ベクトルの例 (一部)

で連鎖型と見なし Group 3 となる. 遷移確率の計算は各関係毎に行う. 例えば, Group 3 の [S-] という構文役割の遷移確率は, Group 3 の長さ 2 の全ての遷移の数 (即ち, 16) に対する [S-] の遷移の数 (即ち, 1) の割合である 0.06 となる. 図 3 のベクトルを図 4 に示す. SS_{G_i} は Group i における構文役割の遷移 [SS] の遷移確率を表す.

4.2 参照表現

参照表現に関して, 本論文ではその先行詞が明示的である指示表現のみを考慮する. “この” や “あの” などの指示形容詞が使われている指示表現はその先行詞が前文に現れることが多い. 従って, 逆に指示形容詞が出現している文の前文にその先行詞が出現していなければ, その 2 文間のつながりは悪いと考えることができる.

このことを考慮するために, 指示形容詞を含む参照表現がその先行詞を前文に含む割合を文書ベクトルの素性として追加する. 対象とする指示形容詞は “この” など 8 種類で, 割合は “指示形容詞+名詞” という表現の出現数に対する, その先行詞が前文に現れている場合の数で求める. これは参照表現が正しく機能している割合と見なすことができる.

Barzilay らの手法では grid を作成する際に共参照解析を行い, 異なる表現であっても同じ要素を指す場合は同一要素と見なしている. これに対して, 本論文では共参照解析は行っていない.

4.3 語彙的結束性に基づいた文中の要素のクラスタリング

Barzilay らの entity grid モデルでは遷移確率を計算する際にそれぞれの要素を独立に扱っている. そのために要素間の関係はモデルに反映されていない. この問題に対して, 各要素を意味的なクラスタリングによってまとめ, 得られたクラスタを 1 つの要素として扱うことで対応する. 本論文ではクラスタリング手法として日本語語彙大系 (池原, 宮崎, 白井, 横尾, 中岩, 小倉, 大山, 林 1997) の意味体系を利用した手法と語彙的連鎖を利用した手法の 2 種類を考える.

要素をクラスタにまとめた際, 同じクラスタにまとめられる要素が 1 文中に複数存在すると, そのクラスタに対して複数の構文役割が存在することになる. このような場合における構文役割の扱いに対して, 次の 2 種類を考える.

手法 1(1st). 構文役割の優先順位³ に基づいて, クラスタに対して 1 つの構文役割を決定する.

³ 優先順位: S>O>X.

図 5 構文役割の選択

手法 2(comb). クラスタ中の構文役割を全て利用し、遷移は全組合せを考える。

図 5 にそれぞれの例を示す。図において、 e_1 , e_2 , e_3 は要素であり、 c_1 は e_2 と e_3 を含むクラスタである。手法 1 では、文 s_1 と s_2 間での c_1 の遷移は [OS] のみとする。これに対して、手法 2 では、構文役割の全ての組合せである [OO], [OS], [-O], [-S] を c_1 の遷移と見なす。

4.3.1 日本語語彙大系を利用したクラスタリング

日本語語彙大系を使って、要素を同じ概念のグループにまとめる。日本語語彙大系は最大 12 段からなる階層的な構造を持つ意味属性の体系を持つシソーラスである。このそれぞれの意味属性をクラスタとして扱う。

テキスト中出现する各要素に対して、その要素が持つ意味属性を日本語語彙大系から検索する。このうち特定の段において同じ意味属性を有する要素を同じクラスタにまとめる。1 つの要素に対して複数の意味属性が存在する場合は、その特定の段の意味属性で見た時に数の多かった意味属性をその要素の意味属性と見なす。

4.3.2 語彙的連鎖を利用したクラスタリング

語彙的連鎖 (Morris and Hirst 1991) とは意味的に関連している語の列である。この語の列をクラスタとして扱う。本論文では Mochizuki らの手法 (Mochizuki, Iwayama, and Okumura 2000) に基づいて語彙的連鎖を求める。

はじめに、単語 X , Y 間の共起スコアはコサイン尺度 (1) によって求める。

$$\cos(X, Y) = \frac{\sum_{i=1}^n x_i \times y_i}{\sqrt{\sum_{i=1}^n x_i^2} \times \sqrt{\sum_{i=1}^n y_i^2}} \quad (1)$$

ここで x_i , y_i はテキスト i において単語 X , Y が出現する回数であり、 n はコーパス中のテキストの数である。次に、2 つのクラスタ C_i , C_j 間の類似度は式 (2) によって求める。

$$\text{sim}(C_i, C_j) = \max_{X, Y} \cos(X \in C_i, Y \in C_j) \quad (2)$$

表 6 構文役割 (拡張)

構文役割	該当する助詞
主語 (S)	“が”
目的語 (O)	“を”, “に”
主題 (H)	“は”
述部要素 (R)	述部に係る他の助詞
その他 (X)	上記以外の助詞
出現せず (-)	文中に出現しない

語彙的連鎖の生成アルゴリズムは1文中の要素のクラスタリングとテキスト全体の要素のクラスタリングの2ステップからなり、テキスト中の全ての文に対してこのステップを繰り返す。1文中の要素のクラスタリングでは、まずテキストから文を取り出し、その文中のそれぞれの要素を1クラスタと見なして、全てのクラスタのペアに対して式(2)によって類似度を計算する。類似度の最も高いペアの類似度が閾値以上であれば、そのペアをマージする。この処理をマージするペアが無くなるまで繰り返す。次に、テキスト全体のクラスタと先ほど生成した文中のクラスタの全てのペアの類似度を同様に式(2)によって計算する。類似度の最も高いペアの類似度が閾値以上であれば、そのペアをマージする。この処理をマージするペアが無くなるまで繰り返す。

4.4 構文役割の拡張

Barzilay ら (Barzilay and Lapata 2008) の entity grid モデルではテキストにおいて顕著な使われ方をする要素をそうでない要素と分けて遷移確率を求めている。これは顕著に表れる要素はテキストの主題を表すことが多く、そのような要素は特別な遷移傾向を持つという仮説に基づいている。主題に関しては、日本語の文章では助詞“は”を用いることでその文の主題を明示的に表すことができる。また、主題かどうかというだけでなく、述部に直接係る要素とそうでない要素では文章の展開への寄与が異なると考えられる。

これらのことをモデルに組み込むために、本論文では Barzilay らの entity grid モデルで用いられている4種類の構文役割を表6に示す主題と述部要素を加えた6種類に拡張する。この構文役割の集合においては、その役割間の優先順位関係を $H > S > O > R > X$ とする (cf. (Walker, Iida, and Cote 1994))。

5 実験と考察

前章で述べた各要素の評価のために2種類の実験を行った。1つは文順序に関するタスクであり、もう1つは自動要約で生成された要約テキストのランキングのタスクである。

表 7 検討するモデル

モデル	説明	次元数	
		no CONJ	+CONJ
Baseline	(Barzilay and Lapata 2008) モデル	252	–
CONJ のみ	接続関係毎の遷移確率の計算	–	731
REF	指示表現に関する素性の追加	253	734
SC(1st)	日本語語彙大系によるクラスタリング (1st)	252	731
SC(comb)	日本語語彙大系によるクラスタリング (comb)	252	731
LC(1st)	語彙的連鎖によるクラスタリング (1st)	252	731
LC(comb)	語彙的連鎖によるクラスタリング (comb)	252	731
SR(H)	構文役割の拡張	688	2031
ALL-SC(1st)	SR(H)+REF+SC(1st)	689	2034
ALL-SC(comb)	SR(H)+REF+SC(all)	689	2034
ALL-LC(1st)	SR(H)+REF+LC(1st)	689	2034
ALL-LC(comb)	SR(H)+REF+LC(comb)	689	2034

実験において性能を評価するモデルと各モデルから生成される文書ベクトルの次元数を表 7 に示す. 各モデルに対して接続関係毎に遷移確率を計算してベクトルを作成した場合 (+CONJ) と, それを行わなかった場合 (no CONJ) の両方で実験を行った. 全てのモデルにおいて頻出する要素とそうでない要素に分けて grid を作成する. その閾値は Barzilay らの手法と同様に 2 とした. Baseline は Barzilay ら (Barzilay and Lapata 2008) の設定に従ったモデルである. 但し, 本論文では Baseline においても共参照解析は行わない.

実験では, 形態素解析には MeCab⁴を, 係り受け解析には CaboCha⁵を使用した.

5.1 予備実験

語彙的結束性の考慮において, 日本語語彙大系を利用したクラスタリングの際に使用する意味属性の段と語彙的連鎖によるクラスタリングの際の閾値をあらかじめ決定する必要がある. 本論文ではこれらの値を予備実験によって決定した.

予備実験は表 7 の SC(1st) と LC(1st) のモデルに対して, 5.2 節と同じタスクを行った. 使用したデータは朝日新聞コーパスの 2003 年の記事のうち, “人もの” というカテゴリに分類されている記事 100 件である. この記事の順番を無作為に並べ替えたものと元の記事を比べて元の記事の方が一貫性があると判定したペアを正解と見なし, その精度が最も良かった値を使用するパラメータとした.

日本語語彙大系で用いられている意味属性体系は最大で 12 段あり 1 が最も抽象的な概念であ

⁴ <http://mecab.sourceforge.net/>

⁵ <http://www.chasen.org/~taku/software/cabocha/>

る。このうち3～9の段に対して実験を行った。結果を表8に示す。

同様に、語彙的連鎖のクラスタのマージに関する閾値について、0.05 から 0.5 まで 0.05 刻みで値を変えて行った実験の結果を表9に示す。

これらの結果から、本実験では語彙的連鎖のクラスタのマージの閾値には 0.35 を使用した。また、日本語語彙大系を用いたクラスタリングは語彙的連鎖のクラスタリングに比べて、精度が低いことが判明したため、以降の実験では文中の要素のクラスタリングでは語彙的連鎖によるクラスタリングについてのみ行う。

5.2 実験 1: テキストの並べ替え

文順序に関するタスクでは、3章で述べた評価方法と同様にオリジナルのテキストと文の順番を並べ替えたテキストとを比較し、どちらが一貫性があるかを正しく判定できた精度でモデルの評価を行った。

学習には SVM^{light6} の ranking SVM モードを使用した。コストパラメータ c の値については各モデル毎に予備実験で使ったデータを使って 10 分割交差検定を行い、最も精度が高いものを使用している。各モデルに対する c の値を表 10 に示す。

表 8 予備実験結果（日本語語彙大系）

段	3	4	5	6	7	8	9
精度	0.584	0.597	0.640	0.602	0.631	0.591	0.634

表 9 予備実験結果（語彙的連鎖）

閾値	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
精度	0.685	0.700	0.718	0.688	0.721	0.689	0.722	0.701	0.701	0.713

表 10 各モデルのコストパラメータの値

モデル	no CONJ	+CONJ
Baseline	100	—
CONJ のみ	—	100
REF	100	100
LC(1st)	10	10
LC(comb)	100	100
SR(H)	10	100
ALL-LC(1st)	100	100
ALL-LC(comb)	10	10

⁶ <http://www.cs.cornell.edu/people/tj/svm.light/>

また、並べ替えテキスト中の文の順番がオリジナルのテキストの文の順番と大きく異なっていれば、一貫性の判定は容易になると考えられる。このことを検証するために、表 11 に示す 5 種類の並べ替えの比較を行う。swap1 はオリジナルのテキストとの差が最も小さく、random はその差が最も大きくなる。mix は swap1, swap2, swap3 の混合であり、swap3 よりは差は小さい。

本実験では朝日新聞コーパスの 2003 年分の記事から、“行政改革”、“医療”、“教育”というカテゴリに該当するもので、1 記事あたり 10 文以上で構成されているものを使用した。データセット中に含まれる記事のカテゴリの割合は均等になるように調整している。オリジナルのテキストと比較する並べ替えテキストに関して、表 11 に示す各並べ替えの種類のそれぞれにおいて、1 つの記事に対して 20 個の並べ替えテキストを生成した。

この比較する並べ替えのテキストの数に関して、Karamanis (Karamanis 2006) はセンタリング理論に基づいた手法に対して、信頼できる結果を得るために 100,000 個の並べ替えテキストを生成して評価を行っているが、我々は entity grid に基づいたモデルを用いており、このモデルでの精度向上を目的としているため、使用する並べ替えテキストの数は Barzilay ら (Barzilay and Lapata 2008) の実験の設定にあわせている。

実験はデータセットに対して 10 分割交差検定を行い、テストデータ中の各ペアにおいてオリジナルのテキストの方が一貫性があると判定されたペアの割合で評価した。

表 12 に 100 記事、300 記事に対して mix の並べ替えを行ったデータを用いた各モデルの結果を示す。“no CONJ” は各モデルにおいて接続関係毎の遷移確率の計算を行わなかった場合、“+CONJ” は接続関係毎の遷移確率の計算を行った場合を示す。

表中の太字の数値は使用したデータにおいて最も良かったものを表し、斜字はベースライン（接続関係を未考慮の Baseline）を下回ったものを表す。また、右肩の記号 $^{**}(p < 0.01)$, $^{*}(p < 0.05)$ は符号検定においてベースラインの精度と有意な差があることを示す。

全体としては、いくつかベースラインを下回っているものがあるものの、多くのモデルにおいてベースラインを有意に上回る結果を得ることができた。接続関係毎に遷移確率を計算したモデル（“+CONJ” の列）の方がそうでないモデル（“no CONJ” の列）に比べて良い結果を示

表 11 並べ替えの種類

並べ替え	説明
swap1	無作為に 2 文選び、それらを入れ替える
swap2	無作為に選んだ 2 文の入れ替えを 2 回行う
swap3	無作為に選んだ 2 文の入れ替えを 3 回行う
mix	swap1～swap3 を同程度に行う
random	文の順番を無作為に並べ替える

表 12 モデル別の結果 (実験 1)

	mix(100)		mix(300)	
	no CONJ	+CONJ	no CONJ	+CONJ
Baseline	0.547	–	0.580	–
CONJ のみ	–	0.551	–	0.573
REF	0.546	0.550	0.579	0.573
LC(1st)	0.567	0.582*	0.585	0.677**
LC(comb)	0.549	0.557	0.614	0.773**
SR(H)	0.558	0.562	0.579	0.560**
ALL-LC(1st)	0.574	0.577*	0.574	0.563
ALL-LC(comb)	0.564	0.594**	0.579	0.586

している。特に各データセットにおいて接続関係のタイプを考慮したモデルが最も良い精度を得ており、文脈の展開を明示的に示す接続表現から得られる接続関係が一貫性の判定に有用であることを示している。

grid から作成される文書ベクトルの素性は構文役割の遷移の組合せの数だけ存在する。従って、構文役割を拡張したモデル (SR(H)) はベースラインのモデルに比べて素性の数が多くなり、データが少なかった時の影響が顕著に表れると考えられる。

また、参照表現については少数の明示的な指示形容詞のみに限定したため、参照表現を考慮したモデル (REF) でもそれほど差が出なかったと考えられる。

語彙的結束性に基づいたクラスタリングを行ったモデルでは良好な結果を得ることができた。

本論文で提案した要素の全てを考慮したモデルは、全ての場合で最良の結果を得ることはなく語彙的結束性のみを考慮したモデルを下回った場合もあった。これは構文役割を拡張したモデルと同様に、全ての要素を考慮したモデルと語彙的結束性のみを考慮したモデルとでは文書ベクトルの次元の数が異なることが影響していると考えられる。

本論文で用いている一貫性モデルではモデルの学習に必要なデータは人間が書いたテキストのみであり、学習のために特別な情報を付与する必要はない。従って、学習データの作成に必要なコストはほとんどない。そこで、データを更に増やして実験を行った。この実験ではベースラインのモデル (Baseline)、構文役割を拡張したモデル (SR(H))、全ての要素を考慮したモデル (ALL-LC(comb))、接続関係毎に遷移確率を計算したモデル (+CONJ, Baseline) と全ての要素を考慮した接続関係毎に遷移確率を計算したモデル (+CONJ, ALL-LC(comb)) のみを使用した。結果を表 13 に、そのグラフを図 6 に示す。

接続関係毎の遷移確率を考慮したモデルはそうでないモデルに比べて、データが増加するにつれて精度が向上することが明らかになった。一方、構文役割を拡張したモデルはデータを増やしていてもベースラインに比べてあまり向上は見られなかった。

表 13 データ増加時の精度 (実験 1)

モデル	100	200	300	400	500	600	700	800	900
Baseline	0.547	0.559	0.580	0.595	0.606	0.612	0.623	0.626	0.624
SR(H)	0.558	0.560	0.579	0.601	0.610	0.607	0.617	0.621	0.621
ALL-LC(comb)	0.564	0.561	0.579	0.599	0.606	0.607	0.616	0.619	0.620
+CONJ, Baseline	0.551	0.534**	0.573	0.618**	0.642**	0.652**	0.669**	0.668**	0.669**
+CONJ, ALL-LC(comb)	0.594**	0.564	0.586	0.633**	0.651**	0.655**	0.668**	0.675**	0.675**

図 6 データ増加時の精度 (実験 1)

表 14 データ別の結果 (実験 1)

モデル		swap1	swap2	swap3	mix	random
no CONJ	baseline	0.532	0.552	0.584	0.547	0.687
	ALL-LC(comb)	0.544	0.566	0.609	0.564	0.725**
+CONJ	CONJ のみ	0.560**	0.560	0.666**	0.551	0.696
	ALL-LC(comb)	0.565**	0.633**	0.642**	0.594**	0.719*

表 14 に並べ替えの種類毎での結果を示す. 比較するテキストの差と問題の難易度との関係については, 差が一番小さい swap1 では精度が最も低く, 一番大きい random では精度が最も高くなっており, 仮説通りの結果が得られた. また, 全てのデータセットにおいて Baseline と比べて本論文のモデルの方が良好な結果を得ることができた.

5.3 実験 2: 要約文書の比較

5.2 節で行った実験では、あるテキストとそのテキスト中の文の順番を並べ替えたものとを比較している。このため比較する 2 つのテキストの単語の出現頻度分布は等しいと言える。しかし、実際にはこのような状況は稀であると考えられる。例えば、自動要約の評価においては同じ元文書から生成された要約を用いてシステムの評価を行う。元文書が同じであっても、システム毎に異なる要約が生成されることがあり、これらの単語の出現傾向は異なると考えられる。

そこで、実際に自動要約システムによって生成された要約を比較し、どちらが一貫性があるかを判定するという実験を行った。実験に使用したデータは NTCIR-4⁷(Kando 2004) のサブタスクであった TSC3 (Text Summarization Challenge 3)⁸に提出された要約である。TSC3 では 11 のシステムが 30 件の元文書に対してそれぞれ長短 2 種類の要約を生成している。このうち実際に要約文が出力されている 657 件の要約を使用した。

それぞれの要約には被験者による評価結果が付与されている。この評価では、被験者は 15 個の Quality question と呼ばれるテキストの質に関するチェック項目 (Hirao, Okumura, Fukushima, and Nanba 2004) が示され、各項目毎にスコアを付ける。この Quality question は主に要約の読みやすさに対する質問で構成されている。本実験では特に一貫性に関する項目のスコアのみに着目し、これらから要約の一貫性に関するスコアを計算し比較を行った。

要約のスコアの決定について述べる。TSC3 で用いられた 15 個のチェック項目のうち、付録 A に示す 8 個の項目のスコアを利用した。

それぞれチェック項目のスコアは各項目の内容に該当する箇所の個数であり、 $qq_2 \sim qq_8$ については qq_1 の項目に当てはまった重複文は除外して数えられている。以上より、要約 S のスコア $score(S)$ を以下の式によって求める。

$$score(S) = \frac{N(qq_2) + \dots + N(qq_8)}{length(S) - N(qq_1)} + \frac{N(qq_1)}{length(S)} \quad (3)$$

ここで、 $N(qq_i)$ はチェック項目 qq_i のスコアであり、 $length(S)$ は要約 S の文数である。 qq_8 の回答は“矛盾している”、“どちらともいえない”、“矛盾していない”のいずれかであり、これらのスコアは順に 1, 0.5, 0 として $N(qq_8)$ の値とした。各スコアは文章中のおかしな箇所の個数であることから、 $score(S)$ は小さい方が良いテキスト、即ち、本実験においては一貫性が高いと考える。

この $score(S)$ を用いて、本実験では同じテキストから生成された異なるシステムによる同じ長さの要約のペアに対し、どちらの要約が一貫性があるかを判定する。従って、タスクとしては実験 1 と同じものとなる。比較する要約のスコアが等しいペアは除外する。

⁷ <http://research.nii.ac.jp/ntcir/ntcir-ws4/ws-en.html>

⁸ <http://www.lr.pi.titech.ac.jp/tsc/index-en.html>

テキスト一貫性の判定を実際に利用する状況では、判定が必要なデータを訓練データに用いることはできず、別に訓練データを用意する必要がある。このことを考慮して、本実験では交差検定ではなく 5.2 節の実験において作成した 300 記事とその mix の並べ替えを訓練データとして用い、それによって得られたモデルを用いて判定を行った。実験に使用した学習器や各パラメータの値は 5.2 節での実験と同じである。

また、5.2 節の実験と同様に、比較するテキストの差による精度の違いの検証も行った。本実験ではそれぞれの要約のスコアの差が大きければ、一貫性の判定は容易になると考えられる。そこで比較する要約のペアのスコアの差を 0 から 2.0 まで 0.5 刻みでの範囲で分割し、それぞれの精度を計算した。

用いたテストデータ全てに対する、各モデルの精度を表 15 に示す。表中の記号、字体の意味については前節の実験と同様である。

学習データとテストデータのドメインが異なるために、前節の実験に比べて全体的な精度は低くなっているが、提案したほぼ全てのモデルにおいてベースラインよりも良い精度を得ることができた。

接続関係を考慮したモデル（表中の“+CONJ”の列）とそうでないモデル（表中の“no CONJ”の列）では、最も良い精度を得られたモデルは接続関係を考慮しない場合でのものであったが、それぞれのモデルにおける接続関係の考慮の有無による違いでは考慮した方が良い精度を示しているものが多い。本実験においても、構文役割の拡張 (SR(H)) や参照表現の考慮 (REF) を組み込んだモデルは、ほとんど改善が見られなかった。これは前節の実験結果と同様であった。

語彙的結束性に基づくクラスタリングにおいても、同様にベースラインを上回る結果を得ることができた。

比較した要約のスコアの差が小さければ、それらの一貫性を判定することが困難になると考えられる。そこでテストデータの各ペアのスコアの差毎の精度を求めた。その結果を表 16 に示

表 15 モデル別の結果（実験 2）

	no CONJ	+CONJ
Baseline	0.502	–
CONJ のみ	–	0.507
REF	0.502	0.507
LC(1st)	0.540**	0.563**
LC(comb)	0.586**	0.557**
SR(H)	0.503	0.505
ALL-LC(1st)	0.552**	0.549**
ALL-LC(comb)	0.589**	0.546**

表 16 スコアの差毎の結果 (実験 2)

モデル		all (6434)	0~0.5 (2986)	0.5~1.0 (2014)	1.0~1.5 (930)	1.5~2.0 (334)
no CONJ	Baseline	0.502	0.500	0.502	0.505	0.515
	ALL-LC(comb)	0.589**	0.532*	0.592**	0.669**	0.775**
+CONJ	CONJ のみ	0.507	0.508	0.502	<i>0.503</i>	0.536
	ALL-LC(comb)	0.546**	0.510	0.547**	0.604**	0.689**

表 17 順位相関 (実験 2)

	no CONJ	+CONJ
Baseline	0.207	—
CONJ	—	<i>0.109</i>
REF	0.210	<i>0.108</i>
LC(1st)	0.239	<i>0.097</i>
LC(comb)	0.218	0.195
SR(H)	0.217	<i>0.114</i>
ALL-LC(1st)	0.242	<i>0.121</i>
ALL-LC(comb)	0.231	<i>0.126</i>

す。1 行目のラベルの括弧の中の数字はその範囲に該当する要約ペアの数である。差が 2.0 より大きいペアについては、該当するペアの数が多くなかったため省略している。

要約のスコアの差と精度の関係については、こちらでも仮説通りスコアの差が小さければ判定は難しくなり、差が大きくなれば判定は容易であるという結果になった。

本実験で用いた TSC のデータは 1 つの元文書に対して複数のシステム要約が存在しており、上述の計算式で求められたスコアに基づいて同じ元文書から生成された要約を順位付けすることができる。そこで要約のペアの比較ではなく、同一文書から生成された要約の順位を推定するという実験を行った。使用したモデルは前述の実験と同じ 300 記事とその mix の並べ替えのデータで学習したものである。評価には Spearman の順位相関係数の平均を使用した。結果を表 17 に示す。

ベースラインに比べて提案したモデルに高い相関を示すものがあったが、全体的に相関は低いという結果が得られた。これは特にスコアの差が小さい場合での判定精度が影響していると考えられる。

本実験結果から、ベースラインと比べると、ある程度実際にテキスト一貫性の判定を行うような設定においても本論文で提案したモデルの方が有効であることが明らかになった。しかし、その精度は高いとは言えず、改良の余地がある。

6 おわりに

本論文では接続関係, 参照表現, 語彙的結束性といった結束装置, また, 日本語に特化した構文役割を **entity grid** モデルに組み込むことで, 結束性を考慮したテキストの局所的な一貫性モデルを提案し, その有効性の検証を行った. 実験から, 提案モデルがテキスト一貫性の高いテキストの判定と自動要約によって作成されたテキストの評価の2つのタスクにおいて, オリジナルの **entity grid** モデルを上回ることを示した.

entity grid によるテキストの一貫性モデルはテキストにおける文中の要素の遷移に着目したものである. 本論文ではその要素の遷移は文脈の展開によって違う傾向を示すという仮説を立て, 文脈展開の種類を明示的に示す働きを持つ接続関係に着目し, その種類毎に **grid** を作成するというモデルを提案した. 実験の結果, 接続関係を考慮したモデルの方がそうでないモデルに比べて良い結果を得ることができ, この仮説がテキストの一貫性の評価において有効であることを示した. また, その他の項目についても一貫性の評価において有効であることを示した.

本論文で採用した **entity grid** に基づいた局所的一貫性のモデルでは, テキストを1つのベクトルとして扱うため, テキスト全体についての一貫性の判定は行えても, テキストの部分についての一貫性の評価は行えない. テキストの部分の一貫性の評価は, 小論文の自動評価など教育の現場での利用において有用であると考えられる. また, 例えばテキストにおいて一貫性の悪い箇所の数を用いるなどによって, テキストの部分の一貫性の評価をテキスト全体の評価に用いることも可能であると考えられる. このことから, テキスト内の部分的な単位での一貫性の評価手法の提案が今後の課題である.

参考文献

- Barzilay, R. and Lapata, M. (2005). “Modeling Local Coherence: an Entity-based Approach.” In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, pp. 141–148.
- Barzilay, R. and Lapata, M. (2008). “Modeling Local Coherence: An Entity-Based Approach.” *Computational Linguistics*, **34** (1), pp. 1–34.
- Barzilay, R. and Lee, L. (2004). “Catching the Drift: Probabilistic Content Models, with Applications to Generation and Summarization.” In *HLT-NAACL 2004: Main Proceedings*, pp. 113–120.
- Elsner, M., Austerweil, J., and Charniak, E. (2007). “A Unified Local and Global Model for Discourse Coherence.” In *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main*

- Conference*, pp. 436–443.
- Elsner, M. and Charniak, E. (2008). “Coreference-inspired Coherence Modeling.” In *Proceedings of ACL-08: HLT, Short Papers*, pp. 41–44.
- Filippova, K. and Strube, M. (2007). “Extending the Entity-grid Coherence Model to Semantically Related Entities.” In *Proceedings of the 11th European Workshop on Natural Language Generation*.
- Grosz, B. J., Weinstein, S., and Joshi, A. K. (1995). “Centering: a Framework for Modeling the Local Coherence of Discourse.” *Computational Linguistics*, **21** (2), pp. 203–225.
- Halliday, M. A. K. and Hasan, R. (1976). *Cohesion in English*. Longman, London.
- Hirao, T., Okumura, M., Fukushima, T., and Nanba, H. (2004). “Text Summarization Challenge 3 - Text summarization evaluation at NTCIRWorkshop 4.” Working Notes of the 4th NTCIR Workshop Meeting.
- 市川孝 (1978). 国語教育のための文章論概説. 教育出版.
- 池原悟, 宮崎正弘, 白井諭, 横尾昭男, 中岩浩巳, 小倉健太郎, 大山芳史, 林良彦 (1997). 日本語語彙大系. 岩波書店.
- 庵功雄 (2007). 日本語におけるテキストの結束性の研究. くろしお出版.
- 板倉由知, 白井治彦, 黒岩丈介, 小高知宏, 小倉久和 (2008). 単語の概念関係を用いた段落一貫性評価指標の有効性. 情報処理学会研究報告 NL-183.
- Kando, N. (2004). “Overview of the Fourth NTCIR Workshop.” Working Notes of the 4th NTCIR Workshop meeting.
- Karamanis, N. (2006). “Evaluating Centering for Sentence Ordering in Two New Domains.” In *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers*, pp. 65–68.
- Lin, C.-Y. (2004). “ROUGE: A Package for Automatic Evaluation of Summaries.” In *Proceedings of Workshop on Text Summarization Branches Out, Post Conference Workshop of ACL 2004*, pp. 74–81.
- Lin, C.-Y. and Hovy, E. (2003). “Automatic Evaluation of Summaries Using N-gram Co-Occurrence Statistics.” In *Proceedings of HLT-NAACL-2003*.
- Miltsakaki, E. and Kukich, K. (2004). “Evaluation of Text Coherence for Electronic Essay Scoring Systems.” *Natural Language Engineering*, **10** (1), pp. 25–55.
- Mochizuki, H., Iwayama, M., and Okumura, M. (2000). “Passage-Level Document Retrieval Using Lexical Chains.” In *Proceedings of RIAO 2000*, pp. 491–506.
- Morris, J. and Hirst, G. (1991). “Lexical Cohesion Computed by Thesaural Relations as an Indicator of the Structure of Text.” *Computational Linguistics*, **17** (1), pp. 21–48.

- Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. (2002). “BLEU: a Method for Automatic Evaluation of Machine Translation.” In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pp. 311–318.
- Soricut, R. and Marcu, D. (2006). “Discourse Generation Using Utility-Trained Coherence Models.” In *Proceedings of the COLING/ACL 2006 Main Conference Poster Sessions*, pp. 803–810.
- 田窪行則, 西山佑司, 三藤博, 亀山恵, 片桐恭弘 (2004). 談話と文脈. 言語の科学 7. 岩波書店.
- Walker, M., Iida, M., and Cote, S. (1994). “Japanese Discourse and the Process of Centering.” *Computational Linguistics*, **20** (2), pp. 193–232.

付録

A スコアの計算に使用した **Quality question**

- qq1.** 同一の, あるいはほぼ重複する文はいくつあるか?
- qq2.** (ゼロ) 代名詞化, 指示表現化すべき箇所はいくつあるか?
- qq3.** 先行詞のない指示表現はいくつあるか?
- qq4.** 固有表現の出現位置がおかしい箇所はいくつあるか?
- qq5.** 同一事物を参照する表現の一貫性という観点から修正すべき表現はいくつあるか?
- qq6.** (前後の文脈も踏まえた上で) 必須要素が欠如している箇所はいくつあるか?
- qq7.** 接続詞が必要・不必要な箇所はいくつあるか?
- qq8.** 時系列の関係が矛盾していないか?

略歴

横野 光 (正会員) : 2003 年岡山大学工学部情報工学科卒. 2008 年同大大学院自然科学研究科産業創成工学専攻単位取得退学. 博士 (工学). 同年東京工業大学精密工学研究所研究員, 現在に至る. 自然言語処理の研究に従事. 情報処理学会, 言語処理学会各会員.

奥村 学 (正会員) : 1962 年生. 1984 年東京工業大学工学部情報工学科卒業. 1989 年同大学院博士課程修了. 工学博士. 同年, 東京工業大学工学部情報工学科助手. 1992 年北陸先端科学技術大学院大学情報科学研究科助教授, 2000 年東京工業大学精密工学研究所助教授, 2009 年同教授, 現在に至る. 自然言語処理, 知的情報提示技術, 語学学習支援, テキスト評価分析, テキストマ

イニングに関する研究に従事。情報処理学会，電子情報通信学会，人工知能学会，AAAI，言語処理学会，ACL，認知科学会，計量国語学会各会員。

oku@pi.titech.ac.jp, <http://oku-gw.pi.titech.ac.jp/~oku/>

(2009 年 6 月 16 日 受付)

(2009 年 8 月 20 日 再受付)

(2009 年 9 月 18 日 採録)