Vol.2012-NL-209 No.1
2012/11/22

社会科学系論文を対象にした要旨作成システム

金子　満生[†]　　恵谷　淳一朗[†]　　韓　東力[†]

論文の内容理解や論文サーベイを行う際に要旨は非常に有用である．しかし要旨・要約には著者の意見が反映されていないものや，前書きのようなものがあるものの，要旨・要約自体がそもそも存在しない論文もある．特に社会科学の分野においてはこのよう事例が多々存在する．そこで，本研究では社会科学系論文に対象を絞り，著者の意見と論文中の主要な内容をともに配慮しながら要旨作成を試みる．先行研究ではすでに重要語句抽出の有効性が確認できたため，本研究では重要語句抽出をもとに，先行研究で大きな問題となった文間・文内の繋がりに焦点を当てながら，より詳細に日本語の文構造について調査した．また，よりジャンルに特化させるために 221 編からなる社会科学系論文データベースを作成し，論文からの情報抽出を行った．評価実験の結果，目標である文の繋がりの強化については良い評価が得られたが，それによる弊害も同時に確認したため今後も改善を行っていく．

A Summary Generation System for Social Scientific Papers

MICHIO KANEKO[†]　ZYUNITIRO EDANI[†]　DONGLI HAN[†]

Summaries are quite useful when one is trying to understand the content of a paper, or conducting a survey with a large number of scientific documents. The situation is even clearer for the domain of social science, as most papers are very long and some of them don’t even have any summaries at all. In this work, we narrow our attention down to the social scientific papers and try to generate their summaries automatically. As we have found the usefulness of important word extraction in a previous study to generate summaries for social scientific papers, we continue using important word in this paper and change our focus to connections between or within sentences in a generated summary. Experimental results show the effectiveness our new method in improving the connections between or within sentences, whereas some problems remain as side-effects and will definitely need to be refined in the future.

1. はじめに

学術論文では，本文の前に要旨・要約が存在することが多い．これにより読み手は論文の概要を短い時間である程度把握することができる．特に社会科学系の論文には文章全体が非常に長いものが多く，すべて読むのに非常に時間がかかるため，論文に要旨・要約が存在すれば論文の内容理解やサーベイの際の取捨選択に役立つ．しかし，論文によっては著者の意見を反映していない要約や，そもそも要旨・要約が存在しない論文もある．そこで本研究では，社会科学系の論文を対象に，ジャンルに特化した要旨の作成を行う．ここで，要旨とは「述べられたことの，最も重要なこと．もしくは肝要な事柄.」であり，要約とは「長い話や文章を短くまとめて要点を明らかにすること．また，まとめたもの.」である[1].

既存の要約研究には，語句抽出と文生成による研究 [2] や重要文抽出と文簡約を併用した研究 [3]，文節重要度と係り受け関係を利用した研究[4]のような文章中の重要な部分を用いて要約を作っている研究が数多く存在する．また，話題の導入部や結論部を用い要約を行う研究[5] や話題の流れに着目した研究[6] のように文章の形で要約を作成している研究もある．他にも冗長性の排除を考慮した研究がある[7]．しかしこれらの研究の中には文章の形に依存してしまうものや，全体の流れはつかめるものの，どの箇所が重要なのかわかりにくくなってしまっているものも見受けられている．

そこで我々は，「重要文抽出中心の方が文生成より情報・論理展開ともにシンプルである」[8]という考えに基づき，重要文抽出を中心に論文の重要な点を簡潔にまとめた文章を作ることを目標に要旨作成システムの構築を試みた[9]．システムによって作成された要旨に対し，「文法的に自然かどうか」,「意味の通る日本語かどうか」と「文の繋がりが自然かどうか」の 3 点においてアンケート調査を行ったところ，前者の 2 つに関しては良好の評価を得たが，「文の繋がりが自然かどうか」に関しては良い評価を得ることができなかった．本研究では先行研究[9]で提案された手法を大幅に変更し，以下の 4 点を重視しながら要旨作成システムの再構築を行った．

- 頻出度や位置情報だけでない重要語句・重要文抽出
- 要旨としての読みやすさ
- 文内・文間の繋がり
- 社会科学系論文の特徴

システムは事前準備を除いて大きく分けると，文内整理，重要度判定，要旨作成の 3 つの段階からなる．具体的な処理については 3～5 章で述べ，6 章以降では実験・評価について述べる．

†日本大学大学院総合基礎科学研究科
Graduate School of Integrated Basic Sciences, Nihon University

2. 事前準備

本研究では語句や文の解析・重み付けを行うために以下の 6 つの辞書ファイルをあらかじめ作成しておく.

- <副詞辞書>:「副詞的表現の諸相」[10]から程度を表す副詞を抽出したもの.
- <文末表現辞書>:「日本語表現文型」[11]から助動詞と同様の働きをする表現をすべて抽出したもの.
- <変換規則対応辞書>:接続助詞と接続詞の対応をまとめたもの
- <必須格辞書>:「EDR日本語共起辞書」[12]を用いて, すべての動詞の格情報から上位 3 つまでをファイルに格納したもの.
- <拡充接続詞辞書>:「国語教育のための文章論概説」[13]から 1 つの事柄に関して拡充して述べる接続詞を抽出したもの.(図 1 と表 1 を参照)
- <連結接続詞辞書>:「国語教育のための文章論概説」から 2 つの事柄を論理的に結び付けて述べる接続詞を抽出したもの.(図 1 と表 1 を参照)

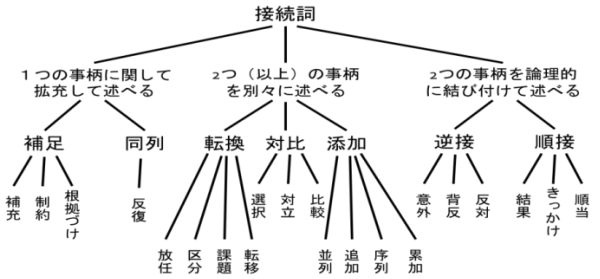


図 1 接続詞の分類

また，社会科学系論文に特化するために以下の 3 つの辞書の作成を行った．1 つ目はキーワード辞書であり,「現代日本政治小辞典」[14]と「現代思想を読む事典」[15]に含まれる見出し語をまとめたものである．2 つ目は名詞句辞書であり，3 つ目は名詞共起情報である．本研究では社会科学系論文に多く出現する名詞句や名詞共起情報を抽出するため，Webから取得した社会科学系の論文 221 編，全 63,056 文からなる社会論文データベースの作成を行った．この社会論文データベースに対し,「名詞句は文構成要素基準の 1 つとなる」という考え[16]に基づき以下の定義に従い名詞句の抽出を行った．抽出された名詞句の総数は 90,365 個である.

- 名詞化した動詞の連用形で終わるもの
- 動詞の連用形に～カタ，～ブリ（ップリ），～ヨウ，～バ，～バショ，～トコロ（ドコロ），～トキ（ドキ）などで終わっているもの
- 形容詞・形容動詞に～サをつけたもの
- 名詞＋名詞
- 用言の連体形＋名詞

表 1 接続詞の細分類

順接	前の内容を条件とするその帰結を導く	順当	だから，それで，したがって，それなら
		きっかけ	すると，と
		結果	かくて，こうして
逆接	前の内容に反する内容を導く	反対	しかし，けれども，だが
		背反	それなのに，そのくせ，しかるに
		意外	ところが，それが
添加	前の内容に付け加わる内容を導く	累加	そして，そうして
		序列	ついで，つぎに
		追加	そのうえ，それに
		並列	また，ならびに
対比	前の内容に対して対比的な内容を導く	比較	というより
		対立	そのかわり
		選択	それとも，あるいは，または
転換	前の内容から転じて，別個の内容を導く	転移	ところで，ときに
		課題	さて
		区分	それでは，では
		放任	ともあれ
同列	前の内容と同等とみなされる内容を導く	反復	すなわち，つまり，ようするに
補足	前の内容を補足する内容を導く	根拠付け	なぜなら，というのは
		制約	ただし，もっとも
		補充	なお，ちなみに

また，話題の流れを見る際に名詞の共起情報が有用であるという考え[17]に基づき社会論文データベースに対し共起情報の抽出を行った．語句A，Bの共起度X（A,B）は相互情報量の考え方に基づき以下の式で表す.

$$X(A,B) = \log \frac{P(A,B)}{P(A)P(B)} \quad (1)$$

ここで，P（A,B）とは語句 A と語句 B が共に 1 文内に出現する確率であり，P（A）と P（B）は語句 A と語句 B がそれぞれ 1 文内に出現する確率を表している．ここで抽出された名詞の共起ペアは 1,141,701 ペアであった．

3. 文内整理

本節以降ではシステムの各モジュールについて詳しく述べていく．文内整理では解析対象の論文を形態素解析したうえ，各文に対して後述の処理を施す．また，語句の重み付けに必要なキーワードの抽出も行う．ここではキーワード辞書に一致した語句をkeywords，名詞句辞書に一致した語句をFkeywordsと呼ぶ．さらに，論文を読み込む際に形態素解析器mecab[18]（以下mecabと呼ぶ）を用いて名詞と複合名詞を抽出するが，これらの名詞・複合名詞をNkeywordsと呼ぶ．但し，ここでの名詞とはmecabで名詞と判断された語句から＜数詞，副詞可能，代名詞，ナイ形容詞語幹，形容動詞語幹，非自立＞とタグ付けられたものを除いたものである．これら 3 種類のキーワードが論文内における出現回数と出現段落数も合わせて記録しておく．以下では，括弧内の整理，三人称削除，文分割と文情報付加の順に詳しく述べていく．

3.1 括弧内の整理

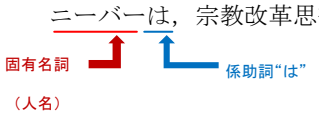
論文を読む際に，丸括弧で挿入されている文を無視しても意味が通じることが殆どであるので，要旨を作成するのに不必要であると判断した．ただし，括弧内のもので 15 文字以下の長さであり，読点を含まないものをその論文独自のキーワードとして抽出する．この 15 文字以下という基準を定めたのはキーワード辞書に含まれる語句の文字数がほとんど 15 文字を超えていないからである．読点を含まないのは，あくまで我々が必要としているのはキーワードになり得る語句であり文ではないので，括弧内の文が抽出されるのを防ぐためである．ここで抽出したキーワードをこの論文独自のキーワードとして Tkeywords と呼ぶ．括弧はキーワード抽出以後の解析に不要なため削除する．

丸括弧以外の括弧に対しては論文内で括弧に囲まれた箇所は，引用文か論文内で特徴的な語句であることが多いので，通常の文と異なる処理が必要である．本来，括弧で囲まれた文は引用文であることが多いため書き手の意見を抽出するということを考えると文そのものを削除すべきであるが，引用文が文中の要素として組み込まれており，これを削除すると文の意味が通らないことや文法的におかしくなるという事態が起こりえるため，括弧のみを消すにとどまった．

3.2 三人称文削除

本研究では極力著者の意見を基に要旨を作成したいため，主語が第三人称の文は要旨に不適切であると考えている．文内に係助詞の“は”か，格助詞の“が”を含み，かつ直前が固有名詞で人名のものか，直前が接尾語で 2 つ前が人名のもの，もしくは“彼”，“彼ら”など明らかに三人称と判断されるものの 3 通りは，三人称文に当てはまるものとして削除しておく．具体的な例を以下に示す．

例.



ここで使用している例文は文献[19]から抽出している．例文には係助詞“は”が存在し，その直前が固有名詞かつ人名なので，この文は三人称が主語の文であると判断され削除される．

3.3 文分割

論文の中には 1 文で複数行に及ぶような長文が数多く存在する．このような長文を解析する際，重要な個所が文の中で偏る可能性がある．そこで，以下の表 2 の規則に従い文分割を行い不要箇所が抽出されないようにする．また，接続助詞の場合は 2 節で述べられた変換規則対応表を用いて分割を行う．表 3 は変換規則対応表の一部である．

表 2　文分割規則

分割前	分割後
動詞（+接尾）　+「、」	原型+「。」+そして+「、」
接続助詞　+「、」	言い切り（原型）+「。」+「接続詞」+「、」

表 3 変換規則対応表の例

	助詞	接続詞
接続助詞	が	だが
接続助詞	て	そして
接続助詞	で	そして
接続助詞	ので	なので
接続助詞	ば	ならば
接続助詞	や	それに
・・・	・・・	・・・

以下に文分割の例を挙げる．ここでは接続助詞“ので”があり，直後が「、」なので，“する”を言い切りの形に変え，「。」を追加し接続助詞“ので”を接続詞“なので”に変えて文を分割する．

例.
～を必要とするので、通常一般の・・・
＝＞ ～を必要とする。なので、通常一般の・・・

3.4 文情報付加

ここでは，重み付けや要旨を作成する際に必要な情報を各文に付加する．最初に結束文の抽出を行う．結束文とは 2 文間の繋がりの強い文のことであり，以下の 4 パターンに当て嵌まる物を結束文とする．

① **疑問詞を含む文と直後の文**
② **指示詞を含む文と直前の文**
③ **「2 つの事柄を論理的に結び付けて述べる」接続詞でつながれている文**
④ **「1 つの事柄に関して拡充して述べる」接続詞でつながれている文**

但し，①の場合は疑問文が最後の文なら抽出を行わず，②の場合は指示詞を含む文が第 1 文目，もしくは指示詞の指す語句が文内に存在していると判断された場合は抽出を行わない．③と④は図 1 の「接続詞の分類」をもとに定めている．

次に各文に対して重要位置との対応付けを行う．著者が社会科学系論文をランダムに 40 編程度選定し，論文内のどの部分で重要な話題が出現しているかを手動で調べてみた．その結果，最後の章や章の最初の段落，最後の段落に重要なことが書かれていることが多いと判明したため，文の位置情報がこの条件にあてはまるものをあらかじめ取得しておく．この際，各文がそれぞれ何章・何段落・何番目に出現したのかという位置情報を付加する．

4. 重要度判定

要旨を作る際にどの文を用いるべきかを決めるため，重要語句抽出，重要文抽出の 2 段階に分けて重み付けを行う．

4.1 重要語句抽出

まず，論文中に出現する語句の重要度を求めるため，keywords，Fkeywords，Nkeywords と Tkeywords の 4 種類のキーワードについて出現頻度を利用した以下の式により語句の重み（以下 Score とする）を計算する．

$$Score = wc \cdot (\frac{wp}{dp} + 1) \quad (2)$$

ここでは単語の出現頻度を wc，単語が出現した段落数を wp，論文内にある全段落の総数を dp とする．但し，すべてのキーワードに対し画一的に重みをつけるのではなく，各キーワードの要旨における重要度により式(3)から式(6)のように差別化を図る．ここでの Score_k，Score_Fk，Score_Nk と Score_Tk はそれぞれ Keywords，Fkeywords，Nkeywords と Tkeywords の重みを表している．

$$Score\_k = wc \cdot (\frac{wp}{dp} + 1) + pos\_\inf \quad (3)$$

$$Score\_Fk = wc \cdot (\frac{wp}{dp} + 1) + pos\_\inf \quad (4)$$

$$Score\_Nk = wc \cdot (\frac{wp}{dp} + 1) \cdot i + pos\_\inf \quad (5)$$

$$Score\_Fk = wc \cdot (\frac{wp}{dp} + 1) + len\_\inf \quad (6)$$

語句が 3.4 節で述べた論文中の重要な位置に出現した場合は，それ以外の場所に出現した語句との差別化を図るためにキーワードの出現回数を表す pos_inf を加算する．また，Tkeywords が論文中の重要な位置に出現した場合は，解析対象の論文で意図的に強調し，なおかつ論文の重要な位置に出現しているので他のキーワードよりも重みを高くするために Tkeywords の文字数である len_inf を加算する．さらに，Nkeywords は他の辞書に載っているものや論文で強調しているキーワードと違い，単なる頻出名詞のため，重要度を低く抑えるべく係数 i（0~1 の間）を掛けて差別化を図る．以上のように計算した各キーワードの score が高い順に上位 5 つを抽出する．この過程で得られた上位 5 つのキーワードを多くの社会科学系論文と同様に，作成された論文要旨とともに結果画面に表示する（図 5 を参照）．

4.2 重要文抽出

本節では，論文中にある各文の重要度を計算するための手法を述べる．文の重要度（以下 bun_score と呼ぶ）を基に重要文ランキングを作成し，それにより要旨に必要な文を取得する．まずは各文に対し、以下の要素を含んでいるかを確認する．

- 強調表現：副詞辞書に存在している語句を含んでいるか.
- 末尾表現：末尾表現辞書に存在している語句を含んでいるか.
- 主題 : 係助詞 “は” の直前の語句が名詞のもの.
- 結束文：結束文を持っているかを判定する.

結束文を持っている場合は，2 文間にわたって話題が継続している可能性が高く，単独の文に比べ情報量が多いと思われるためより高い重要度を付与する．

ここでは主題は名詞であれば何でもよいということではなく，Nkeywords と同様の基準を満たすもののみとする．また，主題もキーワードと同じく論文が何について述べているかを判断する要因の一つになりうるので，出現回数で上位 5 つのものを論文要旨と共に表示する（図 5 を参照）．

文の重要度は式（7）によりで求められる．

$$bun\_score = \sum\{Score\ \ (keywords)\} \times \alpha^{k} \quad (7)$$

基本的には文中に含まれるキーワード（*keywords*）の重みの総和で求まるが，文内でキーワードを特定する際に文字数の多いキーワードから順に探索している．これは，例えばキーワード辞書には「階級」,「政治」というような文字数の少ない語句が存在するが,「階級政治」のように他のキーワードを含むキーワードも存在するため，誤って本来抽出されるべきキーワードが抽出されないという事態を防ぐためである．これに加えて上記の 4 つの要素を含んでいる場合には他の文よりも重要度が高いと考えられるため，重み係数 α（1.1）を掛ける．ここでは，上記の要素を複数含んでいる場合は含まれた要素の数（k 個）だけ α を繰り返して掛け合わせる．以上の手順により計算した重要度の高い順に重要文ランキングを作成する．

5. 要旨作成

この章では実際に要旨作成に用いる文を取得・整形し要旨として結合していく手順を説明する．まず各文に対する処理を行う前に必須格の抽出を行う．必須格とは「ガ」,「ヲ」,「ニ」など述語の表す内容に必要な格である．必須格を含む文節がなければ文章として不自然となってしまう．よってここでは，事前準備で読み込んだ必須格情報を用いて述語に係っている必須格を含む文節を抽出する．

5.1 文簡約

要旨では文内の最も重要なことが述べられるのが望ましいため，以下の図 2 のように文簡約を行い，不要個所を削除しておく．

ここではまず各文に対し Cabocha[20]を用いて係り受け解析を行う．解析結果を受け文節，受け文に直接かかっている係り文節，受け文に直接かかっていない係り文節に分ける．図 3 の例文では「化したのである」が受け文節であり，「言葉は」,「言葉とともに」と「死語と」が受け文に直接かかっている係り文節であり，「階級政治という」と「階級という」が受け文に直接かかっていない係り文節である．

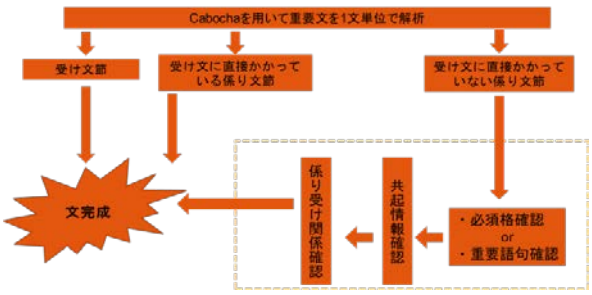


図 2 文簡約の流れ

次に，受け文に直接かかっていない係り文節の中に，先に抽出した必須格を含む文節もしくは重要語句を含む文節があれば，文の構成要素として重要な役割を担っている可能性が高いとして抽出する．また，ここで取得した文節中にある名詞と，名詞共起情報に存在する共起パターンと一致したペアの名詞を含む文節があれば抽出する．これは，共起している名詞は話題の流れをつかむ要因になりうるので，文の構成要素として重視すべきという考え方に基づいている．例えば，語句「サミット」を含む文節が抽出された場合，抽出されなかった文節に語句「自国」を含む文節があれば抽出を行う（表 4 を参照）．最後にこれまで抽出された文節に係られている文節があれば抽出を行う．例えば図 3 の例では，「階級政治という」という文節が抽出されたが，「言葉は」という文節が抽出されなかった場合は，文の意味が通じなくなってしまうため文節「言葉は」も抽出する．

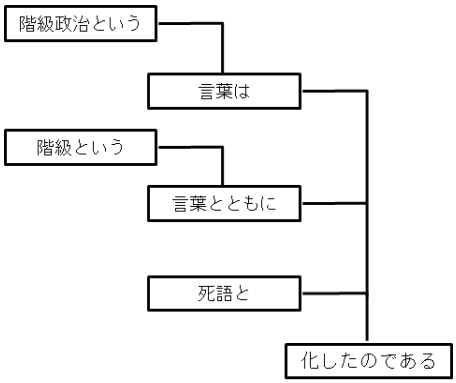


図 3 係り受け解析の例

ここまでの過程で抽出した受け文節，受け文節に直接かかっている係り文節と受け文に直接かかっていない係り文節を元々の文内における出現順に結合し，簡約文として改めて作成する．こうやって簡約された文に 4.2 節で求まった重要文ランキングの順位を継承させる．

表 4 名詞共起情報の例

サミット	ミーイズム	1.588042
サミット	世界	0.458759
サミット	論説	2.043721
サミット	各国	1.628684
サミット	自国	2.365649
サミット	利益	0.780687
サミット	形骸	3.365649
サミット	経済	1.687578

5.2 要旨文取得

この処理では指定された要約率に従い要旨の構成要素である文の取得を行う．システム利用時にユーザーに作成された要旨の割合を指定してもらうので，その割合に基づいて式（8）により必要語数の計算を行う．ここでは N は必要語数を表す．

$$N = \text{指定された割合} \times \text{全文章の文字数} \quad (8)$$

重要文ランキングの順位に基づき上位の文から 1 文ずつ取得していく．今まで取得された文の総文字数より必要語数が多ければ，再び文の取得を行う．結束文を含む場合はそのペア文の文字数も総文字数に加える．必要語数より総文字数が多くなった最初の時点において，もし最後に抽出した文が結束文を含むのならば最後の文を取得し，それ以外なら最後の文を削除し，要旨文の取得処理を終了させる．

5.3 結束文取得

5.2 節で取得していた文を出現順に並び替え，各文に結束文のペアを持つかどうかを確認する．もし結束文のペアがまだ取得されていなければ図 4 のように追加する．これにより話題の連続性が向上し，文間の繋がりの強化を図る．

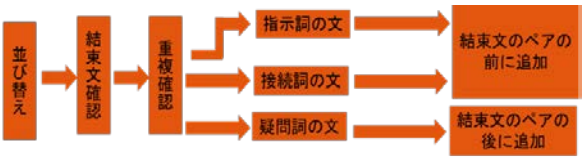


図 4 結束文取得の流れ

5.4 文章組み立て

5.3 節までの処理では，単に文を羅列しただけに留まっているが，ここではさらに読み手に読みやすくするために文章の形に整形していく．

まずは接続詞の調整を行う．元々の出現段落が同じである文は接続詞を残し，それ以外の接続詞は削除する．これにより本来とは違う意味で文が繋がってしまうことや不自然に文を繋げることを防ぐ．

次に文の結合を行う．先行研究[9]では，1 文が極端に短くなることや，文が細切れになり文間の繋がりが悪くなった箇所が存在したためその対策としてこの処理を行う．隣り合う文同士が同一文や，隣接する文であった場合は表 2 に示した文分割後の形式に当てはまれば，「文分割ルール」の逆を用いて文の結合を行う．また，話題の流れを見やすくすれば文同士の繋がりもより強くなるという仮説から主題の確認を行う．直前の文と同じ主題を持つなら，主題の普通名詞を代名詞に置き換える．

上述の整形処理を経た要旨文は図 5 のように位置情報を用いて章が違えば段落を変えて表示する．ここでは指定された要約率で作成した要旨をキーワードと主題の上位 5 つずつと共に表示する．「原文表示」機能では，原文のどの部分を要旨作成に利用しているかを文節単位でテキストの色を変えながら表示している．また，作成する要旨の割合を変更した際，増えた文の色を変えることによりどの要素が加わったかもわかるようになっている．

図 5 要旨出力画面

6. 評価実験

原論文に掲載されている著者の書いた要旨（以下「原文」と呼ぶ）とほぼ同じ割合（字数）で，本研究で作成した要旨，先行研究で構築されたシステム[9]で作成した要旨と Microsoft Word 2003 で作成した要旨（以下「Word」と呼ぶ）の 4 つの要旨を用いて，アンケートによる評価実験を行った．

著者らと同じ言語情報処理分野の大学院生 4 名と学部生 14 名に対し，3 人か 4 人を 1 クループとし合計 5 つのグループに分かれてもらった．上記の 4 パターンの要旨を，どの要旨がどのパターンか分からないように論文原文とともに用意し，30 分程度読んでもらった．その後は 20 分程度グループディスカッションをしてもらい，「文章が文法的に自然かどうか」，「意味の通る日本語かどうか」，「文の繋がりが自然かどうか」，「要旨として適切かどうか」の 4 通りの設問に対し良かった順に並べ替えて相対評価をしてもらった．先行研究のアンケート調査では，システムが作成した要旨を評価したところ，文法も意味も 70%以上の満足度を得たが，文の繋がりについては 65%という比較的に残念な結果となってしまった[9]. 6.1 と 6.2 節では本研究のアンケート調査の結果を先行研究と比較しながら考察していく．

6.1 結果

先行研究で評価が低かった「文の繋がりが自然かどうか」に対する本研究のアンケート結果は以下の図 6 のような分布になった. 原文を除くと, 本研究は最も評価が高く，先行研究と Word を凌いだ結果をなった．このことから，著者が書いた要旨には劣るが，繋がりの増強を目的に導入してきた対策がある程度有効なのではないかと思われる．

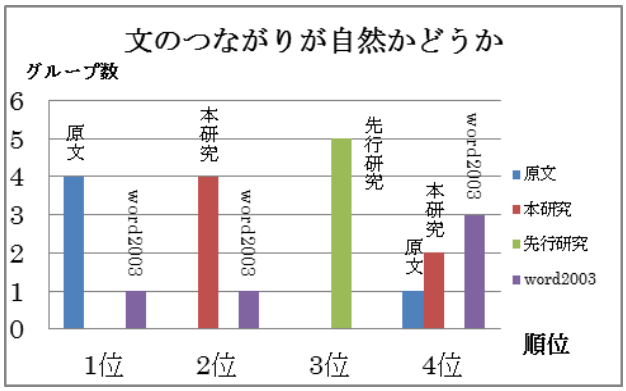


図 6 文の繋がりが自然かどうかの評価

また「要旨として適切かどうか」に関しては図 7 のようになった.「文の繋がりが自然かどうか」と同様の方法で評価したところ，原文が最も高く，本研究と先行研究は同率で 2 位，Word は 4 位であった.

しかし,「文章が文法的に自然かどうか」と「意味の通る日本語かどうか」についてはあまり好ましい評価を得ることができなかった．これについてはアンケート回答後に聞き取り調査を行ったところ,「代名詞が良く出てくる」や「1 文が長い」などいずれも文の繋がりを増強させる対策に由来している結果であることが分かった.

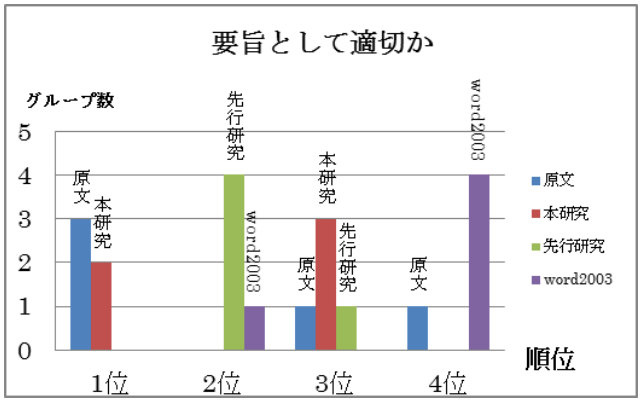


図 7 要旨として適切かどうかの評価

6.2 考察

本研究の主な目標としていた「文間繋がりの増強」については予想通りの評価を得ることができたが,「文章が文法的に自然かどうか」と「意味の通る日本語かどうか」については予想を下回る結果となってしまった．これは先行研究で短くなりがちであった文を，本研究では原論文の長さに近づけたことと，元々の要旨が論文全体の 3〜4%程度と非常に短かったため，その割合に合わせて要旨を作成することにより，論文によっては 2 文のみで成り立つ要旨なども生成されたのが主な原因であると考えられる.

改善案としては，たとえば重要文抽出の際に結束文を考慮される要素から外すことや文章組み立ての際に主題が同じであっても代名詞に置き換えないという手段が考えられる．また，要旨を含む社会科学系論文を収集し，要旨に存在する 1 文の長さを統計的に調査し，分布を参照することにより要旨に相応しい 1 文の長さを求めるという手法も考えられる.

7. おわりに

本研究では社会科学系論文を対象にした要旨作成を試みた．重要文抽出などで論文中の重要な個所を用いて要旨を作成したほか，先行研究に比べ社会科学系論文データベースの作成やデータベースから得られた統計情報の導入，さらに日本語文の要素について詳細な調査を行うことにより文中・文間の繋がりをより自然にすることができた.

しかし，要旨の割合が少なくなると文章全体に対して 1 文が占める割合が大きくなり，それが被験者にとって読みにくい一因となることが分かった．今後は指定された割合に適した要旨の長さについても検討を行っていく.

全体的には，文間の繋がり，文法の正しさと意味の通りやすさの各指標は互いに牽制しているようであるため，システム全体の性能を保ちながら，各指標とも良好になるような均衡点を見つけることが今後の課題となるであろう.

参考文献

1) 梅棹忠夫, 金田一春彦, 阪倉篤義, 日野原重明: 日本語大辞典講談社カラー版第二版, 講談社, (1995).
2) 畑山満, 武美子, 松尾義博, 白井諭: 重要語句抽出による新聞記事自動要約 , 自然言語処理, 9(4), pp.55-73. (2001).
3) 諸岡祐平, 江嵜誠, 高木一幸, 尾関和彦: 重要文抽出と文簡約を併用した新聞記事の自動要約, 言語処理学会第 10 回年次大会発表論文集, A10. P4-04. (2004).
4) 小黒玲, 小関和彦, 張玉潔, 高木幸一: 文節重要度と係り受け整合度に基づく文簡約アルゴリズム, 言語処理学会第 6 回年次大会発表論文集, pp.133-136. (2001).
5) 川端正法, 山本和英: 話題の継続に着目した国会会議録要約, 言語処理学会第 13 回年次大会, pp.696-699. (2007).
6) 市丸夏樹, 日高達: 要約文の話題の流れの最大化による自動要約, 自然言語処理, 12(6), pp.45-61. (2005).
7) 望月源, 奥村学: 読みやすさの向上と冗長性の排除を考慮した重要個所抽出型要約, 情報処理学会研究報告, NL139, pp.17-24. 2000).
8) 望主雅子, 萩野紫穂, 太田公子, 井佐原均: 重要文と要約の差異に基づく要約手法の調査, 情報処理学会研究報告, NL−135, pp.95-102, (2000).
9) 金子満生, 恵谷淳一郎, 松澤由梨枝, 韓東力: 重要語句抽出を利用した要旨作成システム, 言語処理学会第 18 回年次大会発表論文集, F4-1, (2012).
10) 仁田義雄: 副詞的表現の諸相, くろしお出版 (2002).
11) 森田良行, 松木正恵: 日本語表現文型―用例中心・複合辞の意味と用法, アルク (1989).
12) EDR 日本語共起辞書
http://nlp.kuee.kyoto-u.ac.jp/NLP_Portal/lr-cat-j.html
13) 市川孝: 国語教育のための文章論概説, 教育出版 (1978).
14) 内田満(編集): 現代日本政治小辞典, ブレーン出版 (2001).
15) 今村仁司(編): 現代思想を読む事典, 講談社現代新書 (1988).
16) 南不二男: 現代日本語の構造, 大修館書店 (1974).
17) 市丸夏樹, 飛松宏征, 日高達: 話題の流れを保持する自動要

約, 情報処理学会研究報告, NL-160, pp.43-48. (2004).
18) http://mecab.sourceforge.net/
19) 佐久間重: キリスト教神学における歴史認識-ラインホールド・ニーバーによる近代文化についての見解- , 名古屋文理大学紀要 10, pp.5-61. (2010).
20) http://code.google.com/p/cabocha/