

異文化的背景からみる文章要約プロセスの比較研究

香取 真理^{*}

§ 1. はじめに

近年、インターネット等を通じて、日々膨大な情報へのアクセスが可能となり、必要な情報のみを的確かつ迅速に選択する能力が必要不可欠となっている。人間の知的活動を支援する基礎技術として、多角的な観点から自動要約に関する研究が進められており、自然言語処理の一つの研究分野となっている。

本研究は、文章理解過程に深く関わる、人の文章要約過程の解明を目指すものであり、異文化的背景を持つ被験者による、要約文生成過程における言語現象を分析することにより、その特徴を明瞭化しようと試みたものである。

§ 2. 研究の目的

文章の要約という作業は人間の文章理解に関わる重要な認知行動の一つであり、これまでも幾つかの研究が行われてきた。van Dijk & Kintch (1983)、Lehnert(1982)、Kay&Black(1986)、Omanson (1982)、Thorndyke(1977)、佐久間(1989)、邑本(1998)等による、認知言語学的・認知心理学的研究の一部としての要約研究や、Jing and Mckeown(2000)、Endres-Niggenmeyer(1998)、Mani(2001)らによる自動要約システム研究の一部としての、人の文章要約過程の研究である。前者は主にストーリー性のあるオリジナルテキスト(文章)からの要約文章作成過程の研究であり、後者は主に論文や記事等をオリジナルテキストとした、抄録やアブストラクト等からの要約研究である。

近年では、多量なテキストの中から必要なテキストを検索する、テキスト検索システム、携帯端末への情報配信システム、テレビの字幕作成システムや災害時における情報提供システムなど、人間の知的活動を支援する基礎技術として、

多角的な観点から自動要約に関する研究が進められており、要約研究は、自然言語処理の一つの研究分野となっている。

幾つかの言語を処理し、入力と同じ言語で出力する多言語要約 (Multilingual Summarization) システムや、幾つかの言語を処理し、入力と異なる言語で出力する言語横断要約 (Cross-Lingual Summarization) システム等の開発も進められており、実用化されれば、異文化間での情報交換に非常に有効であると思われる。

しかし、長尾、黒橋他 (1998) が指摘するように、自動要約が人の要約と同じように「理解」を有する文章構築を行うようになるまでには、まだ多くの課題が残されている。Inderjeet (奥村、難波、上田訳) (2003) もその著書の中で、「機械がどうすれば(人間と)同じように出来るかに光を当てるため、人間の要約者がどのように行っているかに関する研究がもっと必要である。」と述べている。

香取 (2004a) では、異なる文化的背景を持つ被験者 (日本語母語話者20名・英語母語話者20名) を被験者とし、日本語母語話者による日本文要約、及び英語母語話者による英文要約を調査の対象とした。また、それぞれの要約文を分析すると同時に、被験者へのインタビューを行い、日本語母語話者及び英語母語話者による要約プロセスを定性的・定量的の両側面から調査・解析した。

香取 (2004b) では、主に定性的解析を行い、被験者へのインタビューと要約文分析の結果から、日本語母語話者及び英語母語話者、それぞれに関し、要約アルゴリズムモデルの概要をデザインした。その結果、日本語母語話者被験者はオリジナルの文章を使用しながら「大意」型の要約を行う傾向にあるが、英語母語話者被験者

^{*}青森公立大学准教授

はオリジナルの文章を理解し、他の言葉で言い換える「要旨」型の要約を行う傾向があるという結論を得た。

香取 (2004a) の定量的解析では、要約文章解析に有効であると考えられる、新たなIU 「アイデアユニット(idea unit)」 (以降IU¹⁾) 分割基準を設定し、その妥当性を確認するとともに、ノンパラメトリック検定により、日本語母語話者、英語母語話者両被験者による要約テキストの中に使用されたIUの割合 (IU使用率) に有意な相違がみられるかどうかを検証した。

適合度の検定では、IU使用率の分布に関し、英語母語話者は日本語母語話者と異なる分布を示すという結果が得られた。また、独立性の検定からは、IU使用率が使用母語と高い確率で関係しているという結果が明らかになった。

香取(2004a, 2004b)では、「日本語母語話者」と「英語母語話者」というカテゴリーで分析を行ったが、他の異文化的背景は分析に加えなかった。

本研究では、異文化的バックグラウンドをもつ被験者として、日本語母語話者、英語母語話者被験者を、それぞれ、母語、年代、性別のカテゴリーに分け、異文化的背景をもつ被験者の要約文章生成時にみられる特徴を、オリジナルテキストからのIU使用率と要約率のばらつきに関し、検証する。

§ 3. 研究方法

3.1. 調査概要

被験者60名 (日本語母語話者30名、英語母語話者30名) に対し、調査を行った。被験者は日本語母語話者被験者、英語母語話者被験者共に、大学生、大学院生、若しくは大学・大学院を修了した社会人である。表1は被験者のバックグラウンドによる分類を示している。日本語母語話者には日本語のオリジナルテキストを、英語母語話者には英語のオリジナルテキストを渡し、自由に要約してもらった。オリジナルテキストの量は、双方ともA4サイズの用紙1枚程度 (日本文 1,079文字、英文 484語²⁾) の文章である。

生成される要約文の分量は、オリジナルテキストの50%量を目安とすることを告げるが、厳密

表1.

	使用母語	性別	年代
被 験 者	日本語母語話者 (30名)	男性 (15名)	20代 (5名)
			30代 (5名)
			40代以上(5名)
		女性 (15名)	20代 (5名)
			30代 (5名)
			40代以上(5名)
	英語母語話者 (30名)	男性 (15名)	20代 (5名)
			30代 (5名)
			40代以上(5名)
		女性 (15名)	20代 (5名)
			30代 (5名)
			40代以上(5名)

なものではないことを知らせる。被験者は、自由にメモ等を取りながら要約する。但し、オリジナルテキストの中に理解できない単語等が存在しても、辞書等で確認しないものとした。以下に被験者への指示を記す。

指示 1 ; 原文は1,079文字です。これを約半数500文字ぐらいに要約してください。

All you have to do is to read a text (it contains about 484 words) and summarize it to about 240-250 words (to about 50%).

指示 2 ; もし、分からない単語や表現があっても辞書で調べたりしないでください。

Even if you find some words that you don't understand or don't know, do not consult a dictionary.

指示 3 ; 後ほど、要約文について幾つか質問させていただきます。

Later, I will ask some questions about your summary.

指示 4 ; 制限時間はありませんが、30分から40分ほどで終了するかと思います。要約にかかった時間を計ってください。

There is no time limit to summarize but could you please measure the time it took you to summarize. (Usually it takes 30-40 minutes.)

被験者が生成した要約文章は、香取（2004a）で使用したIUを使用し、調査・検証を行った。後に、要約文の要約率、要約文中のIU使用率に関し、データのばらつきからみる、要約文章の特徴について検証を行った。

3.2. IU分割基準

文章の分析研究を行う場合、研究の目的や意図により、さまざまな条件で文章の「単位」分けが必要となってくる。文章要約に関する先行研究の中でも、文章の「単位」分けは研究者により異なる定義を用いて行われている。また「単位」の名称もさまざまである。

佐久間（1994）は「残存認定単位（Z単位）」という名称を用い、1. 主節、2. 連用節（従属節）、3. 連体（修飾）節、4. 引用節、5. 提題表現、6. 状況表現、7. 文副詞表現、8. 接続表現、9. 独立表現、10. その他の特殊表現（1～9以外）という10種類の分類基準を適用した。

呂本（1998）は分析単位をIUとし、7種の認定基準を設定した。基本的には単文を1つのIUとする分類基準を設定しているが、完全なIUの基準を確立することは極めて困難であるとの見解を述べている。

英文テキストのIU認定基準に関しても、さまざまな基準が検証されて来た。Lehnert（1982）のplot unit や、IUを節(single clause)単位に分割することを基本としたCarrell（1985）、それを修正したIkeno（1996）等がある。Kimura（1999）で使用されたIkeno（1996の分析私信）によるIU認定基準に関しても、やはり節を基準とした分割が試みられているが、不定詞や、副詞句の分割方法にも言及し、具体的な10種の基準を設けている。

しかし、これらはいずれも日本語もしくは英語という一つの言語に関するIU設定基準であるため、日本語オリジナルテキストと英語オリジナルテキスト双方の調査を有する研究においては、両言語の構造的相違から、先行研究のいずれか単独のIU基準を採択することは困難である。

そのため香取（2004a）では、上記の先行研究におけるIU分類基準を参考にし、以下のような基準を設定した。基準設定に関しては、日本語

と英語の構造的相違を認識しつつ、双方に共通しうる、かつできる限り明快な基準設定を試みた。尚、本基準中での「命題」とは益岡、仁田、郡司、金水（1997, p.11）の「ひとまとまりの事態、文の意味内容のうち客体化・対象化された出来事や事柄を表した部分である」という定義を用いる。

a. 日本語テキストにおけるIU分割基準

1. 基本的には単文。
2. 一つのIUには命題が2つ以上入らないようにする。
例) プッシュ政権はイラクを攻撃し政権を転覆させた。
命題1) プッシュ政権はイラクを攻撃した
命題2) プッシュ政権は（イラクの）政権を転覆させた
（例文には上記のように2つの命題が含まれている。よってそれぞれを独立のIUとする。）
3. 句読点がある場合はそこで区切る。
4. 「」（かぎ括弧）、で囲まれている部分はその部分で一つのIUとする。

b. 英語テキストにおけるIU分割基準

1. 基本的には単文。
2. 一つのIUには命題が2つ以上入らないようにする。
例) There is a danger of naming an enemy and invoking a worldwide struggle.
命題1) There is a danger of naming an enemy.
命題2) There is a danger of invoking a worldwide struggle.
3. , (カンマ)、(ピリオド) : (コロン) がある場合はそこで区切る。
4. - (ハイフン)や" (引用符)で囲まれている部分はその部分で一つのユニットとする。
5. 不定詞があり、不定詞以下が3つ以上の単語で構成されている場合、不定詞部分以降から新たなIUとする。
6. 分詞があり、分詞以下が3つ以上の単語で構成されている場合、分詞部分以降から新たな

IUとする。

- 3単語以上からなる前置詞句はそれの一つのIUとする。
- 等位接続詞が複数の語または句を対等に結んでいる場合、等位接続詞以降の最初の語または句のみを一つのIUに含める。それ以降は別のIUとする。

上記の基準に基づき、附録1に示す日本語オリジナルテキスト、英語オリジナルテキストについて、それぞれIU分割を行った。日本語オリジナルテキストは79個のIUに、英語オリジナルテキストは98個のIUにそれぞれ分割された。

附録2 は日本語テキストにおけるIU分割結果を、附録3は、英語テキストのIU分割結果を示している。また、本研究によるオリジナルテキスト選択の理由は、以下の4点である。

- 日本語母語話者、英語母語話者両被験者が同等の背景知識を持つ題材である。
- 日本文、英文共に主題・主張に大きな差異がない。
- ほぼ同量である。
- 難易度に大きな差異が認められない。

§ 4. 研究結果

4.1. 母語別IU使用率と要約率の関係

図1. は日本語母語話者被験者、図2. は英語母語話者被験者に関する、カテゴリー別IU使用率と要約率の平均値をグラフ化したものである。本論文における「要約率」とは、日本語の場合は、要約文中の文字数をオリジナルテキストの文字数で割ったもの（要約文の文字数 ÷ オリ

図1.

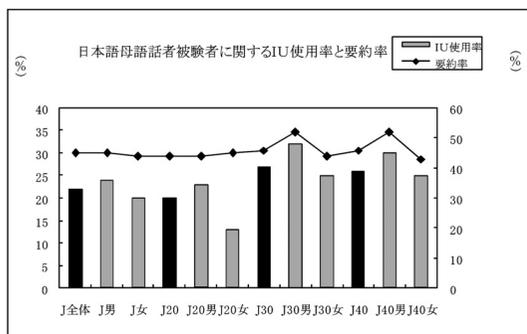
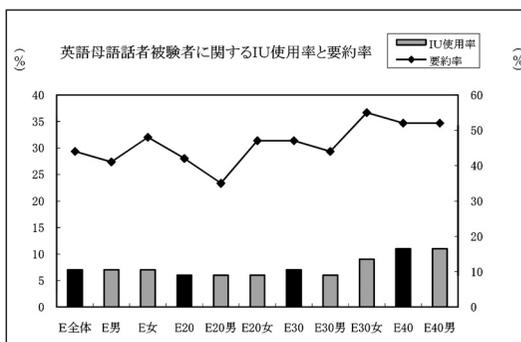


図2.



ジナルテキストの文字数)、英語の場合は、要約文中の単語数をオリジナルテキストの単語数で割ったもの（要約文の単語数 ÷ オリジナルテキストの単語数）とする。

図1. より、日本語母語話者被験者はIU使用率が高く、要約率は、ほぼ一定していることが伺える。換言すると、日本語母語話者は文章を要約する際、オリジナルテキスト中の表現や文章をそのままの形で多数使用し、最終的には被験者の多くがほぼ同じ長さの要約文章を生成する傾向があると言える。

一方、図2. から、英語母語話者被験者は、オリジナルテキストからのIU使用率は低く一定しているが、要約率は年代、性別により多様性があることが伺える。すなわち、英語母語話者は、文章を要約する際、オリジナルテキスト中の表現や文章をそのまま抜き出して使用する頻度は少なく、最終的には多様な長さの要約文章を生成していた結果となった。

図3.は、被験者のカテゴリー別IU使用率と要約率の関係を表したものである。横軸は被験者のIU使用率を、縦軸は要約率のばらつきを示している。データは右上に行くほど、被験者がオリジナルのテキストに沿った長い要約文章を生成した事を示し、左下に行くほど、オリジナルテキストからパラフレーズして短い要約文章を生成したことを示している。

図3.中の「E」は英語母語話者被験者、「J」は日本語母語話者被験者を示している。また「男」は男性「女」は女性被験者、「20, 30」等の数字は被験者の年代を示している。

図3. 被験者カテゴリ別 IU使用率と要約率
(J: 日本語母語話者, E: 英語母語話者)

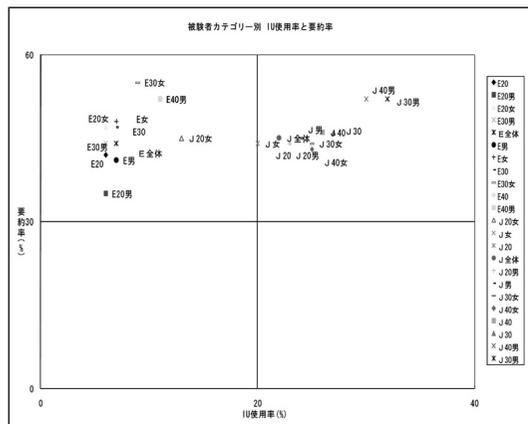


図3. より、日本語母語話者被験者は、図の右上にデータが集まる傾向にあったが、20代女性被験者のみ、例外的にグラフの左側、すなわちIU使用率が低い傾向にあるという特徴がみられた。これに対し、30代・40代の日本語母語話者男性被験者は、全被験者の中でも、最もIU使用率が高く、オリジナルテキストの表現をそのまま使用し、かつ長めの要約文章を生成する傾向があった。また、英語母語話者被験者の中でも、20代女性と20代男性はIU使用率が低い傾向がみられた。

表2. 日本語母語話者に視られるIU使用率のばらつき

	J全体	JM	JF	J20's	J20'sM	J20'sF	J30's	J30'sF	J40's [~]	J40's [~] F
平均	22	24	20	20	23	13	27	25	26	25
範囲	49	47	44	49	47	30	40	40	29	29
標準偏差	13	13	14	13	14	7	17	19	13	16

表3. 日本語母語話者に視られる要約率のばらつき

	J全体	JM	JF	J20's	J20'sM	J20'sF	J30's	J30'sF	J40's [~]	J40's [~] F
平均	45	45	44	44	44	45	46	44	46	43
範囲	26	18	26	19	18	17	19	19	22	22
標準偏差	6	5	8	5	5	6	9	9	11	12

表4. 英語母語話者に視られるIU使用率のばらつき

	E全体	EM	EF	E20's	E20'sM	E20'sF	E30's	E30'sM	E30'sF	E40's [~]
平均	7	7	7	6	6	6	7	6	9	11
範囲	23	23	17	18	6	17	15	15	13	19
標準偏差	6	7	6	5	6	5	6	6	9	11

表5. 英語母語話者に視られる要約率のばらつき

	E全体	EM	EF	E20's	E20'sM	E20'sF	E30's	E30'sM	E30'sF	E40's [~]
平均	44	41	48	42	35	47	47	44	55	52
範囲	58	58	44	58	58	38	37	33	6	6
標準偏差	14	16	10	16	18	11	12	13	4	3

4.2. 異文化的カテゴリにみられるばらつきの特徴

次に、IU使用率と要約率に関し、データの範囲、標準偏差の2点から、そのばらつきにみられる特徴を検証する。

表2. から表5. は母語別にIU使用率と要約率に関し、数値的な代表値(平均、範囲、標準偏差)を算出し、その結果をまとめたものである。

図4. から図7. は、範囲、標準偏差の2点に関し、表2. から表5.の数値を基にグラフ化したものである。

図中の「E」は英語母語話者被験者、「J」は日本語母語話者被験者を示している。また「M」は男性「F」は女性被験者、「20's」は20代被験者を示している³⁾。

4.1.より、日本語母語話者は英語母語話者被験者と比べ、IU使用率が高く要約率が一定であることが分かったが、ここではそのばらつきに関して検証を行う。

表2. 表3. と図4. 図5. より、日本語母語話者はIU使用率のばらつきは比較的大きく、要約率のばらつきは少ないという傾向が伺える。しかし、20代女性被験者だけはIU使用率のばらつきが小さい傾向を示している。4.1.の結果と総合する

と日本語母語話者被験者の中で20代女性被験者だけは文章要約時にオリジナルテキストの表現を抜き出して使用する傾向が少ない結果となった。

図4.

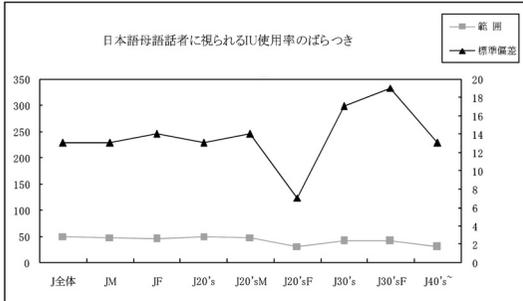


図5.

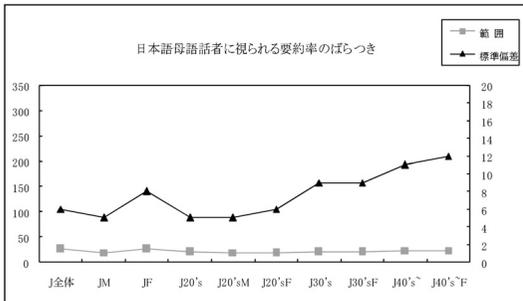


図6.

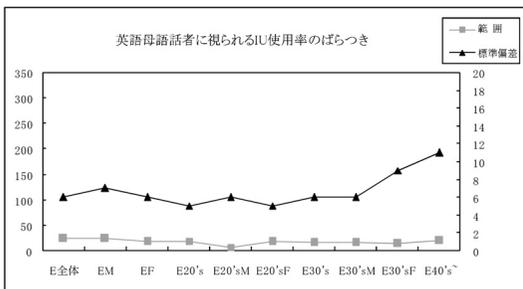


図7.

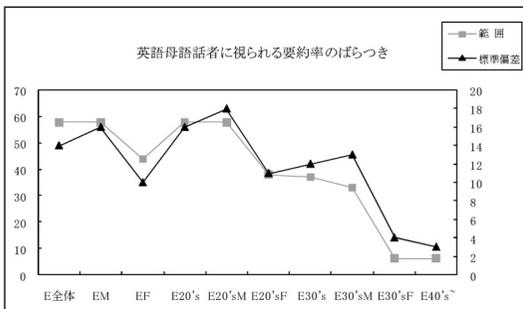


表4. 表5. と図6. 図7. より、英語母語話者被験者は日本語母語話者被験者とは対照的にIU使用率のばらつきは少なく、要約率にばらつきがみられるという傾向があった。つまり、要約時には主にパラフレーズを行いながら、多様な分量の要約文章を生成したことになる。

また、表4. 表5. により、英語母語話者被験者の中でも、20代の女性被験者は、要約率、IU使用率のばらつきが小さい傾向を示す結果となった。

表2. から表5. より、日本語母語話者被験者では、全体的に男性より女性被験者の方がIU使用率、要約率共に、ばらつきが大きく、多様な要約文章を生成する傾向があった。

英語母語話者被験者では、これとは反対に、男性被験者の方が、IU使用率、要約率共に、ばらつきの大きい多様な、要約文章を生成する傾向があった。

しかし、今回の研究では、20代女性被験者は日本語母語話者、英語母語話者共に、他のカテゴリーの被験者と比べると、ばらつきの少ない要約文章を生成する傾向にあった。

§5. まとめ

香取 (2004a) による「被験者へのインタビューと要約文分析の結果から、日本語母語話者被験者はオリジナルの文章を使用しながら「大意」型の要約を行う傾向にあるが、英語母語話者被験者はオリジナルの文章を理解し、他の言葉で言い換える「要旨」型の要約を行う傾向があるということが判明している。

本研究においては、被験者を40名から60名に増やし、母語の他に、年代別と性別というカテゴリーを加え、被験者の特徴を検証した結果、以下のような傾向がみられた。

日本語母語話者被験者はそれぞれのカテゴリーにおける、IU使用率のばらつきが比較的大きいものに対し、要約率はほぼ一定となる傾向がみられた。一方、英語母語話者被験者は、IU使用率のばらつきは小さくほぼ一定であるのに対し、要約率は多様でばらつきがあった。これは、英語母語話者被験者は、オリジナルテキストを自

分の言葉で言い換える「要旨」型の要約を行うという傾向から、要約率を計算しながら要約文章を生成するのが比較的困難であるためとも考えられる。

一方、日本語母語話者被験者は、オリジナルテキストの文章やIUをそのまま使用する「大意」型の要約を行う傾向にあるため、要約率の計算を行いながら、要約文章を生成する作業が比較的容易であると予想される。

また、今回の調査で、日本語母語話者被験者は、全体として、IU使用率のばらつきが大きい結果となったが、20代女性日本語母語話者被験者だけはIU使用率にばらつきが少ない、画一的な要約文章を作成する傾向があった。それに対し、30代女性日本語母語話者被験者はばらつきの大きい多様な要約文章を生成する傾向にあり、日本人被験者全体としては、男性よりも女性被験者の方が、ばらつきの大きい多様な要約文章を生成するという特徴がみられた。

また、年代別では、年代が高くなるに従い、IU使用率、要約率双方共に、ばらつきが大きくなる傾向がみられた。

英語母語話者被験者に関しては、20代男性被験者と30代女性被験者が比較的ばらつきの多い要約文章を生成していた。

今回の研究においては、20代女性被験者は、日本語母語話者、英語母語話者共に他のカテゴリと比べ、より画一的な要約文章を生成しているが、30代女性被験者は、より多様な要約文章を作成する傾向があった。換言すると、女性被験者は、年代ごと特徴的な要約文章を生成していたと言える。一方男性被験者は、女性被験者に比べ、ばらつきの少ない要約文章を生成していた。特に30代・40代以上の日本語母語話者男性被験者は、要約率、IU使用率とも全被験者の中で高い傾向にあり、オリジナルテキストの文章や表現をそのまま使用し、ばらつきの少ない要約文章を作る傾向が強かった。

英語母語話者被験者に関しては、年代が高くなるにつれて要約率のばらつきは、小さくなる傾向があった。しかし、40代以上の男性被験者には、若干ではあるが日本語母語話者の40代男

性被験者と同様、20代・30代の被験者より、IU使用率が高くなる傾向がみられた。

香取(2006)では、オリジナルテキストと要約文章の長さの関係についての検証を行い、要約文章の長さオリジナルテキストの長さの関係は、対数関係にあると予想される結果を得た。検証の対象となった被験者の中で、要約率が低い要約文を生成していたのは、20代前半の被験者であり、インタビューの結果、「日常的に(ほぼ毎日1時間以上)インターネット等を使用している」という回答を得た。

インターネットの検索画面では、検索結果が最大でも約135文字以内で表示されることが多い。今回の調査において、要約率・IU使用率が共に低かった、20代被験者に対するこれらの影響については、今後研究を続けていく予定である。

また、本研究では使用母語に関わらず20代女性被験者はIU使用率が低い傾向にあったが、今回要約のためのオリジナルテキストとして使用した文章は、英文・日本文共に新聞の社説である。インタビューにより、20代女性日本語母語話者被験者からは、「社説は殆ど読まない」という回答を得ている。これに対し、40代男性日本語母語話者被験者からは「新聞の社説欄には良く目を通す」という回答を複数得ている。これらのインタビュー結果と要約文章生成プロセスとの関係についても現在、検証を続けている。

謝 辞

本稿に貴重なご助言を下さった紀要叢書委員の方々にお礼申し上げます。

(2008年6月16日受付、2008年7月4日受理)

注

- 1) IUとは門田、野呂(2001, p.300)によると、統語的な単位であり、統語的尺度の捉え方により様々な種類がある。その単位の再成立をもって読解力の指標とする分析方法がアイデアユニット (IU) 分析である。
- 2) 日本文：2003年9月11日版 朝日新聞 社説『戦争は解決でなかった』。
英文：2003年9月12日版International Herald Tribune、

3) 例：J20's F 日本人被験者20代女性

参考文献

- Bernhardt, E.B. and Kamil, M.L.(1995). Interpreting relationships between L1 and L2 reading: Consolidating the linguistic threshold and the linguistic interdependence hypotheses. *Applied Linguistics*, 16, 15-34.
- Carrell, P. L.(1985). Facilitating ESL reading comprehension by teaching text structure. *TESOL Quarterly*, 19, 727-752.
- Carrell, P.L.(1991). Second languagereading: Reading ability or language proficiency? *Applied Linguistics*, 12, 159-179
- Clarke, M.A.(1980). The short circuit hypothesis of ESL reading -Or when language competence interferes with reading performance. *Modern Language Journal*, 64, 203-209
- van Dijk, T. A., & Kintch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.
- Endres, B.N.(1998). *Summarizing Information*. Berlin / Heidelberg : Springer-Verlag
- 畑山満美子, 松尾義博, 白井論 (2002) 「重要語句抽出による新聞記事自動要約」自然言語処理学会論文誌 『自然言語処理』 4, 55-73.
- Ikeno, O. (1996). The effects of text -structure-guiding question on comprehension of texts with varying linguistic difficulties. *JACET Bulletin* (27), 51-68.
- Inderjeet, M. (2001). *Automatic Summarization*, Amsterdam / Philadelphia: John Benjamins. (=奥村学, 難波英嗣, 上田禎子訳 2003 『自動要約』 共立出版)
- Jing, H. and McKeown,K.(2000). Cut and paste based text summarization. *In Proc. Of the 1st Meeting of the North American Chapter of the Association for Computational Linguistics*, 178-185.
- 香取真理 (2004a) 『異言語母語話者の要約アルゴリズムに関する比較研究』立教大学大学院異文化コミュニケーション研究科 修士論文
- 香取真理 (2004b) 「要約アルゴリズムにおける情報処理アプローチに関する考察ー日本語母語話者, 英語母語話者の場合ー」立教大学大学院異文化コミュニケーション研究科 『異文化コミュニケーション論集』 2, pp.85-94
- 香取真理(2006) 「文章要約プロセスに関する数理解析」日本シミュレーション学会 第25回大会発表論文集 pp.115-118
- Kay,D.S., & Black, J.B.(1986). Explanation-driven processing in summarization: The interaction of content and process. In J.A.Galambos, R.P.Abelson, & J.B.Black (Eds.), *Knowledge structures*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Kimura, Y. (1999). Metacognitive awareness training and reading comprehension of Japanese EFL learners. *JACET Bulletin*, 30, 45-58.
- Lehnert, W. G. (1982). Plot units: A narrative summarization strategy. In W. G. Lehnert & M. H. Ringle (eds.), *Strategies for natural language processing*, (pp. 375-412). Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Main, I. (2001). *Automatic Summarization*. Amsterdam / Philadelphia: John Benjamins.
- 益岡隆志, 仁田義雄, 郡司隆男, 金水敏 (1997) 『岩波講座 言語の科学 5 文法』岩波書店
- 増山繁, 山本和英 (2002) 「テキスト自動要約における新たな展開と展望ー統計的方法, 換言処理, そして・・・」情報処理学会論文誌 『情報処理』 43 (12) , pp.1310-1316
- 呂本俊亮 (1998) 『文章理解についての認知心理学的研究ー記憶と要約に関する実験と理解過程のモデル化ー』風間書房
- 長尾真, 黒橋禎夫, 佐藤理池, 池原悟, 中野洋 (1998) 『岩波講座 言語の科学9言語情報処理』岩波書店
- Omanson, R.C.(1982). The relation between centrality and story category variation. *Journal of Verbal Learning and Verbal Behavior*, 21, 326-337
- 佐久間まゆみ 編 (1989) 『文章構造と要約文の諸相』くろしお出版
- 佐久間まゆみ 編 (1994) 『要約文の表現類型』ひつじ書房
- Thorndyke, P.W.(1977). Cognitive Structures in comprehension and memory of narrative discourse. *Cognitive Psychology*, 9, 77-110

付録 1.

(1) 日本語オリジナルテキスト

ビンラディン様。私の息子は大きな夢を持ち、アメリカで懸命に生きていました。なぜ、あなたは罪もない人々を巻き添えにしたのか。

世界貿易センタービルに対するテロ攻撃で一人息子を失った白鳥晴弘さんは、首謀者とされるビンラディンあての手紙にそう書いた。パキスタンに駐在するアラビア語放送の記者に「本人は無理でも、できるだけ近い人に届けてほしい」と託した。深い憤りと悲しみと無数の「なぜ」を残したあ

の日から、もう2年である。

世界は変わった。アフガニスタンでアルカイダの拠点をつぶし、タリバーン政権を打倒したブッシュ米政権は、次の目標としてイラクを攻撃し、政権を転覆させた。大量破壊兵器をテロリストに渡さないための「対テロ戦」とされた。

テロとの関連が明確で、国際社会の広い支持を得たアフガニスタンでは戦後いち早く暫定政権ができ、復興が動き出した。

ところが、イラク戦争の状況は悪化するばかりだ。国連安保理と欧州の主要同盟国を振り切って開戦し、何とか制圧したものの、米英軍への攻撃がやまない。国連事務所さえテロの標的となった。

大量破壊兵器は見つからず、民衆の生活もままならない。フセイン政権からの解放という成果を帳消しにしかねない混乱の中で、イラクが「憎悪の思想に」支えられた新たなテロの巣窟になったかのような事態が出現している。

力で世界を牛耳ろうとする米国への憎しみ。その背景にある貧困。泥沼のパレスチナ問題。同時多発テロの温床として、多くのことが語られてきた。

だが、国際社会は、テロをどう撲滅できるのか今でも立ちすくんでいる。米国がその軍事力に物を言わせて戦争に勝っても、テロをなくせるわけではなかった。

様々な手段でテロを抑え込みながら温床をなくしていくという、長くたゆまぬ努力が基本ということではあるまいか。

白鳥さんは、子息の遺産や補償金を基にアフガニスタンの子どもたちを助けるプロジェクトを立ち上げようとしている。「テロをなくすには、そこから始めるしかない」と考えたからだ。

国際社会もそんな思いに応じてほしい。最も大事なのは緊密な協調態勢だろう。米国はこの原点に返るべきだ。欧州とも、アラブ・イスラム諸国とも連携してテロ撲滅のための戦略を立て直すことだ。

パレスチナやインドネシア、チェチェンと、それぞれ異なる政治的背景を持つテロでも、一つのテロが別な地のテロを呼ぶ。とくにパレスチナ紛争の政治解決はテロ撲滅の視点からも急務である。

テロリストの根を絶つために、イスラム諸国自

身の民主化努力も求められる。テロのない世界への新たな出発点として、この9・11を位置づけたい。

(2) 英語オリジナルテキスト

Two years after a gang of reactionary religious zealots flew commercial airplanes into the World Trade Center and the Pentagon, President George W. Bush commonly speaks about a diffuse global war on terrorism, and a former CIA chief, James Woolsey, has even described the conflict as World War IV.

There is a danger in this imprecise way of naming an enemy and invoking a worldwide struggle of indefinite duration. The danger is that repetition of these misleading definitions will be used to rationalize an antiterrorist strategy that embroils Americans unnecessarily in other countries' domestic power struggles. To fall into this trap is to play into the hands of Osama bin Laden and his associates, who would like to provoke the global holy war they preach.

Instead of loose talk that acts like America's moral enemy is an abstract noun—"terrorism" in all its forms and manifestations—Bush would be wise to distinguish Al Qaeda and the groups affiliated with it from Islamist movements that may be trying to overthrow regimes in their own countries but have not declared war against the United States.

When U.S. intelligence agents and armed forces turn up in countries such as Mauritania, Niger, Chad, Mali, Djibouti or Uzbekistan, they appear to be waging the global war against all Islamists that bin Laden invokes, and Washington appears to be validating the grandiloquent ideological claims of Al Qaeda.

There is a genuine need for intelligence and law-enforcement cooperation against Al Qaeda and its affiliates—American's declared enemies. But a promiscuous entanglement in the internecine conflicts of countries ruled by

vicious dictators risks a strategic blunder.

Pursuing Bush's indiscriminate war on terrorism, U.S. military and intelligence personnel are currently on the ground, cooperating with the corrupt and repressive regimes of, among others, Islam Karimov in Uzbekistan and Maaouyah Ould Sid Ahmed Taya of Mauritania. These undemocratic rulers exaggerate the threat of small Islamist groups in their lands to excuse thuggish suppression of all dissent.

If Washington allows itself to be identified with these and other despotic clients in the war on terrorism, opposition movements against despised local rules could be transformed into anti-imperialist struggles against America, as the reviled foreign power behind

the local tyrant and his torturers.

This is precisely what bin Laden and his Egyptian partner Ayman al-Zawahiri allege in their statements for public consumption—that "crusader America" is the "far power" standing behind collaborationist regimes in Muslim countries that the Al Qaeda ideologues define as the "near power."

Nothing would better suit Al Qaeda's recruiting tactics than to be able to point to U.S. backing for the Karimovs of the Muslim world as proof that America is at war with all Muslims, not merely the fanatics of Al Qaeda who target Americans. Bush must not nourish that fantasy by confusing a campaign against one aberrant Islamist faction with a world war against all terrorists.

付録2. 日本語テキストにおけるIU分割表

1	ビンラディン様。	41	事態が出現している。
2	私の息子は大きな夢を持ち、	42	力で世界を牛耳ろうとする
3	アメリカで懸命に生きていました。	43	米国への憎しみ。
4	なぜ、	44	その背景にある貧困。
5	あなたは罪もない人々を巻き添えにしたのか。	45	泥沼のバレスチナ問題。
6	世界貿易センタービルに対するテロ攻撃で	46	同時多発テロの温床として、
7	一人息子を失った白鳥清弘さんは、	47	多くのことが語りられてきた。
8	首謀者とされるビンラディンあての手紙こそ書いた。	48	だが、
9	パキスタンに駐在するアラビア語放送の記者に	49	国際社会は、
10	「本人は無理でも、できるだけ近い人に届けてほしい」	50	テロをどう撲滅できるのか
11	と託した。	51	今でも立ちすくんでいる。
12	深い憤りと悲しみと	52	米国がその軍事力に物を言わせて
13	無数の「なぜ」を残したあの日から、	53	戦争に勝っても、
14	もう2年である。	54	テロをなくせるわけではなかった。
15	世界は変わった。	55	様々な手段でテロを抑え込みながら
16	アフガニスタンでアルカイダの拠点をつぶし、	56	温床をなくしていくという、
17	タリバン政権を打倒したブッシュ米政権は、	57	長けたゆめめ努力が基本ということではあるまいか。
18	次の目標としてイラクを攻撃し、	58	白鳥さんは、
19	政権を転覆させた。	59	子息の遺産や補償金を基に
20	大量破壊兵器をテロリストに渡さないための	60	アフガニスタンの子どもたちを助ける
21	「対テロ戦」	61	プロジェクトを立ち上げようとしている。
22	とされた。	62	「テロをなくすのは、そこから始めるしかない」
23	テロとの関連が明確で、	63	と考えたからだ。
24	国際社会の広い支持を得たアフガニスタンでは	64	国際社会もそんな思いにこたえてほしい。
25	戦後いち早く暫定政権ができ、	65	最も大事なのは緊密な協働態勢だろう。
26	復興が動き出した。	66	米国はこの原点に戻るべきだ。
27	ところが、	67	欧州とも、
28	イラク戦争の状況は悪化するばかりだ。	68	アラブ・イスラム諸国とも連携して
29	国連安保理と欧州の主要同盟国を振り切って開戦し、	69	テロ撲滅のための戦略を立て直すことだ。
30	何とか制圧したものの、	70	バレスチナやインドネシア、
31	米英軍への攻撃がやまない。	71	チェチェンと、
32	国連事務所さえテロの標的となった。	72	それぞれ異なる政治的な背景を持つテロでも、
33	大量破壊兵器は見つからず、	73	一つのテロが別な地のテロを呼ぶ。
34	民衆の生活もままならない。	74	とくにバレスチナ紛争の政治解決は
35	フセイン政権からの解放という	75	テロ撲滅の視点からも急務である。
36	成果を帳消しにしかない混乱の中で、	76	テロリストの根を絶つために、
37	イラクが	77	イスラム諸国自身の民主化努力も求められる。
38	「憎悪の思想」	78	テロのない世界への新たな出発点として、
39	に支えられた	79	この9・11を位置づけたい。
40	新たなテロの巣窟になったかのような		

付録3. 英語テキストにおけるIU分割表

1	Two years after	51	But a promiscuous entanglement
2	a gang of reactionary religious zealots flew commercial airplanes	52	in the interecine
3	into the World Trade Center and the Pentagon	53	conflicts of countries ruled by vicious dictators
4	President George W. Bush commonly speaks	54	risks a strategic blunder
5	about a diffuse global war on terrorism	55	Pursuing Bush's indiscriminate war on terrorism
6	and a former CIA chief	56	U.S. military and intelligence personnel are currently
7	James Woolsey	57	on the ground
8	has even described the conflict as World War IV	58	cooperating with the corrupt and repressive regimes
9	as World War IV	59	of among others
10	There is a danger	60	Islam Karimov in Uzbekistan and Maouyah Ould Sid Ahmed Taya of Mauritania
11	in this imprecise way	61	These undemocratic rulers exaggerate the threat
12	of naming an enemy	62	of small Islamist groups
13	and invoking a worldwide struggle	63	in their lands
14	of indefinite duration	64	to excuse thuggish suppression
15	The danger is that repetition	65	of all dissent
16	of these misleading definitions will be used	66	Washington allows itself
17	to rationalize an antiterrorist strategy	67	to be identified with these and other despotic clients
18	that embroils Americans unnecessarily	68	in the war on terrorism
19	To fall into this trap is	69	opposition movements
20	to play into the hands of Osama bin Laden and his associates	70	against despised local rules
21	who would like	71	could be transformed
22	to provoke the global holy war	72	into anti-imperialist struggles against America
23	they preach	73	as the reviled foreign power
24	Instead of loose talk	74	behind the local tyrant and his torturers
25	that acts like America's moral enemy is an abstract noun-	75	This is precisely
26	terrorism	76	what bin Laden and his Egyptian partner Ayman al-Zawahiri allege
27	in all its forms and manifestations-	77	in their statements
28	Bush would be wise	78	for public consumption
29	to distinguish Al Qaeda and the groups affiliated with it	79	-that
30	from Islamist movements	80	crusader America
31	that may be trying	81	is the
32	to overthrow regimes	82	far power
33	in their own countries	83	standing behind collaborationist regimes
34	but have not declared war against the United States	84	in Muslim countries
35	When U.S. intelligence agents	85	that the Al Qaeda ideologues define
36	and armed forces turn up in countries such as Mauritania	86	as the
37	Niger	87	near power
38	Chad	88	Nothing would better suit Al Qaeda's recruiting tactics
39	Mali	89	than to be able to point to U.S.
40	Djibouti	90	backing for the Karimovs
41	they appear	91	of the Muslim world as proof
42	to be waging the global war	92	that America is at war with all Muslims
43	against all Islamists	93	not merely the fanatics of Al Qaeda
44	that bin Laden invokes	94	who target Americans
45	and Washington appears	95	Bush must not nourish that fantasy
46	to be validating the grandiloquent ideological claims of Al Qaeda	96	by confusing a campaign
47	There is a genuine	97	against one aberrant Islamist faction
48	need for intelligence and law-enforcement cooperation	98	with a world war against all terrorists
49	against Al Qaeda and its affiliates		
50	-American's declared enemies		

A Comparative Study of the Human Summarizing Process Among Different Cultural Backgrounds

Mari Katori

Abstract

The human system of summarizing a text is one of the methods of natural language processing. It is considered to be a process of reconstructing text. This paper examines the differences in the human system of summarizing between people of different cultural and language backgrounds. In this research, a sample of subjects consisting of 30 Japanese speakers and 30 English speakers, were requested to summarize a text in their mother tongue. Then, each summarized text was analyzed by the author noting their use of mother tongue, age and gender.

Some differences were found between Japanese and English speakers in their summarizing methods. Specifically, the majority of native Japanese speakers summarized each paragraph closely following the original text, maintaining the original words and expressions. In contrast, most of the native English speakers constructed the summarized text by paraphrasing, not necessarily keeping the organization of the original paragraph intact. Additionally, differences between age and gender, noted as well.