5

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…ー

増山　繁
豊橋技術科学大学知識情報工学系　masuyama@tutkie.tut.ac.jp

山本 和英
長岡技術科学大学電気系　yamamoto@fw.ipsj.or.jp

はじめに

現在，インターネット等を通じて大量の情報が溢れており，目まぐるしく変化する状況に対応して素早い意思決定を行うためには，大量の情報の中から必要な情報を的確に取捨選択することが必須である．したがって，重要な情報のみを選択して提供することにより，要点を迅速に把握することを助けるテキスト自動要約は，人間の知的活動を支援する基礎技術としてますます重要性を増している．それに伴い，テキスト自動要約に関して，内外で活発に研究されるようになってきた[11]．人間であれば，要約を作成するとき，まず，内容を理解し，次に，理解した内容に基づいて新たに文章を作成する．しかしながら，現状の計算機による意味理解に関する技術は，人間が要約を行うのと同様の過程を行うのに十分なほどには発展していない．そこで，現状の技術で可能な限りで要約を作成する技術が開発されてきた．

少し前までは，テキスト自動要約に関する研究は，1950年代後半のLuhnを先駆けとする，重要文抽出に関するものが多かったが，現在では，日本語を対象とするものでも，形態素解析ツールのJUMAN，茶筌等，構文解析ツールのKNP等の普及（本誌2000年11月号，「特集：使いやすくなった自然言語処理のフリーソフトー知っておきたいツールの中身ー」参照）により，重要文抽出のみでなく，構文解析結果を積極的に利用した文内要約も行われるようになった．それらは，要約規則を人手で作成しているものが多い．一方，計算機の処理速度と記憶容量の急激な増大，および，大量の機械可読テキストが利用可能になってきたことから，音声認識や情報検索で比較的早くから用いられており，最近自然言語処理でもよく用いられるようになってきた統計的方法の利用が注目されるようになってきた．そこで，それについて，次章で解説する．

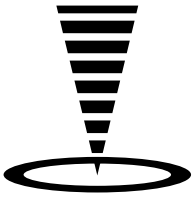
現状のテキスト自動要約は，重要文抽出に典型的に見られるように，「重要部分」を何らかの方法で選択して，または，それと本質的には同じであるが，重要でない部分を削除して要約を作成しているものが大部分を占める．一方，先に述べたように，人間の作成した要約は，内容を理解した上で，頭の中で整理して，もう一度文を作成するという手順を踏んでいる．そこで，より自然な要約を生成するためには，このような原文の一部を抜き出してそのまま構成する以上のより深い解析に基づく要約作成手法を考える必要がある．そのような要約手法の1つとして重要である換言について，引き続き紹介する．

最後に，将来の社会における知的活動支援のための基本的要素技術としての観点からテキスト自動要約について今後研究を進めるべき方向を展望する．

統計的方法

統計的方法は，自然言語で規範とすべきものは膨大な用例であるとの立場に立つものである．人間の場合，たとえば英語を勉強するとき，和文とその英訳を収集した対訳集があれば，それを暗記することで類似の文例が出題されれば試験で合格点をとることができる．それと同様に，要約文と原文の対応コーパスが十分な量あれば，決定木等の教師ありの統計的機械学習手法の適用が可能である[11]．今後，対応コーパスが利用可能になってくるとともに，次第に教師ありの機械学習手法が広く用いられるようになると思われる．しかしながら，性能の良いテキスト自動要約を行うためには対応コーパスによる学習は分野ごとに行う必要がある．また，言語は生き物であり，用語，表現法等は常に変化していく．さらに，対応コーパスは作成するのに膨大な人手と費用を要するため，将来にわたっても対象とする応用分野で必ずしも利用可能とは限らない．そこで，テキスト自動要約を対応コーパスでない，一般のコーパスを用いて行う統計的方法も提案されてきている．人間の場合，対訳集がなくても膨大な用例に触れて，そこから何らかの法則性（傾向）を学ぶことで言語を習得することができる．文章の達人になるには，大量の読書が役に立ち，外国語を習得するには，一定期間，その言語に浸り切るのが有効である．それと同様に，計算機でも大量のコーパスがあれば統計的方法で有用な言語知識を獲得できる可能性がある．本章では，まず，対応コーパスが利用可能な場合の統計的方法を，次に，対応コーパスが利用可能でない場合でも使える統計的方法を紹介する．なお，対象を日本語とするものに絞った．

まず，日本語における対応コーパスを用いた統計的方法による要約の初期の試みとして，加藤，浦谷[5]は，NHKテレビニュース原稿とそれに対応する文字放送原稿を使用して，局所要約知識の自動獲得を行っている．局所知識は，原稿と要約文の差分としての置換規則と，その置換規則が適用できるかどうかを決める置換条件からなっている．ここでいう置換は，次章の換言に当たる．置換規則はDPマッチングにより最適な単語対応を求めることにより獲得する．一方，置換条件は置換規則が適用できる可能性のある文字列の前後それぞれn文字の組を，それぞれ正例，負例の区別を付けて並べたものである．それに最も近い条件が正例である場合には置換可能，そうでなければ置換不可能と判断する．この手法により，アメリカ→米，総理大臣→首相，外務大臣→外相，など，典型的な短縮置換が取得できていると報告されている[5]．



文献7）では，SVM（Support Vector Machine）を重要文抽出に適用した結果が報告されている．SVMは，他の教師あり学習法に比べ比較的少ない学習データで高い精度を達成できるという特長がある．学習用データとしては，TSC1（本特集，福島らの記事参照）の重要文抽出タスクの正解データを用いている．SVMは，解析時間がかかるが，実時間性を要求されない応用では，将来，正解データが蓄積されてくるに従い，有望な手法の1つとなるであろう．

対応コーパスが存在しない場合の統計的要約法として，堀，古井らは，音声認識結果を対象とし，要約を，ある評価関数を最大化することで認識単語列から単語列を適切な要約となるよう抜き出す問題としてモデル化し，動的計画法により解を求める一連の研究を行っている[6]．つまり，テレビニュース，ないし，講演の音声認識結果の各発話文から相対的に重要な単語を，指定した要約率で抽出し，それらを接合することで要約文を生成する自動要約の枠組みを提案している[6]．そこでは，次のスコアを用いて要約の候補となる単語列に対する評価関数を与え，動的計画法を用いてその評価関数を最大化する単語列を要約としている．すなわち，原文における相対的な単語の重要度を示す単語スコア，要約文内の単語連鎖の適正度を表す言語スコア，認識結果に含まれる認識誤りを除外するため，音響的，言語的に信頼度の低い単語が要約文に含まれないようにするための信頼度スコア，係り受け関係にない単語連鎖にペナルティを与える単語間スコアを用いる[6]．

次に，文中の重要でない部分を統計的方法で認定することで削除する要約手法を紹介する．テキスト自動要約は，一般に複数の手法を組み合わせて行うことが多い．文内の一部分を削除することはそのための重要な要素技術の1つである．そこで，動詞を含んだ節の削除可能性を決定する統計的方法を紹介する．

これまで，文内の非重要部分の削除の研究は，ヒューリスティクスを用いたものが多かった．たとえば，TSC1に参加して良好な成績を収めた要約システムYELLOW[9]では，人手で作成した36個の規則を用いて名詞に対する2重修飾の削除を行っている．しかしながら，そのような規則を作成するには膨大な時間と手間がかかる．また，網羅性を保証するのは難しい．しかも，

そのような作業には，深い言語知識と高い言語能力を持った人材が必要である．しかしながら，そのような人材はいつでも確保できるとは限らない．それに対し，統計的方法によると，大量のコーパスから知識を抽出するので，比較的容易に，一定の水準の結果を得ることができる可能性がある．

文章を読んでいて，省略されやすいところと省略されにくいところがあるのに気づく．たとえば，ある名詞は滅多に修飾されない．したがってその名詞を修飾する必要性は高くなさそうだ．また，ある名詞はそれを修飾する動詞が限られている．したがって，動詞を含む修飾節がなくても大体どのような内容か想像がつく．以上のことから，動詞を含む連体修飾節からなる修飾節の削除されやすさは十分なコーパスがあれば統計的に定量化できそうだ．そこで，統計的手法で取り組む対象として動詞を含む連体修飾節からなる修飾節の削除問題を取り上げた．

動詞を含む連体修飾節からなる修飾節と被修飾語である名詞との関係に着目してその削除可能性を実例を見ながら考える．

[例 1]

A, B両国は首脳会談を行い，両国民が待ち望んでいた自由貿易協定を締結することで合意した．□

動詞を含んだ連体修飾節を考える．例 1 では，「両国民が待ち望んでいた」がそれに該当し，「自由貿易協定」が，その被修飾語である．

(1) 動詞を含む連体修飾節が修飾している名詞に対して，その名詞を修飾する動詞の種類が限定されているとき，動詞を含んだ連体修飾節は被修飾名詞から連想されやすいので削除しやすい．

—たとえば，「科学者にとって最高の栄誉であるノーベル賞は…」の「ノーベル賞」を修飾する語は比較的限定されている．したがって，「科学者にとって最高の栄誉である」は削除できる可能性が高い．

名詞に対する動詞による修飾の多様性は，各被修飾名詞に対して，動詞を含む連体修飾節の動詞で名詞が修飾される頻度をコーパスから取得してエントロピーを計算することで測定する．

(2) もともと修飾されにくい名詞に対する修飾語は，削除してもよいことが多い．

—これは，各名詞に対するコーパス中での修飾頻度を測定する．国名，地名等，周知の固有名詞はこれに該当することが多い．

(3) 重要な名詞は，それだけで意味が分かることが多いので，その修飾語は削除可能である可能性が高い．

—これは，被修飾名詞のIDF（たとえば，文献 2)，3) 参照）を用いて，すなわち，コーパス中で万遍なく大部分の文書に出現するよりも，特定の文書に偏って出現する語の方が重要として計る．さらに，それが複合名詞の場合，それを構成する名詞の数も考慮する．

「バルカン半島では，人々を不幸にする内戦が長く続いた」において，「内戦」だけで意味が分かるので，「人々を不幸にする」は削除できる可能性が高い．実際，「内戦」は，特定の記事に集中して出現するのでIDF値が高い．

以上のような観察に基づき，「修飾多様性」「修飾されやすさ」の指標と修飾先名詞の重要度をコーパスから算出する．一方，修飾動詞節の削除可能性を検討するのに，これまでは，非修飾名詞，および，それとその修飾動詞節との関係に着目したが，次に，修飾動詞節自身の重要度に注目する．

(4) 動詞を含んだ連体修飾節自身が重要でなければ削除しやすい．

—それを反映するため，まず，動詞を含む連体修飾の動詞にかかっている連用表現の数が多い場合，それは重要である可能性が高いことに着目する．さらに，動詞を含む連体修飾節の重要度を測る際，文脈も考慮する．すなわち，それに含まれている情報が以前の文にも含まれている場合，既知の情報なので，重要度が高くなく，削除可能と認定されやすくする．一方，それが以降の文に含まれている場合，削除可能と認定されにくくする．

文献 12) では，以上に基づく手法を提案し，その性能について予備的検討を行い，良好な結果を得た．

ある修飾表現が削除できるかどうかを類似の文例と比較して一方でその修飾表現がなければ削除できる可能性があると判定することは自然であろう．そこで，次に，削除可能な連用修飾表現の認定を行う統計的手法を紹介する．

[例 2]

1. A国とB国は，長期にわたる交渉の末，国境協定を締結した．
2. 両国は，相互不可侵条約を締結したが…□

具体的には，連用修飾表現に対し，同一の動詞に係り，同一の格助詞，類似した名詞を持つ連用修飾表現がコーパス中に存在するとき，また，被修飾動詞に対して，コーパス中でその動詞に多く係る格助詞を含むとき削除されにくくする．ただし，削除対象は，2つ以上の連用修飾表現を含む場合のみとする．これは，そのうちの一部を削除しても情報欠落が少ないからである．

上記の例2において，1.の修飾成分である「長期にわたる交渉の末」は2.において対応する修飾成分がないので，省略可能である．

なお，類似度は，コーパス内の動詞と名詞の出現に関する相互情報量から測る方法を用いた．さらに，文脈を考慮するため，その連用修飾表現に含まれている情報が以前の文にも含まれているなら，削除可能と認定されやすくし，逆に，その情報が以降の文に含まれている場合や，重要な情報が含まれている場合には，削除可能と認定されにくくなるように工夫をしている．詳細は，文献10）で報告した．

自然言語処理における統計的方法については，本格的な教科書[2)]が参考となる．また，情報検索の知識が役に立つ．たとえば，文献3）を参照されたい．

換言処理

これまでの要約処理は，入力テキストを部分的に削除（あるいは部分的に選択）することで要約を実現してきた．その理由は，最も容易に要約問題をモデル化しやすいためである．しかし，本来要約は表現削除（あるいは表現選択）問題だけではなく，表現変換問題であるはずである．ここでは，テキストを表現変換，つまり別の表現を使って言い換えることによって要約を行うことを考える．このような表現変換を実現する処理をここでは換言処理と呼ぶ．

例 3-1: この問題はそう簡単には解けそうにない
例 3-2: これは難問だ

本来ならば計算機にこのような要約を行ってほしい．例3-1から例3-2のような換言は非常に難しそうだが，以下の各処理手順に分解すれば，それぞれの言い換えは，工夫すればなんとかできそうだ．

例 3-1: この問題はそう簡単には解けそうにない
例 3-A: これはそう簡単には解けそうもない問題だ
例 3-B: これは容易でない問題だ
例 3-C: これは難しい問題だ
例 3-2: これは難問だ

◎情報の選択と圧縮

ところで，要約処理は2種類に大別できる．すなわち，1つは情報の選択で，もう1つは情報の圧縮である．大量の情報の中から一部の重要なものを取捨選択して残りを捨ててしまうのが情報の選択で，情報をできるだけ損なわずに情報の密度の薄い部分をより密度の濃い形態で置換するのが情報の圧縮である．果物を例にとれば，お店に並んでいる果物の中からおいしそうなものを選ぶこと（果物の選択）と，ジューサーで果物を絞り出して濃縮すること（果物の圧縮）に相当する．

例1で示したのは情報の圧縮である．原文の持つ情報をできるだけ失うことなく，より短く表現しようとしたものである．このように，換言技術は情報を圧縮する際に非常に重要な技術である．いわば，性能のよいジューサーである．

ところで，換言処理は情報圧縮時のみならず，情報選択時においても重要な役割を果たす．たとえばテキストから重要と思われる文を選択するとしよう．もし各文が長いと1つの文の中に重要な表現とそうではない表現が含まれている場合があり，文単位で選択するとこれを分離することができない．そこで，やむを得ず，その文を選択する（もしくは選択しない）ことになる．

このような場合，換言処理を行うことで長文は短文，あるいはより短い表現単位に分割でき，これによってより満足度の高い選択が可能になる．すなわち，換言処理はテキストの分割と統合を可能にすることで情報選択の自由度が向上する．果物の例では，1皿単位から1個単位への販売単位の変更，あるいは大きな果物を切り売りしてくれるサービスに相当する．これは便利である．

◎過去と現状

換言処理は，最近になって始まったわけではなく，以前からさまざまな研究が行われてきた．たとえば，機械翻訳の前処理，テキストの推敲／校正などは以前から行われてきた．近年になって，換言処理そのものが独立した研究対象としても認識されるようになり，聴覚障害者向け読解支援のための換言など特定の応用向けの研究も始まってきた．ここで特筆すべきは，これらの多くが日本（すなわち日本語）の研究であることである．詳しく

[原文] 遠山敦子文部科学大臣は 18 日、閣議後に記者会見を行ない、政府の地方分権改革推進会議が公立学校の教員給与の国庫半額負担を見直す中間報告を打ち出したことについて「かなりの事実誤認がある。義務教育に国が負う責務は大きい」と述べ、反対する姿勢を示した。

[要約] 文科相が 18 日の会見で、地方分権改革推進会議中間報告に反対姿勢。公立学校の教員給与の国庫半額負担の見直しについて。

[換言の途中経過]

遠山敦子文部科学大臣
→ 文部科学大臣
→ 文部科学相
→ 文科相

18 日、閣議後に記者会見を行ない
→ 18 日、閣議後の記者会見で
→ 18 日の記者会見で
→ 18 日の会見で

…会議が…を見直す中間報告を打ち出したこと
→ …会議が打ち出した中間報告
→ …会議の中間報告
→ …会議中間報告

…を見直す中間報告
→ 中間報告は…を見直す
→ 中間報告で…の見直し

…について（…と述べ）反対する姿勢
→ …に反対する姿勢
→ …に反対の姿勢
→ …に反対姿勢

図 **-1**　今後実現可能と予想される要約例

は文献 8）に譲るが，換言処理の分野では日本が最も進んでいるといっても過言ではない．

それでは現在の換言技術でどのような要約が可能となるか考えてみた．たとえば，図**-1** に示した要約例は作例ではあるが，個別の換言技術は現在の研究成果でも十分可能であり，今後このような要約が自動的に生成できる可能性は高い．最大の難関はこれら個別の換言知識を 1 人で作ることが現実的でないことであり，今後関係者全員が協力して換言知識の共有と蓄積を進めるべきである．

今後の展望

テキスト自動要約は，膨大な情報の中から真に必要とされる情報を必要なときに提供することで人間の知的活動能力を大幅に強化するための技術と位置付けられる．そのことから，これからテキスト自動要約が必要とされる応用分野として，たとえば，以下が考えられる．

(1) 携帯端末への情報配信

(2) **WWW** からのデータの要約による提示

これらについては，本特集の中川らによる記事に譲る．

(3) 聴覚障害者のための **TV** 字幕生成

これについては，本特集の江原らによる記事を参照されたい．

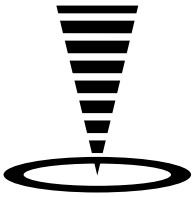
(4) 災害時における情報提供

高度成長期は，比較的地震活動が静穏な時期であったが，最近，活動期に入ったと聞く．そのことから，震災時における情報提供について研究しておく必要性が高まってきている．また，ニューヨークのテロ事件のような突発的な事件の際，高層ビル中において避難経路等に関する状況に応じた実時間の情報を提供することで被害を最小限にとどめることができる可能性がある．このような，緊急時における情報提供目的のための要約においては，重要な情報の欠落を極力避けること，また，正確な情報伝達が必要なため，曖昧さがない，子供，外国人等の情報弱者にも分かりやすい表現を用いる必要がある．前章で言及した聴覚障害者向け読解支援のための換言処理が参考となろう．

(5) 音声要約

音声要約の研究は，テレビ，ラジオの番組等の音声コンテンツを要点だけ音声のまま提供するために必要であり，高さや強さ，スピード等音声の韻律的特徴を用いて発話の強調部分を認定して選択する研究等が以前から行われていた（文献 1）など）が，日本でも最近活発に研究されるようになってきている．たとえば，日本音響学会 2002 年秋季研究発表会講演論文集だけで数編収録されている．今後，音声を対象にしていることの特長を活かし，話速変換の技術と併用するなどして，聞き取りやすく，しかも重要な点は逃さないようにするなど，よりユーザの必要性に応えるような技術を開発する必要がある．

応用を考えることは要約研究にとって非常に重要である．基礎研究の立場から見ても，応用を定めると，評価が明確になり，どのようなことを重点的に研究すべきか



が見えてくる．また，応用ごとに望ましい要約が異なることが多く，それぞれ固有のシステムの開発が必要な場合が多い．そのいずれの場合でも，本稿で紹介した統計的手法，および，換言手法は基本的な要素技術として有効と思われる．

基礎技術面については，たとえば，以下が重要と考えられる．

(1) 複数文書要約

複数文書要約においては，また，イベントを抽出して，それを時系列的に並べるなどの技法が提案されている．しかしながら，要約技術として見た場合，そもそも，時間的経過の要素を除いて，単一文書の要約と複数記事の要約との間に，本質的な差があるのか？ コーパスサイズの大きさ，および，重複部の多さなどの，単に量的な差ではないのか？ この点は，十分見極める必要がある．その上で，複数文書要約のための技術を開発していく必要がある．なお，複数文書要約についての詳細は，本特集の難波らの記事や，サーベイ[11]に譲る．

(2) 統計的方法

本稿で紹介した統計的方法は，文中の削除可能な部分を認定するためのものであった．しかしながら，より自然な要約を生成するためには，このような情報の選択に基づく手法のみでなく，前章で述べた「情報圧縮」を行う統計的要約手法を開発する必要がある．そのための試みの1つとして統計的方法を換言の取得に適用することが考えられる．先に紹介した文献[5]は，対応コーパスを用いた換言取得の研究として興味深い．複数の文に対して1つの文からなる要約を作成するための統計的方法を開発することも重要な課題である．数十万以上の語彙を持ち，しかも，語彙自身が時間とともに変遷していく自然言語においては，十分な統計をとれるだけのコーパスが用意できないというデータスパース性の問題は常につきまとう．したがって，まず，基礎的研究として，統計的方法によってどこまで良い要約の作成が可能かの限界を見極める必要がある．それとともに，辞書，シソーラス等の知識や，言語学の知見等に基づくヒューリスティックスを併用することにより，より良い要約を生成する手法を開発することが重要である．

(3) 換言

なぜ換言処理が最近になって注目されるようになったのかを考える．伝統的な自然言語処理の捉え方では，ある表現に対してその表現の意味を理解するという考え方が主流であったと思う．すなわち，まず「意味」（あるいはフレーム，述語論理など）という名の記号集合を定義し，自然言語表現という記号集合から意味という記号集合への記号変換を行うことで自然言語を理解するとし，この変換結果を何らかの方法で加工することで要約を行おうとしてきた．ところが現実世界で流通する膨大なテキストを対象とする場合は，

(a) 表現と意味の間の相互変換が難しい，

(b) そもそもあらゆる自然言語表現に対応できる意味記号の定義が難しい，

という理由でこのアプローチには勢いがない．

一方，換言処理は自然言語表現という記号集合相互の変換処理である．これにより，先ほど挙げた問題のうち，意味記号の定義をする必要がないため少なくとも問題 (b) は存在しない．一方 (a) についても言語→意味→言語と2度変換を行うよりも言語→言語のほうが簡単そうに見える．さらに人工的な記号は皆が満足することは難しいが，換言知識は言語表現間の対応であるため，いったん作ってしまえば言語資源としての価値や流通性は高い．以上により言語の意味は言語で表現したほうが結局は都合がよいという考え方が徐々に支持を集めてきており，換言処理の盛り上がりもこの流れの1つであると思う．今後は換言知識の蓄積と共有が進むであろうし，またそうなることを願っている．

(4) 意味理解の結果を用いた文生成

人間の要約過程により近い要約を実現する上で，長期的には重要だが，上記 (3) でも述べたように，現時点では難しい．特に，意味を表現するためのモデルが，コンピュータ上に載せるという観点から見た場合，永い伝統を持つ論理学に属するもの，および，1960年代後半から1970年代に提案された本質的には等価であるQuillianの意味ネットワークとMinskyのフレーム以来，大きなブレークスルーがなかったといえるかもしれない．むしろ，要約のために必要な意味理解とは何か，また，そこから見たときの有効な意味表現とは何かを追求することで，意味理解の研究に貢献できる可能性がある．

(5) 要約読者の要求を反映した要約

人間が作成する要約は十人十色である．この理由は，要約作成の経験やスキルをどの程度持っているかが異なるからというのもあるが，そもそも原文のど

こに興味を持ってどこを重要と考えるかが読者によって異なっているためである．あるとき突然テキストを渡されて，「これを要約してください」と言われても，誰が何のために読む要約なのかが分からないと困らないだろうか．このように，本来要約は目的志向（誰かが何かの目的のために読むから要約を作成する）なので，計算機によって要約する際もこれらの情報を与えなければ，要約読者の満足するものはできないはずである．

以上をやや大きく捉え，読者の要求を反映した要約と考えると，最も実現に近づいているのは「要約率」であり，多くの研究において要約率を可変にできるような枠組みがとられている．これ以外については，一部の研究でモデルの提案が行われているが，まだ研究が進んでいるとは言いがたい．この理由を考えると，そもそも読者の要求とは何か，あるいは，これをどう表現するかというのが難しい．個別の事例では，たとえば巨人－阪神戦で巨人と阪神のどちらに興味があるか，などのように容易に定義可能なものがあるが，これを一般化するのはきわめて困難である．あるいは読者がテキストに関してどの程度の知識を持っているかによって要約は変化するはずであるが，個別の知識の有無を計算機に逐一入力させるのは不可能である．

以上のように実現には多く課題を残しているが，(4) で述べた意味の理解の問題とも同様，最終的には避けて通れない問題であるので，工学的には実世界に有用なモデルを部分的にうまく切り出すことで要約に反映させるかたちで模索していく必要があるだろう．

(6) 言語学，心理学，脳科学等の知見を取り入れた手法の考案

これらで得られた知見は，必ずしも計算機に載せるのが容易とは限らないが，うまく取り入れることで，より良い要約ができる可能性がある．一方，たとえば，工学的に考案された手法に，言語学的根拠を与えることができれば興味深い．

なお，以上の方法は，排他的なものではなく，たとえば，統計的手法と換言を併用することで，より良い結果が得られると思われる．また，要約結果の評価法の確立も非常に重要であるが，これは，本特集の福島らの記事に譲る．

多忙な現代人にとって，Readers Digest誌のように，1冊の単行本を1つの記事に縮めてくれる要約システムがあるとありがたい．しかしながら，そこに至るまでにはまだまだ解決しなければならない課題が多い．その時点で利用可能な技術を駆使することで，できる限り望ましい要約を生成し，直面する要求に応える必要があることはいうまでもない．しかしながら，それにとどまらず，機械翻訳が自然言語処理技術の発達に大きく貢献したように，テキスト自動要約の研究から自然言語処理における新しい技術が生まれることを願ってやまない．

本稿，特に，今後の展望で述べた考えは，言語処理学会第4回年次大会ワークショップのパネルディスカッション（1998），東三河開発懇話会 産学官交流サロン講演「情報検索とテキスト要約－インターネット時代の情報洪水を生き抜く－」（1998），軽井沢ワークショップの招待講演[4]（1999），第3回西日本地域国立高等専門学校協会技術職員特別研修（情報系）講演，「情報検索，自動テキスト要約，情報抽出－情報洪水は宝の山－」，豊橋技術科学大学マルチメディアセンター（2000）等で次第にかたち作られてきたものである．これらの機会を与えていただいた方々に深謝の意を表したい．

参考文献

1) Chen, F.R. and Withgott, M.: The Use of Emphasis to Automatically Summarize a Spoken Discourse, Proc. ICASSP-92, pp.I-229-232 (1992).
2) Manning, C.D. and Schütze, H.: Foundations of Statistical Natural Language Processing, The MIT Press (1999).
3) Baeza-Yates, R. and Ribeiro-Neto, B.: Modern Information Retrieval, Addison Wesley (1999).
4) 増山　繁: テキスト自動要約技術の現在とその展望（招待講演），第12回 回路とシステム（軽井沢）ワークショップ講演論文集, pp.493-498 (1999).
5) 加藤直人，浦谷則好: 局所的要約知識の自動獲得手法，自然言語処理, Vol.6, No.7, pp.73-92 (1999).
6) 堀　智織，古井貞熙: 講演音声の自動要約の試み，話し言葉の科学と工学ワークショップ講演予稿集，2001年2月28日～3月1日, pp.165-171 (2001).
7) 平尾　努，前田英作，松本裕治: Support Vector Machineによる重要文抽出，情報処理学会研究報告，情報学基礎 63-16 (2001).
8) 乾健太郎: 言語表現を言い換える技術，言語処理学会第8回年次大会チュートリアル資料，pp.1-21 (2002).
9) 大竹清敬，岡本大吾，児玉　充，増山　繁: 重要文抽出，自由作成要約に対応した新聞記事要約システムYELLOW，情報処理学会論文誌「データベース」，Vol.43, No.SIG2 (TOD13), pp.37-43 (2002).
10) 酒井浩之，篠原直嗣，増山　繁，山本和英: 連用修飾表現の省略可能性に関する知識の獲得，自然言語処理，Vol.9, No.3, pp.42-62 (2002).
11) 奥村　学，難波英嗣: テキスト自動要約に関する最近の話題，自然言語処理，Vol.9, No.4, pp.97-116 (2002).
12) Sakai, H. and Masuyama, S.: Unsupervised Knowledge Acquisition about the Deletion Possibility of Adnominal Verb Phrases, Proceedings of the Workshop on Multilingual Summarization and Question Answering,（COLING2002 Post-Conference Workshop）, pp.49-56 (2002).

（平成14年10月30日受付）