

日本語文書読解支援系 QJR の検討

亀田 雅之

(株) リコー・研究開発本部 情報通信研究所

kameda@ic.rdc.rioh.co.jp

日本語文書の読解支援機能として、1) 文書のふり分け選択を支援する Screening 支援、2) 走り読み/速読のポイントをリードする Skimming 支援、3) 長い文や複雑な文の分析的な読みを支援する Analytic Reading 支援を検討した。

Screening 支援では、文書の特徴付ける擬似キーワード群を簡易に抽出し表示する。Skimming 支援では、文中の主要文節や主題句などを速読ポイントとして提示し、Analytic Reading 支援では、骨格文節等への縮退木構造表示で複雑な文の構造の把握を助ける。これらの支援機能は、既に開発した簡易日本語解析系 QJP の解析結果を利用した。

また、上記の支援について、評価実験を行い、支援の有効性と課題を確認した。

Support functions for Reading Japanese text

Masayuki KAMEDA

Information and Communication R & D Center, RICOH COMPANY, LTD.

2-3, Shin-Yokohama, 3-chome, Kohoku-ku, Yokohama, 222

This paper proposes a set of functions for Japanese text reading support.

Keyword candidates help in screening texts. Highlighting of key bunsetsu-phrases within sentences promotes rapid reading or skimming of a text. Hierarchical tree structures support analytic reading of long and complex sentences such as patent sentences.

These functions were developed using the analysis results from a quick Japanese parser, QJP.

Experiments with these support functions have brought their effects and the problems to light.

1. はじめに

情報化/ネットワーク化の進展に伴い、電子化された文書情報が急激に増大しており、大規模な文書データベースの検索や要約といったニーズが高まっている。これらは、文書情報へのアクセスの支援と考えられるが、この他に、より基本的な文書情報へのアクセスである「文書を読む」ための支援が考えられる。電子化文書の閲覧のためのプリミティブなツールとしては、UNIX 上の 'more' や 'less' コマンドがあり、文字列検索機能も備わっている。また、ハイパーテキストや HTML 文書のブラウザは、指定された関連情報へのアクセス機能をもった高度な読みの支援ツールである。

本稿では、電子化文書へのアクセスとして、上記のいずれとも異なり、我々の日常のさまざま「読み」の場面に即し、かつ自然言語処理を有効に利用した「文書を読む支援」-読解支援-を検討する。

こうした「読み」の支援については、ほとんど取り組まれていないが、ここでは、応用系開発のために開発した軽量・高速な日本語解析系「簡易日本語解析系 QJP」[1] から着想した。QJP の単語分割結果から簡単に抽出した名詞群が文書の特徴をよく表していること、解析結果の確認のための木構造表示で文の骨格を比較的容易に把握できることに気付いたことが本検討と提案の動機である。

我々の日常的な「読み」の場面はさまざまあり、「読み」方としては、さがし読み（特定項目に目を移し、断片情報をさがす）、眺め読み（新聞紙面を一覧したり、本をめくって大ざっぱに把握し、必要な記事や章をさがす）、飛ばし読み（要点と結論を選んで読み、必要に応じて細部にあたる）、走り読み/斜め読み（細部には注意せず、重要部分を押さえ、あらましをつかむ）、一般的な読み（物語や記事を順を追って読む）、じっくり読む（内容を正確に理解・記憶するための読みで、細部にも注意を払う）等に分類できる[2]。「さがし読み」～「走り読み」に関連しては、速読テクニック¹がある[2]。また、新聞記事や本の章節内の最初と最後の段落、各段落内の最初と最後の文を飛ばし読む Skip Reading 手法も知られている[3]。

上記分類を踏まえ、日常の「読み」を支援する、以下の4つの読解支援を検討し、提案する。

Screening 支援：文書のふるい分け支援

適切なキーワード群は文書の特徴付け、内容の類推に役立つ。Screening 支援では、文書の特徴付ける用語を高速・簡易に抜き出し、提示することで文書ふるい分けの判断材料を提供する。

Skip Reading 支援：飛ばし読みの支援

上述の Skip Reading 手法を参考に重要位置にある文を抽出・表示し、飛ばし読みをリードする。

Skimming 支援：走り読みの支援

文書を読み進む際に「～が～を～した」という部分を拾っていくことができると、文の読みの流れを中断せずに素早く読み進める。Skimming 支援では、構文解析結果を利用し、文の骨格をなす主要文節等を抽出し、速読ポイントとして示し、走り読みを支援する。

Analytic Reading 支援：分析的な読みの支援

特許文のように、1文が長く、複雑で、じっくりと読むような場面で、しばしば長い修飾部分を括弧でくりながら読むことがある。Analytic Reading 支援では、解析した構文構造を適切にまとめ上げて表示することで、複雑な文の構造の把握を支援する。

本稿では、モチベーションとなった QJP について概略を示した後に、各支援²について説明し、最後にこれらの支援機能を対象にして行なった感覚的な評価実験について報告する。

¹ 1字1字を追って読む「逐次読み」に対し、速読テクニックは、1回の停留で瞬間的に多数の文字を視野に入れ、さらに1行の停留回数を減らし、1停留で1行から数行を把握する等の眼球運動の機械的な動作訓練によって読みの速度向上を図るものである

² 一連の支援を総称して「日本語文書読解支援系 QJR(Quick Japanese text Reading support functions)」と呼ぶ

2. 簡易日本語解析系 QJP

QJP は、漢字かな混じり日本語文を形態素列、単語列に分割し、さらに文節間の係り受け構造を解析する系で、軽量[辞書類 100KB 弱(約 5000 語)、処理系 180KB、必要メモリ 260KB]、高速[100~500 単語/秒(80486/25MHz~Pentium/60MHz)](以上 DOS 版データ)でかつ未知語や長文に対する頑強性をもつ。

QJP により、小規模辞書だけで高速に単語分割が得られ(図1)、例文1からは[日本語、単語、切れ目、膠着言語、処理、形態素解析、関門]等のキーワード候補となる名詞群を容易に抽出することができる。

また、QJP ワークベンチでは、文節の係り受け構造の単純な木構造だけでなく、ノード縮退して表示する機能(図2)があり、文の修飾構造の把握を助ける。

前者の高速な名詞群抽出が Screening 支援、後者の構文構造の縮退表示が Skimming 支援と Analytic Reading 支援の着想のきっかけとなった。

例文1：日本語のように単語間に切れ目を置かない膠着言語の文の処理において、形態素解析は第一の関門である。

【単語リスト】			
* [1] (0)	日本語	(日本語)	(41) 名詞
. [2] (6)	の	(の)	(51) ノ=格助
. [3] (8)	ように	<ようだ>	(60. 22) ヨウダ=ダ用b
* [4] (14)	単語	(単語)	(41) 名詞
. [5] (18)	間	(間)	(43) 形式名詞
. [6] (20)	に	(に)	(51) ニ=格助
* [7] (22)	切れ+目	(切れ目)	(41) 名詞
. [8] (28)	を	(を)	(51) ワ=格助
* [9] (30)	置-か	<置く>	(15. 11) 動:五カ未a
. [10] (34)	ない	<ない>	(60. 4) ナイ=ク体
* [11] (38)	膠着言語	(膠着言語)	(41) 名詞
. [12] (46)	の	(の)	(51) ノ=格助
* [13] (48)	文	(文)	(41) 名詞
. [14] (50)	の	(の)	(51) ノ=格助
* [15] (52)	処理	(処理)	(41) 名詞
. [16] (56)	に	(に)	(51) ニ=格助
. [17] (58)	お-い	<おく>	(75. 22) オク=五カ用b
. [18] (62)	て	(て)	(55) テ=接助
. [19] (64)	.	(.)	(92) 読点
* [20] (66)	形態素解析	(形態素解析)	(41) 名詞
. [21] (76)	は	(は)	(52) ハ=係助
* [22] (78)	第一	(第一)	(46) 数名詞
. [23] (82)	の	(の)	(51) ノ=格助
* [24] (84)	関門	(関門)	(41) 名詞
. [25] (88)	で	<だ>	(60. 23) ダ=用c
. [26] (90)	あ-る	<ある>	(75. 3) アル=五カ終
. [27] (94)	.	(.)	(91) 句点

図1. QJPによる単語分割結果

例文2：指定された相手に指定された時刻にメッセージを届けること

【係り受け木構造】		【縮退木構造】	
1 [1] (0)	指定された相手に	1 [2] (1)	指定された相手に
2 [2] (1)	相手に	2 [4] (1)	指定された時刻に
3 [3] (0)	指定された時刻に	3 [5] (1)	メッセージを
4 [4] (1)	時刻に	4 [6] (2)	届けること
5 [5] (1)	メッセージを		
6 [6] (2)	届けること		

図2. QJPワークベンチによる係り受け木構造表示

3. 日本語文書読解支援機能の検討

本節では、1に示した4支援のうちSkip Reading支援を除いた¹3支援について述べる²。

(1) Screening 支援

日常、我々は、手元に入ってくる文書や日々の新聞等を適当にふるい分けして選択している。これは、主として文書のタイトルや記事の見出しを見て判断しているが、さらに、漢字やカタカナの部分だけを「見て」判断することがある。日本語文では字種の使い分けがあり、ひらがな部分は助詞や助動詞が多く、漢字やカタカナ部分は内容語が多い[5]ので、漢字やカタカナ部分を見ることで、文書の種別や分野等を判断できる。これは、文書を特徴付けるキーワードを簡易的、代替的にさがしていることに相当する。しかし、字種だけに頼ると、副詞や動詞、重要性の低い1漢字の和語名詞なども含まれ、ノイズが多い。

一方、通常、検索のために付与するキーワードを抽出するには、精度を高めるために頻度情報とともに単語の構成やキーワード性の分析等を行う[6][7]。しかし、本検討で考える支援レベルでは、本格的なキーワード辞書等を使って高いコストをかけるのはなじまない。頻度情報は、容易に得られる情報であるが、文書全体を処理することが必要である。一方、単語長がキーワード性の簡易的な手掛かりになる[8]。即ち、複数の単語が複合した語長の長い用語は、それだけ、特殊性、専門性が高まっていると考えられ、文書の特徴付けるといふ観点で重要である。単語長は、文書全体の処理なしで、直ちに得ることができる。

こうした検討から、Screening 支援の一つの方法として、QJPの単語分割結果から一般の名詞を抜き出し、さらにキーワード候補(「擬似キーワード」と呼ぶ)を単語長の閾値に基づき簡易に抽出し、擬似キーワードだけを原文と同じ配置で表示する(図3, 図4)方法を提案する。この表示には擬似キーワードしかないので、文として読み進むことができず、逆に表意性の漢字を主体とした擬似キーワード群が散りばめられた「絵」として視覚に訴え、分野の適否や興味の対象か否かの判断を直感的に導く。必要に応じて原文の表示に切り替えることで、選択した文書の読みや着目したキーワードの部分の原文に移ることができる。

¹Skip Reading 支援は、評価実験に含めているが、今回は段落と文の位置情報に基づき、段落と文を抽出・表示した簡便な方法なので特に述べない。ただし、Skip Reading 支援は、抄録文の抽出[4]とも関連が深く、今後言語処理を含めた高精度化が必要である。
²技術的な観点から、Analytic Reading 支援を Skimming 支援の前に述べる

翻訳すべき言語からなる文を単語に分割し、各単語に対応する概念記号と、各概念記号間の関係を示す意味関係記号から概念構造を定め、その概念構造に基づいて翻訳すべき言語を目標言語に翻訳する翻訳システムにおいて、利用者が選択する概念記号及び意味関係記号を受け入れる手段を設け、既に選択された概念記号と意味関係記号に基づき、翻訳すべき言語からなる文の概念構造を定め、その概念構造から他言語の文を生成することを特徴とする機械翻訳方式各単語.....概念記号.....各概念記号.....意味関係記号.....概念構造.....目標言語.....ユーザー.....概念記号.....意味関係記号.....概念記号.....意味関係記号.....概念構造.....概念構造.....他言語.....機械翻訳方式
---	--

図3. 例文3と擬似キーワード抽出表示

通常兵器関連の工業製品 輸出規制が始動 4ヶ国対象	通常兵器の部品や加工機械に転用できる工業製品の輸出規制が二十日、日本でも始まった。英米などの主要先進七カ国(G7)の合意に基づいた規制であり、イラクなど四カ国を対象にして、対共産圏輸出統制委員会(ココム)のリストを準用する形。G7は既に対象となる品目、国を広げるための話し合いを始めており、冷戦終結で変わる新たな輸出規制に発展しそうだ。 規制対象となる国は、イラン、イラク、リビア、朝鮮民主主義共和国(北朝鮮)の四カ国である。北朝鮮は、既に共産圏として特定地域に指定されているため、新たに追加されるのは三カ国である。また、イラクは経済制裁で禁輸措置が取られている。今回の措置で輸出に大きな変化が出るのはイランとリビアの二国になりそうだ。 輸出貿易管理令などに基づいて定められたコンピュータや工作機械などの規制品目を輸出する時には、通産省に許可申請を行う。その際、特定地域に指定されている国に対しては、明らかに民生用とわかる場合でなければ許可が下りず、事実上、規制品は輸出できない。
通常兵器関連...工業製品 輸出規制..... 4ヶ国.....	通常兵器.....加工機械.....工業製品.....輸出規制.....主要先進七カ国.....G7.....イラク.....対共産圏輸出 統制委員会.....ココム.....リスト.....G7.....話し合い.....冷戦終結.....輸出規制..... 規制対象.....イラン.....イラク.....リビア.....朝鮮民主主義共和 国.....北朝鮮.....北朝鮮.....共産圏.....特定地域 イラク.....経済制裁.....禁輸措置.....イラン.....リビア..... 輸出貿易管理令.....コンピュータ.....工作機械規制品目.....通産省.....許可申請.....特定地域.....民生用.....規制品.....

図4. 例文4と擬似キーワード抽出表示

多くの文書があり、バッチ的処理まで可能な場合には、頻度まで計数し、頻度情報と単語長による評価に基づき順序付けした文書ごとの擬似キーワードの一覧表示(図5)のリストも文書のふるい分けを助ける。

3:輸出規制(12) イラク(9) 輸出(6)
2:工業製品(8) 特定地域(8) イラン(6) リビア(6) 北朝鮮(6) G7(4)
1:対共産圏輸出統制委員会(11) 朝鮮民主主義共和国(9)
輸出貿易管理令(7) 通常兵器関連(6) コンピュータ(6) 通常兵器(4)
加工機械(4) 主要先進(4) 冷戦終結(4) 規制対象(4) 経済制裁(4)
禁輸措置(4) 工作機械(4) 規制品目(4) 許可申請(4) ココム(3)
リスト(3) 共産圏(3) 通産省(3) 民生用(3) 規制品(3)

図5. 擬似キーワード一覧表示(例文4)

(2) Analytic Reading 支援

2に示したように、QJP ワークベンチは、係り受け構造を文節単位に表示するだけでなく、一定レベルの文節ノードに下位の文節ノードを縮退して表示できる(図2)。また、縮退文節群のうちでそのヘッドノードから一定深さ以下の文節の表示を抑止する機能をもつ(図6下)。こうした縮退表示は、例文3(図3)のような長文(50文節)の文構造を捉えるのに有効である。また、図6のように段階的に構造の細部を見ていくこともできる。これらの縮退表示は、そのまま Analytic Reading 支援機能ということができる。

しかし、この縮退木構造表示は、あくまで構造的に上位のノードを重要視し、そのノードに対し縮退や表示抑止を行なっているだけである。

ここでは、骨格文節という概念を導入し、骨格文節ノードに縮退を施す。骨格文節は、基本的には「～が～を～し、～が～を～した」という文の基本要素を木構造から認定したものである。次のようないずれかの条件を満たす文節ノードを骨格文節とした。

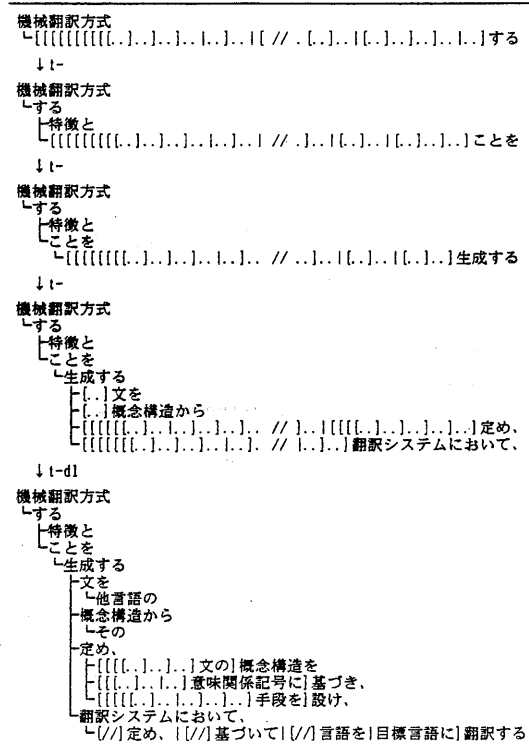


図6. 段階的縮退木構造表示(例文3)

表1. 骨格文節の抽出

- A : ヘッドノード
- A- : Aの直下ノード
- By : 1ノード以上を配下にもつ連用性用言ノード(機能性用言の場合は2文節以上)
- Bt : 2ノード以上を配下にもつ連体性用言ノード
- B- : 用言ノードを直接修飾する直下ノード(並列するノードが全てリーフの場合は除く)
- C1 : BからAのパス上の各ノード
- C2 : Bの各兄弟ノード

Aは文のヘッドであり、Byは文の基本要素のうちの連用性用言であり、Btで複雑な連体性の用言も含める。B-は用言の格要素等の修飾要素である。C1は修飾関係の連続性を満たすための選択であり、C2は木構造表示上の必要な制約である。

図7に例文3の木構造上で骨格文節を示す。

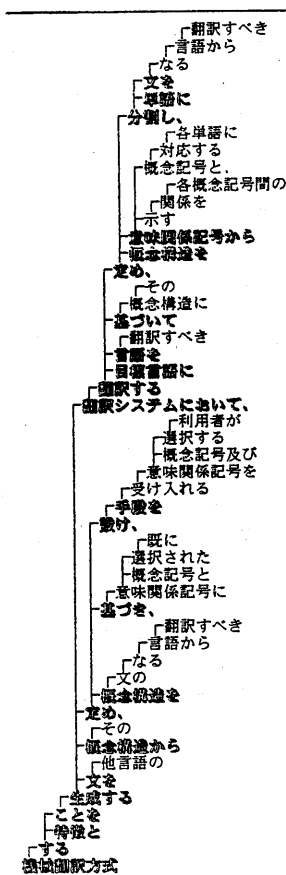


図7. 骨格文節(例文3)

(3) Skimming 支援

主要文節

いわゆる速読テクニックは、機械的な停留により読み進む方法であるが、これに対し Analytic Reading 支援で提案した骨格文節を停留点として識別表示する方法を検討する。骨格文節を停留点（速読ポイント）として示すことにより「～が～を～し、～が～を～した」という文の読みの流れを保証しており、読み手が修飾句も同時に視野に入れて読みを進めていくことで、効果的な走り読みができる。

この支援は、木構造表示でなく原文表示上で識別表示を行うので、骨格文節の C2 のような木構造表示上の制約は満たす必要はない。骨格文節に対し、こうした制約や重要でない要素を取り除き、一方、有効な要素を追加し、表 2 のような主要文節に移行する。

ヘッド (1) と連用性用言文節 (2) 及び修飾の連続性を保証する文節 (X) は、骨格文節の場合と同様である。ヘッドと用言を修飾する要素 (Y,Z) から重要度の低い副詞性の文節等は除く。また、連体性用言句を受けて抽象度を上げると考えられる形式名詞 (3)、主題を示す「は」、文の繋がりを示す接続詞文節 (4) を追加する。

翻訳すべき言語からなる文を単語に分割し、各単語に対応する概念記号と、各概念記号間の関係を示す意味関係記号から概念構造を定め、その概念構造に基づいて翻訳すべき言語を目標言語に翻訳する翻訳システムにおいて、利用者が選択する概念記号及び意味関係記号を受け入れる手段を設け、既に選択された概念記号と意味関係記号に基づき、翻訳すべき言語からなる文の概念構造を定め、その概念構造から他言語の文を生成することを特徴とする機械翻訳方式 文を単語に分割し、.....意味関係記号から概念構造を定め、.....言語を目標言語に翻訳する翻訳システムにおいて、.....手段を設け、.....概念構造を定め、.....概念構造から.....文を生成することを特徴とする機械翻訳方式
---	---

図10. 例文3と主要文節抽出表示

<p>通常兵器関連の工業製品輸出規制が弛緩 4ヶ国対象</p> <p>通常兵器の部品や加工機械に転用できる工業製品の輸出規制が二十日、日本でも弛緩した。英米などの主要先進七カ国 (G7) の合意に基づいた規制であり、イラクなど四カ国を対象にして、対共産圏輸出統制委員会 (ココム) のリストを準用する形。G7 は既に対象となる品目、国を広げるための話し合いを始めており、冷戦終結で変わる新たな輸出規制に発展しそうだ。</p> <p>規制対象となる国は、イラン、イラク、リビア、朝鮮民主主義共和国 (北朝鮮) の四カ国である。北朝鮮は、既に共産圏として特定地域に指定されているため、新たに追加されるのは三カ国である。また、イラクは経済制裁で禁輸措置が取られている。今回の措置で輸出に大きな変化が出るのはイランとリビアの二国になりそうだ。</p> <p>輸出入貿易管理令などに基づいて定められたコンピュータや工作機械などの規制品目を輸出する際には、通産省に許可申請を行う。その際、特定地域に指定されている国に対しては、明らかに民生用とわかる場合でなければ許可が下りず、事実上、規制品は輸出できない。</p>
--

図11. 主要文節抽出表示 (例文4)

表 2. 主要文節の抽出

<ul style="list-style-type: none"> ● 1: 文のヘッドの文節 ● 2: 連用性の用言文節 ● 3: 連体性の用言文節と それに修飾された連用性の形式名詞の文節 ● 4: 「は」、接続詞の文節 ● X: 選択された文節に修飾される文節 ● Y: ヘッドを直接修飾する文節 (除: 副詞性、機能性用言の文節) ● Z: 選択された用言文節を直接修飾する 連用性体言の文節 (除: 副詞性の文節)
--

図 11, 図 10 に主要文節による Skimming 支援表示の例を示す。図 11 では、原文上で強調された主要文節をポイントに走り読みしながら、必要に応じて不足する情報を原文から補うことができる。また、図 10 では、主要文節により長い特許文の大枠を捉えることができ、複雑な文や文書全体を詳細に読む前準備としての支援ともなりうる。

「は」/「が」句

文脈上、係助詞「は」と格助詞「が」が重要な役割を果たしているのが知られている [9]。「が」は主節内で用いられた場合は文脈上の新情報を導入し、「は」は旧情報の中から主題、提題を提示する機能があり、「読む」の上で重要な情報となる。

この観点から、Skimming 支援のもう一つの機能として、「は」と主節の「が」の句をマーキングして、話題の流れや転換を追いながら読むための支援を提案する。この場合、話題の詳細を提示するために、単独の文節でなく、修飾句を含む全体をマーキングする。

当初、「は」と「が」の句のみを識別していたが、評価実験でのコメントから修飾先の用言文節もマーキングするようにした (図 12)。

<p>通常兵器関連の工業製品輸出規制が弛緩 4ヶ国対象</p> <p>通常兵器の部品や加工機械に転用できる工業製品の輸出規制が二十日、日本でも弛緩した。英米などの主要先進七カ国 (G7) の合意に基づいた規制であり、イラクなど四カ国を対象にして、対共産圏輸出統制委員会 (ココム) のリストを準用する形。G7 は既に対象となる品目、国を広げるための話し合いを始めており、冷戦終結で変わる新たな輸出規制に発展しそうだ。</p> <p>規制対象となる国は、イラン、イラク、リビア、朝鮮民主主義共和国 (北朝鮮) の四カ国である。北朝鮮は、既に共産圏として特定地域に指定されているため、新たに追加されるのは三カ国である。また、イラクは経済制裁で禁輸措置が取られている。今回の措置で輸出に大きな変化が出るのはイランとリビアの二国になりそうだ。</p> <p>輸出入貿易管理令などに基づいて定められたコンピュータや工作機械などの規制品目を輸出する際には、通産省に許可申請を行う。その際、特定地域に指定されている国に対しては、明らかに民生用とわかる場合でなければ許可が下りず、事実上、規制品は輸出できない。</p>

図12. 「は」/「が」句抽出表示 (例文4)

4. 評価実験

上記の各支援機能の初期プロトタイプを試作した段階で、これらの支援方式(Skip Reading 支援を含む)の有効性を確認するため、10人を対象に支援表示を印刷した資料により評価実験を行った。

方法

実験では、1組が10記事(約5000字)からなる新聞記事資料を10組ほどを用意し、各組ごとに原文記事群そのもの(以下「原文記事」)、Screening 支援(擬似キーワードリスト[cf. 図5])、Skip Reading 支援(先頭/末尾の文節の先頭/末尾の文の抽出表示)、Skimming 支援(a. 主要文節[cf. 図11], b. 「は」/「が」句[cf. 図12]の各強調表示)の各表示結果を印刷した資料を用意する。Analytic Reading 支援については、a. 段階的縮退表示[cf. 図6], b. 拡張木構造縮退表示[cf. 図8], c. 複数文のF型縮退表示[cf. 図9]の3種類の資料を用意する。

以下、対象者ごとに次の実験を行う。

0. 通常の読み速度の測定: 原文記事を通常の読み方で読んでもらい、所要時間を測定する。

1. Screening: 原文記事とScreening 支援資料をもとに文書ふるい分けの目的で素早く記事を選択してもらい、各所要時間を測定する。後に各々の原文記事を改めて読み、選択すべきだった記事を選択してもらう。

2. Skip Reading: 原文記事とSkip Reading 支援資料を概要把握の目的で飛ばし読みしてもらい、各所要時間を測定する。後に各々の原文記事を改めて読み、概要把握の満足度を振り返り5段階で評価してもらう。

3. Skimming: 原文記事とSkimming 支援資料(2種類)を内容把握の目的で走り読みしてもらい、各所要時間を測定する。後に各々の原文記事を改めて読み、内容把握の満足度を振り返り5段階で評価してもらう。

4. Analytic Reading 支援: 3種類のAnalytic Reading 支援表示資料を提示し、提示した支援表示が参照できる場面で、原文の読解/理解への有用性を5段階で評価してもらう。

この他、最後に全般のコメントをもらった。

0～3までの記事群は同一対象者では重複がない。

表3. 評価実験データの集計結果(XグループとYグループの比較)

実験項目\グループ	Xグループ[7人]		Yグループ[3人]	
0. 通常の読み速度	903字/分 (675-1307)		1440字/分 (3241, 1574, 878)	
	読み時間比	向上点/満点	読み時間比	向上点/満点
1. Screening 支援	10%/ 53%	+2.4, +1.2/100	13%/ 76%	-16.6, -5.5/100
2. Skip Reading 支援	27%/ 81%	+0.10/5	33%/126%	-1.10/5
3. Skimming 支援-a. 主要文節	53%/125%	+0.68/5	94%/106%	-0.47/5
b. は/が句	48%/110%	+0.12/5	82%/ 92%	-1.03/5
	有用度		有用度	
4. Analytic 支援-a. 段階的	2.58/5		+4.34/5	
b. 骨格文節	+3.86/5		+4.17/5	
c. 複数文F	+3.58/5		3.0(*4, *4.1)/5	

※1 時間比: 通常の読み時間/原文の目的読み時間に対する支援の目的読み時間の比

※2 向上点: 1 = 原文と支援での各記事選択の再現率, 適合率[百分率]の差

2, 3=原文と支援での各目的読みの満足度[5段階評価(1-5)]の差

※3 有用度: 目的読みに対する支援の有用(予測)度[5段階評価] *:高評価

結果

10人のデータを整理する中で、1～3の支援が有効に働くグループ(X)と逆に働くグループ(Y)を確認したので、2グループに分けて集計した(表3)。

Xグループは、Screening 支援で、原文でのScreening に対して記事選択の適合率, 再現率を下げずに時間を短縮させ、Skimming 支援-主要文節形式で、原文でのSkimming に対して時間の増加はあるものの、満足度が原文よりかなり高まった。しかし、Skip Reading 支援とSkimming 支援-「は」/「が」形式は、軽微な効果しか見られなかった。

Yグループは一人を除き、もともと速読技術を身に付けているとみられ、1～3の速読系の支援はいずれも逆効果に働いている。Analytic Reading 支援は、全般的に評価が高いが、特に、Yグループで高い。

考察

(対象者数が少ないが)「読み」の速い人を除くと、速読系支援の有効性が確認できる。Analytic Reading 支援は場面は限られるが、全般に支援効果が期待できる。

Screening 支援では、キーワードリストだけでなく、原文と合わせた提示が望ましいという指摘があった。

Skip Reading 支援は、今回の対象記事では、位置情報による文抽出だけでは、十分な効果がなかった。今後、言語処理を施した工夫が必要である。

Skimming 支援では、原文に対する時間の増加があるが、これは、走り読みの場合、原文では読みが流れて読み落としが多いのに対し、支援表示ではリズムカルに適確な読みができるためであることがコメントから明らかになった。満足度の高さ(Xグループの主要文節形式)を裏付け、有効性が確認できる。

ただし、主要文節形式はマーキング箇所が多いという指摘がある。現在の主要文節は、一般に、連体修飾要素の少ない短い文に対しては過剰になる。今後の検討が必要である。

「は」/「が」形式は、満足度では効果が見られなかったが、わかりやすいという好意的なコメントが多かった。特に、修飾先の用言句のマーキングの要望が多く、評価実験の後に用言句のマーキング機能を追加した。この結果、満足度は向上していると考えられる。

5. まとめと今後の展開

従来、余り取り組まれていなかった文書の読解支援機能として4つの支援機能を検討し、本稿では、1) 擬似キーワードの抽出表示により文書のふり分け選択を支援する Screening 支援、2) 文中の主要文節等を速読ポイントとして走り読みをリードする Skimming 支援、3) 縮退木構造表示により長い文や複雑な文の分析的な読みを支援する Analytic Reading 支援の3支援を提案した。本支援は、簡易日本語解析系 QJP をベースとすることで、大規模な辞書や処理系を必要としないのにもかかわらず、比較的高度な読解支援機能を提供する。

提案した支援の評価実験により、各支援の効果の特性と今後改良・検討すべき課題のいくつかが明らかになった。

現在は、擬似キーワード評価の改良と主要文節、重要文の評価方法を検討しており、これらの成果を日本語文書読解支援系 QJR に反映していく予定である。

尚、QJR の主な機能は、QJP ライブラリと併用する QJR ライブラリとして実装し、ワークベンチ形態での機能確認やビューワ等への一部機能の搭載を行っている。

参考文献

- [1] 亀田雅之：軽量・高速な日本語解析ツール『簡易日本語解析系 QJP』、言語処理学会 第1回年次大会, 1995.
- [2] 佐藤泰正：速読の科学, 講談社, BLUE BACKS B732, 1988.
- [3] 田中雄二：田中の現代文速読速解法, 三省堂.
- [4] 渡辺日出雄：新聞記事の要約のための一手法, 言語処理学会 第1回年次大会, 1995.
- [5] 大樫仁司, 小池和弘, 金枝上教史：字種切り法による形態素解析の一改良, 情報処理学会 第43回全国大会, 1992.
- [6] 小川泰嗣, 別所, 岩崎, 西村, 広瀬：短単位キーワードに基づくテキストデータベースシステム, 情報処理学会 データベースシステム研究会 90-6, 1992.
- [7] 小川泰嗣, 望主, 別所：複合語キーワードの自動抽出法, 情報処理学会 自然言語処理研究会報告 97-15, 1993.
- [8] 久保田淳市：キーワード抽出装置, 特開昭 63-244259, 1987.
- [9] 大野晋：日本語の文法を考える, 岩波書店, 岩波新書 黄 53, 1978.

例文出典

- 1：久光徹, 新田：情報処理学会 第42回全国大会, 1991
- 3：公開特許公報(昭60)
- 4：朝日新聞, 1993.1.21