

談話構造と読解過程に注目した英文精読支援システムの開発

井佐原 均

郵政省通信総合研究所 関西先端研究センター

isahara@crl.go.jp

1 はじめに

インターネット等の普及に伴い、情報の発信・受信のそれぞれの場面において、英語に触れる機会が増えてきている。情報受信における言語障壁を克服する手段として、英日翻訳システムや要約システムを応用した英文読解を支援する環境が考えられてきた。しかしながら要約は、きちんと読むべき文献を見つけるためのスクリーニングの役割は果たすが、長い英文から詳細な知識を得るには不十分である。また既存の機械翻訳システムでは、その出力文章を読むだけで十分に知識を得られるだけの精度が得られない。従来の英日機械翻訳システムは、その過去における主たるユーザが翻訳会社等であったこともあり、最終出力として、正しい日本語文章を想定していた。しかしながら、原言語の文章を目標言語の文章に変換するという「翻訳」は、個人の情報受信の支援には必ずしも適切なものではない。他言語で書かれた文章を、一旦母国語の文章にした上で情報を得るのではなく、非母国語文章から直接情報を取得することが必要となる。

本稿では、英語文章の談話構造を把握し、人間の読解過程に沿って談話構造に関する情報を表示することにより、英文の効率良い精読を支援するシステムを提案する。ここでの目標は日本語として正しい文章を生成することではなく、情報を効率よく収集することを可能とすることにある。

情報受信における言語障壁の克服のためには、従来型の英日機械翻訳ではなく、人間の知的活動に密接につながった支援環境の構築を行う必要がある。このような環境の実現に向けて、日本人の英文読解過程を調査し、それを支援する手法を検討することは、今後の自然言語処理研究において必須となると思われる。

2 英文精読支援システム

我々は、人間の知的活動を支援することを目標に討論型のネットワークニュース記事を対象とする知的ニュースリーダを開発している [井佐原, 1997]。これは、インターネットにおける情報利用のヒューマンインタフェースとして、記事の要約ではなく、記事群中の話題の流れや類似記事に関する情報を提示し、ユーザが興味を持った話題についての情報を持つ記事(だけ)をきちんと読むことを支援することを目的としている。ここでは、要約では十分な情報を与えられない場合があるということを前提としている。今回提案する英文精読システムも、要約を読むだけでは、あるいは速読では不十分であり、ある程度の長さを持つ文書をきちんと読む必要に迫られたユーザを支援することを目的としている。例えば、研究者が自分の専門分野についての論文を読む場合を想定している。

これまで英文読解支援としては、要約(した和文)を提示する、単語(群)の訳をふる、といった方法が提案されてきた。英日機械翻訳の弱点を克服し、単語訳よりも高度な支援を行うことを目標に、長い単位(句)での訳を付与する試みもなされてきた [九津見, 1996]。また、読解力をつけるための教育システムとして、速読支援機能を持つ読解支援システムも提案されている [國近, 1997]。

本稿では、辞書引き機能(訳り機能)に加えて、文章中の構成要素の相互関係に基づいて文章中の話題の流れを提示する機能や、人間の読解過程を調査した結果に基づく支援機能(必要な場所へのジャンプ等)を加えた英文精読支援システムを提案する。英語で書かれた文章を、文単位の翻訳が可能な独立した文が集まったもの

とみなすのではなく、流れを持った文章として、それを読み進むのに適切な支援をするものである。

文章中の構成要素の相互関係を記述する研究としては、談話構造を記述し、談話構造中の情報がどのように文の解釈に影響を与えるか、文の解釈が談話文脈の更新にどのような寄与をするか、についての研究がなされている [Kamp, 1993; Groenendijk, 1991]。本システムは、このような談話構造の把握と更新の考え方に基づいた支援を目指している。人間が文章を読み進む場合には、その文章の既に読んだ部分で得られた知識を活かして読み進むのであり、文章の各部分に関して提示する情報も、この知識を前提とするかどうかによって異なってくる。

文章構造の理解を支援する機能としては、文構造表示、指示表現の指示対象表示、修飾関係表示、談話の流れの表示等が考えられる。本システムでは、知的ニュースリーダのために開発したキーワードの流れによる話題の転換・分岐の判断とその表示法 [内元, 1998] を応用することにより、文章の各部分の意味的あるいは文脈の特徴を前後の部分との変化に注目して表示する。

3 心理実験による読解過程の調査

このような、要約ではない、速読支援でもない英文精読支援環境の開発に当たっては、まず、日本人の英文読解過程を調査し、そこで必要とされる支援機能を明確化することが必要である。また、精読支援の実現のためには、英文読解支援の簡易モデルである「英単語への訳振り」システムに対し、どのような機能をつけ加えなくてはならないかを検討し、必要とされる個々の機能についての検討を行うことが必要となる。

読解過程を調査する場合、要約作成等のタスクを与えて、作業中の発声を記録する方法 (think aloud) がある (たとえば [Seguinot, 1996]) が、英文を理解している過程と、要約を作成している過程とが混在すること、発声することによる影響が考えられることから、本研究では、以下に述べる手法を採用した。

英文精読支援システムに必要な機能を調査するため、人間の読解過程を記録するシステムを作成し、心理実験を行う。実験システムの画面を図1に示す。ここでは、ページをめくり、辞書を引き、下線を引きながら、読み

進んでいくといった、人間が通常、印刷された文書を対象に行う読解過程と出来るだけ近い状況を実現することを目指している。ただし、本システムではレイアウト情報や図表の表示はできない。

ウィンドウは、全体に濃い青で着色され、読みづらくなっている。マウスで指定した行の前後数行ずつが読みやすいように白地の画面となる。また、任意の行にマークをつけることが可能であり、その行は水色になる。これは通常の英文読解において、文章上にアンダーラインを引くことに相当する。被験者はこのシステムを使って、画面上で与えられた英文を読み、重要だと思った行に、あらかじめ定められた数のマークを付けることが要求される。システムは、読解開始からの時間と現在読んでいる場所 (マウス操作で白地にされた部分) とを記録する。読解中に辞書引きしたい単語 (群) が現れた場合には、それをクリップボードに入れ、辞書引きすることができる。このようにして検索した語も記録しておく。

このように画面の一部のみを読みやすい表示にし、その部分の動きを記録するツールはこれまでに提案されてきた [関, 1995] が、本ツールでは読解の「真剣さ」を保証する手段としての「マーク付けタスク」を行う機能もツール内に組み込み、さらには、辞書引きの履歴も記録している。

この実験では、重要な文に定められた数のマークを付けるというタスクで、定められた行数の要約を作成することを模倣している。このタスクを与えることにより、読解は文書の最初から終わりに向けて単純に読み進むのではなく、数行前、あるいはマークを付けたところに遡ったりしながら行われる可能性がある。

読解過程の記録において、文脈確認のため後戻りして読んだ部分、単語を辞書引きした部分、構文の理解のために読解に時間がかかった部分等が、難解な部分であると考えられ、「精読支援」をすべき箇所となる。

4 おわりに

談話構造に注目して、英文の精読を支援する環境を提案した。本稿で述べた考え方と、心理実験の結果に基づいて、実用レベルの英文精読支援環境の開発を引き続き行なっていく予定である。

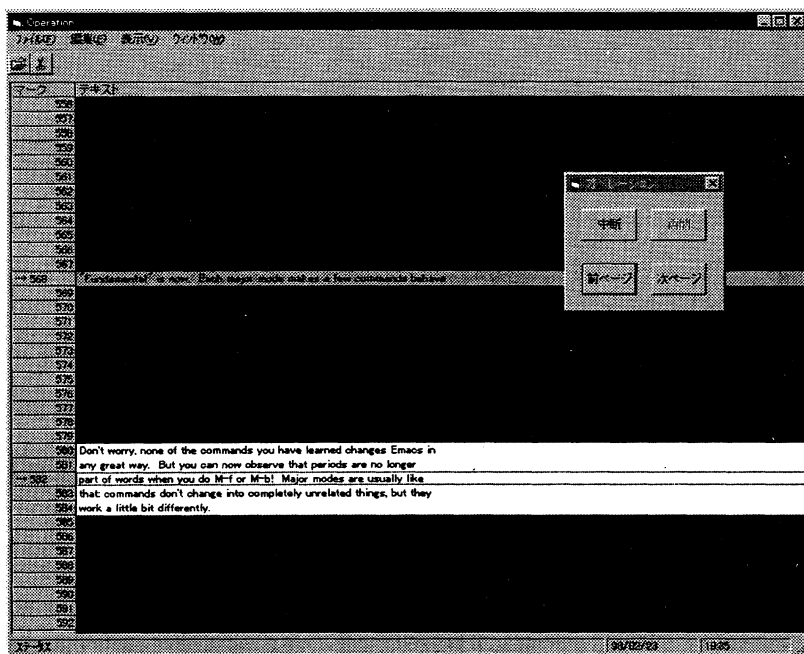


図1 読解過程モニタリングツール

謝辞

本研究の実施に際しては、(株)東洋情報システムの高橋誠氏、渡辺浩明氏、シャープ(株)の福持陽士氏、佐田いち子氏のご協力を得た。記して、ここに感謝する。

参考文献

- [井佐原 1997] 井佐原均, 小作浩美, 内元清貴. 討論型ニュースグループを対象とする知的ニュースリーダーの開発, 情報処理学会自然言語処理研究会, NL119-3, 1997.
- [九津見 1996] 九津見毅, 奥西稔幸, 佐田いち子. 英日機械翻訳における速読支援のための日本語生成, 言語処理学会第2回年次大会, 1996.
- [國近 1997] 國近秀信ほか. 英文読解支援システムと速読支援機能の教育的効果, 情報処理学会人文科学とコンピュータ研究会資料 33-2, 1997.
- [Kamp1993] Hans Kamp and Uwe Reyle. *From Discourse to Logic: Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*, Kluwer Academic Publishers 1993.
- [Groenendijk1991] Jeroen Groenendijk and Martin Stokhof. *Dynamic predicate logic*, Linguistics and Philosophy, 14, 1991.
- [内元 1998] 内元清隆, 小作浩美, 井佐原均. キーワードによるネットワークニュース群の構造化, 言語処理学会第4回年次大会, 1998.
- [Seguinot1996] Candace Seguinot. *Some Thoughts About Think-Aloud Protocols*, Target 8:1, 1996.
- [関 1995] 関友作, 赤堀侃司. ディスプレイ表示法の考案によるテキスト読解過程の測定, 信学技報 ET95-94, 1995.