

# 東芝ソリューション株式会社

本社住所	〒105-6691 東京都港区芝浦 1-1-1(東芝ビルディング)
URL	<a href="http://www.toshiba-sol.co.jp/">http://www.toshiba-sol.co.jp/</a>
展示名	<b>可読性診断技術</b> —産業日本語ライティング支援への活用—
お問合せ先	担当部署 IT 技術研究所 TEL 042-340-6366 FAX 042-340-6013 E-mail kano.toshiyuki@toshiba-sol.co.jp

## ■アピールポイント

理解しにくい文の特徴を考慮した診断規則に基づいて、文の可読性を診断します。

診断結果に基づいて文を修正することにより、読みやすい文章を作成することができます。また、機械翻訳の前処理に利用すると、誤訳しやすい表現を指摘するので、簡単に修正して翻訳精度を向上することができます。

産業日本語ライティング支援ツールとしての活用を検討しています。

## 【産業日本語との関連】

読み手に誤解されやすい表現のチェックを目的として、可読性(理解しやすさ)を判定する技術を開発している。この可読性の判定基準は、産業日本語のライティングルールと共通する点が多く、ライティング支援ツールとして活用できると考えられる。

## 【詳細】

技術文書は、その内容が正確であるだけでなく、内容が読み手に正しく効率的に伝わるのが望ましい。また、機械翻訳等の機械処理においても、書き手の意図が正しく理解され、正確に処理できるのが望ましい。ここでは、記述された文章の理解しやすさを、その文章の可読性とよぶ。可読性の低い文章は、読み手に誤解される危険性が大きく、内容を理解するのに時間がかかる。また、機械処理においても、書き手の意図とは異なる意味に解釈される可能性が高い。可読性診断技術は、文章の理解しやすさを診断し、可読性が低い箇所を検出する技術である。

専門分野の文章の翻訳を翻訳者に依頼する場合、翻訳者が必ずしもその分野に精通しているとは限らない。そこで、翻訳者が内容を正しく理解できるような文章、すなわち可読性の高い文章が求められる。可読性が低い文章は、誤訳を生じさせ、翻訳された文章の修正や再翻訳などの作業増加を招く。

機械翻訳においても、可読性が低い文章では、良い翻訳結果が期待できない。現在の機械翻訳製品は、簡潔でわかりやすい文章に対しては、実用レベルに達した翻訳結果を出力する。しかし、可読性が低い文章を入力すると、正しい翻訳は難しくなり、翻訳の精度が低下する。機械翻訳には人間並みの言語理解能力を期待できないので、翻訳者に依頼するときより一層、可読性の高い文章が求められる。

機械翻訳により誤訳が発生する文の例を次に示す。

[例文 1] 顧客用文書をカラー、社内用文書を白黒で印刷する。

[日英翻訳結果 1] Color and the document for in-house use are printed for the document for customers in black and white.

[日中翻訳結果 1] 在黑白印刷顾客用文书彩色、公司内部用文书。

この例文 1 の場合、「カラー」の直後に述語がないため、機械翻訳は「カラー」の係り先を正しく認識できず、原文の意味とは異なる意味に翻訳してしまう。印刷機能として「カラー」と「白黒」の選択が有ることを知っている人にとっては、その知識で補完することにより、この文を正しく解釈できるが、機械翻訳では「カラー」と「白黒」の関係を捉える前に、「カラー」とその直後の「社内用文書」との並列関係を機械的に捉えてしまい、誤訳を導いている。

このような可読性が低い文を、事前診断で発見できれば、文章作成者に対して修正を促すことにより、人間翻訳や機械翻訳の精度と効率の向上が期待できる。

可読性診断システムでは、構文解析結果を用いて、可読性診断機能群に属する個別の診断機能により、文の可読性を順に診断する。現在取り組んでいる可読性診断機能を次の表に示す。

診断機能名称	説明	※
曖昧な係り受け	係り受け関係の解釈が複数ある箇所を指摘する。	2
複合語	辞書に登録されていない複合語で、直訳では訳しにくい接辞(「可」、「未」、「無」など)を含むものを指摘する。 例: ペット可賃貸マンション、文書管理システム未導入部門	3
述語の省略	述語動詞が省略されている箇所を指摘する。	3
主語の省略	主語が省略されている述語動詞を指摘する。	3
目的語の省略	目的語が省略されている述語動詞を指摘する。	3
主語と述語が離れている	述語動詞から離れている主語を指摘する。	2
目的語と述語が離れている	述語動詞から離れている目的語を指摘する。	2
長い修飾部	修飾部が長い箇所を指摘する。	2
述語の数	述語動詞が多く含まれている文を指摘する。	1
助詞「は」	格の曖昧性がある助詞「は」を指摘する。	2

※産業日本語ライティングルールとの対応 1=短文化、2=構文明確化、3=語彙明確化  
可読性診断システムにより、例文 1 に対して次のような診断結果が得られる。

「カラー」の直後の「で印刷し」が省略された可能性がある。

利用者は、診断結果を参考にして「で印刷し」を補足し、例文 2 のように修正できる。修正結果を機械翻訳すると、正しい意味の訳文が得られる。

[例文 2] 顧客用文書をカラーで印刷し、社内用文書を白黒で印刷する。

[日英翻訳結果 2] The document for customers is printed in color and the document for in-house use is printed in black and white.

[日中翻訳結果 2] 以彩色印刷顾客用文书，以黑白印刷公司内部用文书。

この可読性診断技術を、産業日本語ライティング支援ツールとして活用できないか、評価を行った。適用可能性を検証するために、特許文書の翻訳を想定した評価実験を実施した。その結果、誤訳しやすい文の 7 割を検出することができた。また、「主語の省略」診断を除いた評価では、指摘箇所の 85%で誤訳が発生しており、誤訳の危険性を診断できる見通しが得られた。

展示会場では、可読性診断技術により文章中の可読性が低い箇所を検出するデモを展示する。