

可読性診断技術

東芝ソリューション株式会社 IT技術研究所

2012年2月29日



東芝グループは、持続可能な 地球の未来に貢献します。

東芝ソリューション株式会社 TOSHIBA SOLUTIONS CORPORATION



業務文書の品質

内容の品質

※文書を設計するとき決まる。例えば、記載の中身、詳細度など。

業務文書 の品質 表現の正確性

※記載した表現と記載しようとした表現が異なる。 例えば、誤字脱字、かな漢字変換の誤りなど。

表現の品質 ※文書を執筆する とき決まる。 表現の可読性

※記載した表現が人に理解しやすいかどうか、 コンピュータで処理しやすいかどうか。

表現の整合性

→ ※記載した情報に整合性があるかどうか。 例えば、引用先と引用元の不一致、数値の不整合、 表記揺れなど。



機械翻訳への活用例

原文

「顧客用文書をカラー、社内用文書を白黒で印刷する」

診断なし

診断あり

診断結果:

「カラー」の直後に「で印刷し」が省略されているようです。



可読性診断システム

修正後:

「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」



診断結果に基づく修正



機械翻訳システム

英訳結果

Color and the document for in-house use are printed for the document for customers in black and white.



在黑白印刷顾客用文书彩色、公司内部用文书。



英訳結果

The document for customers is printed in color and the document for in-house use is printed in black and white.

中国語訳結果

以彩色印刷顾客用文书,以黑白印刷 公司内部用文书。



可読性診断技術の活用シーン

- 機械翻訳の前処理
 - 曖昧性のある表現が指摘される
 - 原文を修正し、翻訳の精度を向上
- 翻訳コストの低減
 - 分かりやすい翻訳原文が作成できる
 - 翻訳者からの問合せ、翻訳者の誤解による誤訳が減少
- 人が理解しやすい文章の作成
 - 仕様書などの技術文書で、正しい情報伝達を支援
 - 理解度が増し、業務効率が向上



可読性診断技術の構成

翻訳対象文書

日本語構文解析エンジン

可読性診断 システム



構文解析結果

規則	内容
曖昧な	係り受け先の候補が複数あり、どれ
係り受け	が正しいのか判断しにくい
複合語	辞書に登録されていない複合語に 訳しにくい接辞が含まれている
述語の省略	連用中止となる述語が省略された 並列構造

. . .

診断規則



診断結果に対する 確認と修正

ユーザ



診断モジュール

- ・曖昧な係り受けの診断
- ・複合語の診断
- ・述語の省略の診断
- ・助詞「は」の診断

. . .

コメント[2]: <述語省略>

「カラー」の後に「で印刷し」が省略されているようです。

[述語] で印刷し

付加情報





技術的特徴

- 翻訳しにくい文の特徴を考慮した診断規則に基づいて、 文章の可読性を診断
- 日本語構文解析の結果を使用し、 文の構造から誤訳しやすい箇所や曖昧な箇所を検出
- 翻訳の精度を向上するための付加情報や選択肢を生成

コメント [2]: <述語省略> 「カラー」の後に「で印刷し」が省略されているようです。 「述語] で印刷し

付加情報

昨日保存されたデータが削除された

コメント[3]: <曖昧な係り受け>

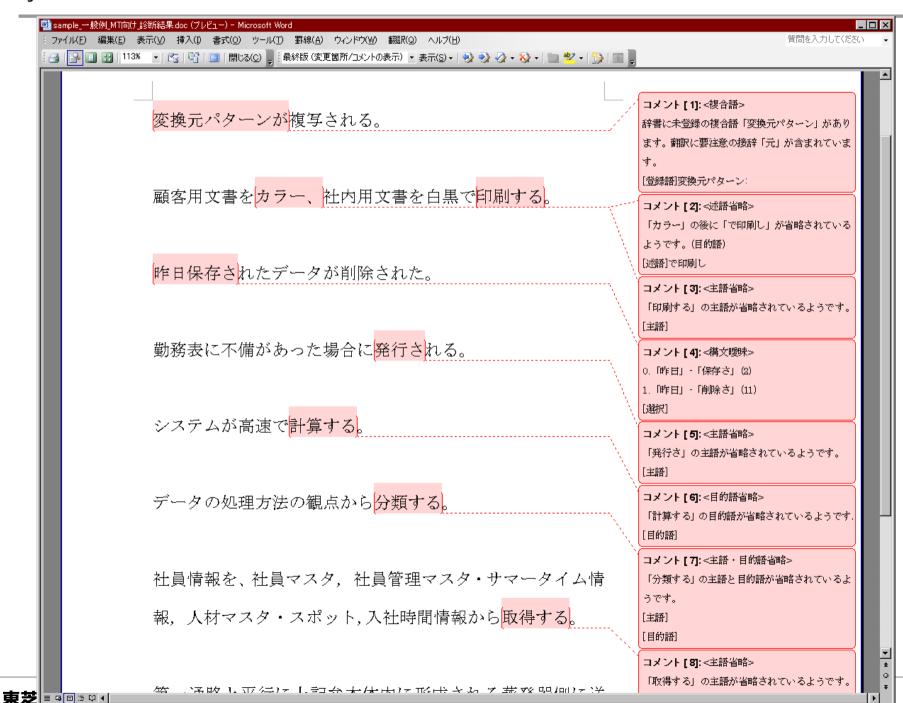
「昨日」-「保存」 「昨日」-「削除」

- 1 ページ

1 セクション 1/1

位置 35mm

可読性診断結果の出力例



1 行 15 桁 記録 変更 拡張 上書 日本語 🛂

Corporation



産業日本語との関連

- 可読性診断の判定基準と産業日本語ライティングルールとの共通性
- ライティングの支援システムや支援環境としての活用



ご清聴ありがとうございました