

ホンヤク出版社のソリューション

翻訳前に「改行」や「空白」などの 制御文字を事前に取り除く方法とは？



「伝える」をデザインする

 株式会社 **ホンヤク出版社**

翻訳前に「改行」や「空白」などの 制御文字を事前に取り除く方法とは？

現在、産業翻訳の分野では CAT (Computer Assisted Translation : コンピューター支援翻訳) ツールは欠かせない存在になっています。

CAT ツールは過去に翻訳した同じフレーズを再利用したり、訳抜けをチェックしたり、用語集を即座に検索したりといった翻訳に欠かせない作業がスムーズに行えます。しかし、それができるのも入力となる原文が整っていないければなりません。例えば、もし、原文が途中で改行して 2 分割されていたら、過去に翻訳した同一文とのマッチングに失敗して、折角の再利用のチャンスを見逃してしまいます。そこで、前処理として、「改行」と「空白」といった制御文字を翻訳前に事前に取り除く必要があります。しかし、これが意外と面倒でかなり時間を取られてしまいます。

弊社には、ボタン一つで、この問題を解決する技術があります。

課題

あるお客様からマニュアルの翻訳を依頼されました。

全 9,000 行のマニュアルでしたが、原稿としてテキストファイルが支給されるとあって、CAT ツールを使えば翻訳は順調に終わると思われました。

ところが、実際にテキストファイルを見ると、あちこちに「改行」や「空白」、「タブ」といった制御文字が使われていて文章が分断されていました。この状態では CAT ツールを使用しても文が途切れているので、過去の類似翻訳文の呼出しなど CAT ツールの機能を生かした翻訳ができません。

なぜ、このようなファイルになってしまったのか。お客様に事情を聞いてみると、私たちに支給されたテキストファイルは Word ファイルから作られたもので、その Word ファイルはイラストが多用されていました。そのイラストの位置を調整するために、文章内に「改行」や「タブ」を挿入したとのことでした。

解決策

通常であれば手作業で不要な「改行」や「空白」を取り除かなければなりませんが、ここで弊社が開発した「文章整形ツール」を使用すれば、余分な制御文字を削除する一方で、必要な改行は残し、本来あるべき日本語文章の姿に変換することが出来ます。

このツールは弊社が持つ AI 技術（自然言語処理技術）の語彙の接続係数を算出する技術を使用しています。この技術は、適用できる文章のカテゴリーに制限は無く、動作も軽いという特徴があります。ユーザーは処理したい文を入力したら、スタートボタンを押すだけです。

まとめ

「文章整形ツール」はボタン一つ押すだけで不要な「改行」、「空白」などを削除して、翻訳しやすい文に整形します。地味なツールですが、その効果は絶大です。弊社ではこのような人的負荷を軽減しつつ、CAT ツールの性能を最大限引き出す「ちょっとしたツール」を開発し、制作現場に貢献しています。ご興味がある方は、ぜひ一度ご相談ください。

(2019年7月1日)

ホニャク出版社のソリューション一覧

- Web を簡単に多言語化する方法
- Word だって自動組版できるんです
- 改訂履歴の見られる電子マニュアルはいかがですか？
- Illustrator 多言語版の時短テクニック
- 動かない AI を生み出さないため、不揃いなデータをデータ整形（データクレンジング）で統一しませんか？
- 翻訳前に「改行」や「空白」などの制御文字を事前に取り除く方法とは？
- 用語集を自動作成！同時に表記揺れも検出！
- 翻訳資産管理
- マニュアル診断から、その先へ
- 英訳する日本語原稿は何に気をつけて書く？
- 翻訳費用削減事例
- Microsoft Access を使ったデータ管理

上記に類似したお悩みごとはございませんか？

弊社がお力になれるかもしれません。

各種ソリューションは弊社の Web サイトからご覧いただけます。

是非ご確認くださいませ！

QR コード読み取りまたは
「ホニャク出版社」で検索



ページ上方のメニューから
「ソリューション」にアクセス

「ソリューション」ページ URL : <https://www.honyaku-shuppan-sha.co.jp/solution.html>

お問い合わせ

03-3355-4411 (9:00~17:30)

Web サイトからもお問い合わせいただけます。

トップページの「お問い合わせ」ボタンからフォームに入力してご送信ください。

商標/コピーライト

Microsoft Word、Excel、PowerPoint は米国およびその他の国における米国 Microsoft Corp. の登録商標です。

Adobe、Adobe FrameMaker、Adobe InDesign、Adobe Acrobat、Adobe Photoshop、Adobe Illustrator は Adobe Systems Incorporated (アドビシステムズ社) の商標です。

Trados は SDL TRADOS 社の登録商標です。

XDocForm は株式会社ホニャク出版社の登録商標です。

その他、本書に表示・記載されている各社の会社名・サービス名・製品名等の商標は、それぞれの会社の商標もしくは登録商標です。

表紙写真提供：Marc Babin(<https://unsplash.com/@marcbabin>)