

複合機における多言語メッセージの分析と再利用手法

坂本佳史, 合田麻衣, 福岡直明, 竹内広宜, 原口智史

デジタル複合機の操作パネルに表示するメッセージは多言語化が推進され 25ヶ国語に及ぶ言語をサポートしている。近年、デジタル複合機に搭載される機能が以前に比べて格段に増加したこと、ヘルプ機能や状態表示などユーザーフレンドリーな情報の提供が求められることで多種、多用なメッセージにより情報をユーザーに提供している。これによりデジタル複合機に用いられるメッセージ数は年々増加、2005年の製品との比較では約2.3倍にも膨れあがっている。

本論文では、デジタル複合機の実際の開発において多言語の翻訳メッセージ・データを解析、再利用を適用する事で翻訳データの再利用率の知見を得た。更に再利用率の向上を目的に調査・解析を行うことで高い再利用率を実現する為の日本語メッセージの形態を特定した。

Analysis and the reuse techniques of the multilingual message in the MFP development.

Yoshifumi Sakamoto Mai Gohda Naoaki Fukuoka Hironobu Takeuchi and Satoshi Haraguchi

In late years, the digital Multi function peripheral/printer(MFP) supports 25 languages. The messages which MFP provides to the user increase year by year. The reason of the increase of the message that is the increase of the function of MFP, a help function and an offer of user-friendly information in the indication which is in a state are required. The number of messages which we put on MFP increased to about 2.3 times in comparison with 2005. In this paper, We analyzed multilingual translation message data in real development of the MFP and got knowledge of the reuse rate of translation data by applying reuse. We specified the form of the Japanese message to realize a higher reuse percentage.

1. はじめに

近年、デジタル複合機(MFP: Multi function peripheral/Printer)は多機能化が進み、それに伴い操作パネルに表示するメッセージ数は増大している。更にユーザーの利便性を向上するためにヘルプ機能の拡張、エラーやメンテナンスのメッセージに代表されるMFPの状態を提示するメッセージの充足に伴ってメッセージ数の増加に拍車をかけている。更にMFPはその市場の大半が海外であることから、多くの言語に対応したメッセージを表示することが必須の機能であり現在、対応する言語の数は25ヶ国語に及ぶ(表1)。

実際に翻訳のブレが発生することである。翻訳のブレとは同一の日本語、もしくは文の断片を翻訳したにも関わらず翻訳後の複数の外国語の文において異なる単語、異なる表現に翻訳されることを意味する。これはすなわち翻訳の品質低下である。もう一つの課題は翻訳対象となるメッセージ数の増大に伴う翻訳作業の急激な増大である。翻訳作業の急激な増大は翻訳費用の増加、翻訳作業期間の増加、翻訳作業に関わる要因の増加を招く。これまで翻訳メッセージの作成は製品開発プロジェクトに大きな影響を及ぼす作業では無かったが翻訳作業の急激な増大により製品開発プロジェクトに対する新たな制約として問題視されるようになってきている。また翻訳作業の増大それ自体が翻訳の品質低下を招いている。

これら課題に対して、現状のMFP操作パネルの多言語メッセージの開発においては、過去に開発した機種種のメッセージ資産である既存メッセージリストを用いて翻訳結果を再利用することで対応している。しかしながら、既存メッセージリストの管理単位は、操作パネルにおいて提供する機能単位、もしくは機器の状態の単位で管理している。これによりメッセージは一連の複数の文から構成される

対応言語一覧

日本語	イタリア語	シンママク語	ポーランド語	アラビア語
英語	オランダ語	フィンランド語	ポルトガル語 (ポルトガル)	ヘブライ語
ドイツ語	ロシア語	ギリシャ語	ルーマニア語	韓国語
フランス語	ポルトガル語 (ブラジル)	ハンガリー語	スウェーデン語	中国語(中国・香港)
スペイン語	チェコ語	ノルウェー語	トルコ語	中国語(台湾・香港)

表1. MFPの対応言語一覧

MFPに搭載するメッセージ数は今後も増大すると考えるがそのメッセージの作成において以下、2つの課題がある。1つは日本語から他言語に翻訳する

場合があり,再利用する際に不要なメッセージを含む場合がある(表2)。

Japanese	English	Arabic
読み込み制限ページ数に達しました。このジョブを中止します。アラビヤ語のメッセージは必要ありません。	Maximum number of scanned pages. The job will be canceled. Are you sure?	تجاوزت الحد الأقصى لعدد الصفحات المسحوبة. سيتم إلغاء العمل. هل أنت متأكد من أنك لا تحتاج إلى الرسائل العربية؟

表2 翻訳メッセージの再利用における問題点

英語の場合,メッセージが不要と判断することは比較的容易であるがそれ以外の言語の場合には,その言語に精通した翻訳担当者の判断が必要となり,再利用における生産性を著しく低下させる要因となっている。

本論では上記問題を解決する為に翻訳データの基本データとなる日本語のメッセージを文単位で構成,管理する事により対訳である外国語の翻訳データの高い再利用率を実現した。更に日本語のメッセージとその英語の対訳を調査する事で更に翻訳データの再利用率を高める為のメッセージデータの変更を提案する。

2. メッセージの分析

MFP に搭載する日本語メッセージと英語対訳を分析した結果,以下の特徴を得た。

- i)ユーザーに対する指示が多く,使用場面が限定されているため,指示を表すメッセージ文の意味が複数あることが少ない
- ii)メッセージ対訳から深い処理を行わずに文対訳を抽出でき,新規メッセージに対して,既存の文対訳データを組み合わせることで翻訳を作成できる
- ii)複数の場面で使用する指示が多いため,指示を表すメッセージの再利用性が高い。メッセージを文に分割して管理することで,再利用性が更に高まる

そこで既存の日本語メッセージにおいて使用頻度が高く,1文によってメッセージを構成し,その英文対訳も1文によって構成するメッセージを110個抽出した。このメッセージはメッセージ総数8,169個の1.4%に相当する。

3. メッセージの作成

この抽出したメッセージを新規のメッセージ作成において文単位での再利用を実施した。その結果,全体のメッセージのうちこれらの文の組み合わせによって作成する事が可能となったメッセージ数は20.7%である。また新規メッセージの部分に対して再利用の適用が可能となったメッセージは29.3%であった。

4. 再利用範囲の拡大

次にメッセージの再利用の範囲を拡大するために全てのメッセージに対する調査を実施した。調査の観点は再利用を阻害する以下の2つの要因である。

- i) 日英の対訳文の数が不一致-対訳間で文の数が異なると翻訳文を再利用できない。また句点によって連結された文章は再利用性が低下する
- ii) 翻訳のブレ -1つの日本語文に対して複数の英訳が存在する

調査には自然言語処理ソフトウェアを用いた。日英の対訳文の数が不一致である翻訳データは全体の4.0%,翻訳のブレが存在する翻訳データは6.9%であった。これら対訳を修正する事により更に翻訳データの再利用率を向上する事が可能と考えられる。

5. まとめ

ポスターセッションにおいてこの事例の紹介と具体事例,更に翻訳データの再利用を促進するために本論の手法と結果を反映した入力支援ツールの発表を行う予定である。

参考文献

- [1] 伊川洋平, 金山博, 竹内広宜, 渡辺日出雄, 三品拓也, 秋本仁志, 清水淳也: オフィス文書のための翻訳支援ツールの実装と評価, 人工知能学会第23回全国大会発表論文集, 3I4-3 (2009).
- [2] 竹内広宜, 荻野紫穂, 中田武男, 坂本佳史: テキスト分析技術を用いた開発関連文書の品質分析, 2009年後期プロフェッショナル論文(2009).
- [3] テクニカルコミュニケーション研究会: わかりやすいマニュアルを作る 文章・用字用語ハンドブック, 日経BP出版センター (1995).