

大学向OSSの翻訳に適用する共通翻訳メモリの開発と国際コミュニティへの展開

常盤, 祐司 / TOKIWA, Yuji

(雑誌名 / Journal or Publication Title)

科学研究費助成事業 研究成果報告書

(開始ページ / Start Page)

1

(終了ページ / End Page)

5

(発行年 / Year)

2016-06

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 1 日現在

機関番号：32675

研究種目：基盤研究(B) (一般)

研究期間：2013～2015

課題番号：25280127

研究課題名(和文) 大学向OSSの翻訳に適用する共通翻訳メモリの開発と国際コミュニティへの展開

研究課題名(英文) Development of a common translation memory to translate open source softwares for higher education and dissemination to global community

研究代表者

常盤 祐司 (TOKIWA, Yuji)

法政大学・情報メディア教育研究センター・教授

研究者番号：70434181

交付決定額(研究期間全体)：(直接経費) 10,200,000円

研究成果の概要(和文)：大学教育で利用されているオープンソースソフトウェアの多くは英語圏で開発されているため日本語への翻訳が必要となる。大学では複数のシステムにより教育支援のためのIT環境を構築しているが、システム間で用語が異なっているとユーザに違和感を抱かせる。この課題に対して、大学教育で頻出する用語および用例を包含する大規模な共通翻訳メモリを作成することによって一貫性のある翻訳を実現した。また、翻訳メモリによる翻訳では実現が困難なコンテキスト依存の翻訳を可能とするgettext形式のPOファイルを併用し、様々な教育支援システムをクラウドベースの翻訳支援システムを用い同一のプロセスで翻訳できる可能性を示した。

研究成果の概要(英文)：As most of the open-source softwares used in tertiary education was developed in English-speaking countries, translating the language to Japanese is a necessity. Though it is anticipated that an IT environment to support education through multiple systems will be built in universities in the future, differences in terminology between systems might create a sense of uneasiness in users. To resolve this issue, we can implement consistent translations by creating a large-scale common translation memory that includes terms and examples commonly used in university education. Additionally, by also using gettext-format PO files that are able to create context-sensitive translations, which cannot be achieved with translations from a translation memory, we demonstrated the possibility to translate various educational support systems using the same process with a translation support system.

研究分野：教育工学

キーワード：翻訳メモリー 学習支援システム オープンソースソフトウェア 国際化 大学教育 授業支援システム

1. 研究開始当初の背景

本研究は IT により大学教育を支援するシステムの研究・開発の一環として位置づけられる。この分野は教育工学として位置づけられてきたが、オープンソースソフトウェア(以下、OSS)による全学的なレベルでの大規模な学習支援システムの開発が行われるようになり、新たな研究領域としての定義が必要になっていた。2008 年には大学教育用 OSS の利用率が商用ソフトウェアを抜いてトップになり、同じく 2008 年には情報処理学会誌にて「e ラーニングの広がり」という特集号が発行された。2010 年には情報処理学会に CLE (Collaboration and Learning Environment) 研究会が設置され大学情報基盤の研究・開発が研究領域として定義された。国外ではカナダ The University of British Columbia にて開発されていたシステムが 1997 年に WebCT として商用化されたことを皮切りに、米国を中心にヨーロッパおよびオーストラリアの各大学で OSS を活用した教育システム開発プロジェクトが開始された。また本研究で扱う学習支援システムの Sakai CLE (以下、Sakai)を開発している Sakai コミュニティは 2003 年に設立されている。これらより国外では 2000 年前後にこの分野の開発がスタートしており、我が国のスタートを CLE 研究会設立の 2010 年とすると、大学教育を支援する OSS の開発は国外が 10 年程度先行していると言わざるをえない。

これまで本研究が関わる研究領域の動向について述べてきたが、本研究はこうした研究領域のなかでも各国が自国の言語にシステムを適用させるために実施しなければならない Localization (以下、l10n) の領域に位置づけられる。l10n の対象には「ソースコードに埋め込まれたメッセージ」および「Documentation としてのマニュアル」があり、いずれも翻訳がその主たる作業となる。文脈のある文章から構成される「Documentation としてのマニュアル」の翻訳は Google Translate、国内では情報通信研究機構 (NICT) にて開発された言語グリッドなどの機械翻訳で処理できるレベルになっており、本研究の Feasibility Study にて行った日本マイクロソフトでのヒアリングでもマニュアルは機械翻訳にて実用的に行われていた。しかしながら Web ページとして構成される「ソースコードに埋め込まれたメッセージ」は画面に表示されて初めて意味をもつ文章となるため機械翻訳には限界がある。また、この分野は大学教育用 OSS を開発している国際コミュニティでも非英語圏内の国がリーダーシップをもって推進していく必要があるが言語の壁に阻まれており l10n のプロセスは必ずしも確立できていない。

本研究のメンバは大学教育用 OSS に関する研究・開発を行ってきており、各大学の

要件にあったシステムの開発を行うとともに、Sakai あるいは Moodle の日本コミュニティの主要メンバとして年次国際カンファレンスにて日本の活動を発表するなど、大学教育用 OSS に関し国内外での展開を行ってきた。その一環として日本 Sakai コミュニティ (以下、Ja Sakai) においては新たな Sakai のリリースに向けた翻訳について継続性の観点から組織的な翻訳が望ましいとして 2012 年 8 月に翻訳をテーマとした 3 日にわたるアンカンファレンスを開催した。ここでは現状の翻訳の課題が整理されたが、Sakai だけでなくすべての大学教育用 OSS についても同様の課題があることが明らかになった。大学における OSS 利用が今後さらに拡大していくことを考えると、なるべく早い時期に大学連携によるプロジェクトを立上げこれらの課題を解決することが急務であるという結論に至った。

2. 研究の目的

Sakai および Moodle などの大学教育用 OSS の翻訳はベンダーではなく用語に慣れ親しんだ大学関係者が主体的に実施することが望まれる。例えば Course という単語を翻訳する場合に、授業科目、授業、科目、コース、講義、コマなどという候補のなかでコンテキストに応じて最適語を選択することは学外者には困難であろう。本研究はこの領域において、Sakai、Moodle だけでなく、今後大学教育に供されるすべての OSS の翻訳に対して共通の用語および共通プロセスを適用するための方法を確立し、それを国内外のコミュニティに展開することを目的としている。

3. 研究の方法

翻訳の課題は「A course management system is useful for instructors and students.」という文章からなる翻訳セグメントを訳す場合に、「コース管理システムは教員と学生にとって有意義である。」あるいは「授業支援システムは担当者と受講者にとって意味のあるものである。」など複数の翻訳候補が存在し一意的に訳せないことである。その解決策のひとつが翻訳セグメントの辞書とも言える図 1 に示すような翻訳メモリを使った翻訳である。本研究は、大学教育に供される OSS のうち多くの大学で利用されている学習支援システムである Sakai、Moodle、および ePortfolio システムの Mahara を翻訳対象として限定し、それらの翻訳を行う際に利用する大学向 OSS 共通翻訳メモリ(日本語版)を開発するものである。また、本研究で採用するコミュニティによる翻訳方法を国内外のコミュニティに展開することによって、本研究で対象外とした大学向 OSS についても同様のプロセスを適用できるようグローバルのレベルにて翻訳方法を確立する。

```

<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE tmx... "http://www.lisa.org/tmx/tmx14.dtd">
<tmx version="1.4">
<header adminlang="en" ... o-tmf="Transifex.com"/>
<body>
<tu>
<tuv xml:lang="en">
<seg>Not a member of the designated site</seg>
</tuv>
<tuv xml:lang="ja">
<seg>指定されたサイトのメンバではありません</seg>
</tuv>
</tu>
<tu>
<tuv xml:lang="en">
<seg>Select new Category for Page</seg>
</tuv>
<tuv xml:lang="ja">
<seg>ページに新しいカテゴリを選択</seg>
</tuv>
</tu>
(以下, <tu> ~ </tu>の組が続く)
</body>
</tmx>

```

図 1 翻訳メモリ事例

はじめに国際的に利用されている Sakai、Moodle、Mahara を対象として、各システムにおける英語表示箇所を翻訳支援ツールにより英日翻訳し、それぞれの翻訳メモリを作成する。翻訳支援ツールとしてはコミュニティによる翻訳をゴールとするため Web 形式であり、かつ翻訳メモリを利用できる翻訳システムである Transifex を用いる。次に作成された 3 つの翻訳メモリをひとつの翻訳メモリに統合し、重複する翻訳セグメントに対して複数の翻訳候補がある箇所を修正し 3 システム共通の翻訳メモリを作成する。この際約 30,000 組の翻訳セグメントが想定されるので、それらが含まれる翻訳セグメントを検索する共通翻訳メモリ作成支援ツールを開発する。

これまでに得られた共通翻訳メモリを使って Sakai に対して共通の用語を用いた翻訳を行うとともに、翻訳のワークフローを確立する。Sakai のソースコードは研究期間中に GitHub で管理されるようになったため、GitHub からソースコードをダウンロードし、翻訳対象ファイルをモジュールごとに PO ファイルとして整理し、それらを Transifex にアップロードするツールを開発する。PO ファイルは図 1 で示した翻訳メモリに類似した翻訳した結果がまとめられているファイルである。PO ファイルは翻訳メモリとは異なり、それを利用する学習支援システムの情報を有している。Transifex には先に開発した共通翻訳メモリをアップロードしておき、これを利用して翻訳を行う。この際、翻訳メモリでは類似した翻訳セグメントしか表示されないため、単語単位で用例を検索するシステムを開発する。すべての翻訳を完了した後、翻訳ファイルを PO ファイルとしてダウンロードし、それを Sakai のソースコードに展開するツールを開発する。最後にこれらのソースコードをビルドして、サーバの所定の位置に配置し、サーバを再起動して Web 画面上で翻訳箇所をチェックする。そして確認がで

き次第、翻訳前後のソースコードから Sakai のソースコードを管理する GitHub プロジェクトに Pull Request し、国際的に管理されているソースコードに翻訳結果を反映する。

ソースコードダウンロードから Web 画面上での翻訳チェックまでの一連の手順は翻訳の度に行なう必要があるが、継続的インテグレーションツールとして広く利用されている Jenkins により毎日定期的に行う。

以上により一連の翻訳ワークフローを示したが、GitHub、Jenkins、Transifex などの汎用性のあるシステムに加え、その間を埋めるシステムを開発し、包括的な翻訳環境を構築し、実際の Sakai の翻訳を行い、国際的に管理されている Sakai ソースコードに新たな翻訳を反映する実践的な方法を採用する。

4. 研究成果

(1) Sakai 翻訳

新たに開発した翻訳ワークフローにて Sakai の翻訳を 2 サイクル行い、その時点にて翻訳すべきソースコードをすべて翻訳した。それらの翻訳は GitHub にて国際的に一元管理されている Sakai ソースコードに反映した。日本語化した Sakai の画面事例を図 2 に示す。すべてのメッセージが日本語に翻訳されているほか、左列にあるナビゲーションメニューの説明も日本語に翻訳されていることがわかる。本ソースコードは Sakai コミュニティが管理し公開されている GitHub が

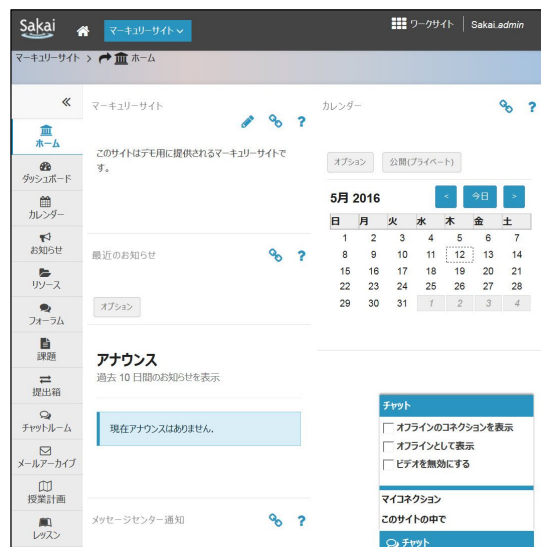


図 2 日本語化した Sakai 画面事例
ら誰もがダウンロードできる。

(2) 大学教育用共通翻訳メモリの開発

Sakai、Moodle、Mahara を Transifex にて翻訳した際に生成されるそれぞれの翻訳メモリを結合し、下記の浄化を適用して大学向け共通翻訳メモリを開発し、プロジェクト Web サイトにて公開した。開発された翻訳メモリのサイズは図 1 で示した <tu> で定義される英日翻訳セグメント換算で 33,554 組となった。

- 同一ソースが異訳された翻訳の統合
- 大文字小文字を識別する Case Sensitive により同一の単語に対して複数の翻訳がある場合の統合
- #, *, {}などの数学記号だけで構成されたソースになっている翻訳の削除
- Go, or, at などソースが2文字以下の単語の翻訳の削除
- 句読点(、)とカンマ, ピリオドの混在を統一

Sakai, Moodle, Mahara とは異なる学習支援システムを新たに翻訳する際にこの大学向共通翻訳メモリを利用すれば、多くの用語は Sakai, Moodle, Mahara と同じように翻訳することができる。

(3) 翻訳基盤および翻訳プロセスの整備

本研究開始時には Web ベースの翻訳基盤は一般的ではなかったが、研究期間中にソースコード管理システムとして Subversion に代わり GitHub が広く利用されるようになり、翻訳支援システムも Crowdin あるいは本研究で利用した Transifex という Web 対応のシステムが使われるようになった。本研究では、研究成果の横展開を念頭に置き、こうした汎用的な Web アプリケーション開発で利用される最新のサービスを採用した。

ここでは図3を参照して確立した翻訳ワークフローを説明する。なお、図中 TMX というマークが右肩に付記されているシステムは本研究にて開発したシステムを示す。

ワークフローのスタートは、国際コミュニティで管理する GitHub から英語のソースコードを引き出し、それを Web 版翻訳支援システムの Transifex に翻訳対象として設定するプロセスから始まる()。Transifex には開発した共通翻訳メモリを設定する。ソースコードの文字列が共通翻訳メモリにある英語文字列と完全一致した場合には設定により翻訳済となる。翻訳者は翻訳メモリから推奨される翻訳および過去の翻訳から得られた PO ファイルをデータベース化し、それを検索する POBrowser を併用し翻訳を進める()。Transifex に蓄積される翻訳は、Jenkins の設定により定期的にダウンロードされ()、さらにビルドおよびサーバに配置される()。翻訳者は Web 画面を通じて翻訳結果を確認する。すべての翻訳を完了後、翻訳前後のソースから Pull Request 用データを生成し、国際コミュニティが管理する GitHub に対して Pull Request を行う()。

開発した翻訳基盤および翻訳プロセスを示したが、本研究で開発したシステムは GitHub およびプロジェクト Web サイトにて公開しており、GitHub および Transifex とともにオープンソースプロジェクトでは無償で利用できるため OSS プロジェクトでの本システムの活用が期待できる。

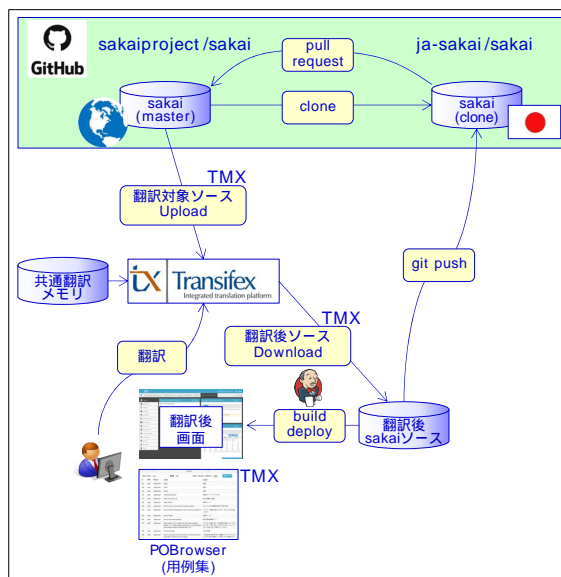


図 3 確立した翻訳ワークフロー

(4) 国際的コミュニティへの展開

Transifex による翻訳環境ではソースコードを Transifex にアップロードすれば、日本語に限らず他の言語の翻訳環境が設定できる。本プロジェクトでは Sakai コミュニティの 2013 年次国際大会にて研究成果のデモを行い、そこで興味を示したスペインコミュニティが本システムを利用することになった。さらに 2015 年度には同じ国際大会にてスペインコミュニティと共同セッションを行った結果、トルコおよびスウェーデンのコミュニティも本システムを利用して翻訳を行うことになった。

また、本研究で開発した翻訳基盤は Sakai だけでなく Moodle および Mahara などの翻訳にも適用できるため、それぞれの国内外の研究会にて提案を行った。さらに、国内のある大学にて開発しているオープンソースプロジェクトでは日本語で開発したシステムを英語版に翻訳した。

(5) 今後の研究課題

本研究のスタート時の課題として、ユーザが複数のシステムを利用して学習する際に、同一英語ソースに対して異なる日本語翻訳があるとユーザが違和感を覚えるという仮説の設定をした。しかしながら、実際に感じる違和感は単なる翻訳だけの課題では収まるわけではなく、現時点では下記のように整理しており、今後の研究課題となっている。

概念を知らないことからの違和感

学習支援システムおよび SNS などのツール特有の用語に起因し、学習支援システムでは、「問題プール」, 「問題バンク」, SNS では「ウォール」, 「タイムライン」, 「シェア」, 「コネクション」などが該当する。学習者がこれらの用語になじみがないと違和感を覚える。

日本語として不自然なことによる違和感受身形の多用および文化的差異に起因し、例として「ダウンロードできるファイル」「大学プロフィール」などがあり、それぞれ「添付ファイル」「大学について」などと翻訳することが望ましい。

英語をカタカナ化していることによる違和感
言語の意味のまま日本語として扱っているため、似て非なる意味になってしまっていることがある。例として、「アクティビティ」「テンプレート」「フィールド」「フォーム」「サイト」などが該当する。

5. 主な発表論文等

〔雑誌論文〕(計 1 件)

常盤祐司、出口大輔、宮崎誠、平岡齊士、喜多敏博、梶田将司、教育用オープンソースソフトウェア群のローカライゼーションと共通翻訳メモリの開発 - 一貫性のある用語による教育支援システムを目指して -、情報処理学会デジタルプラクティス、査読有、Vol.6 No.2、2015、pp79-88、<http://ci.nii.ac.jp/naid/170000121436>

〔学会発表〕(計 14 件)

常盤祐司、藤井聡一郎、出口大輔、梶田将司、Open Apereo 2015 Conference 参加報告、情報処理学会、2015 年 12 月 6 日、JR 福井駅 AOSSA(福井県・福井市)

Makoto Miyazaki、Naoshi Hiraoka、Yuji Tokiwa、Daisuke Deguchi、Toshihiro Kita、Shoji Kajita, A proposal for technique to use common terms among multiple systems - Common translation memory by Mahara, Moodle and Sakai CLE, mahara UK, 2015 年 11 月 10 日, Southampton, UK

長岡千香子、平岡齊士、喜多敏博、Moodle とポートフォリオ・SNS との連携のコツ - 翻訳カスタマイズとプラグイン、MoodleMoot Japan 2015、2015 年 2 月 20 日、京都産業大学(京都府・京都市)

常盤祐司、藤井聡一郎、宮崎誠、出口大輔、平岡齊士、梶田将司、Open Apereo 2014 Conference 参加報告、情報処理学会、2014 年 10 月 24 日、東北大学(宮城県・仙台市)

常盤祐司、宮崎誠、出口大輔、平岡齊士、梶田将司、Open Apereo 2013 Conference 参加報告、情報処理学会、2013 年 12 月 15 日、琉球大学(沖縄県・中頭郡)

〔その他〕

ホームページ等

<http://www.sakaiproject.jp/tmx/>

<https://github.com/JaSakai>

6. 研究組織

(1) 研究代表者

常盤 祐司 (TOKIWA, Yuji)

法政大学・

情報メディア教育研究センター・教授

研究者番号：70434181

(2) 研究分担者

出口 大輔 (DEGUCHI, Daisuke)

名古屋大学・情報連携統括本部・准教授

研究者番号：20437081

喜多 敏博 (KITA, Toshihiro)

熊本大学・eラーニング推進機構・教授

研究者番号：20284739

宮崎 誠 (MIYAZAKI, Makoto)

畿央大学・教育学部・特任助教

研究者番号：60613065

梶田 将司 (KAJITA, Shoji)

京都大学・

学術情報メディアセンター・教授

研究者番号：30273296

平岡 齊士 (HIRAOKA, Naoshi)

熊本大学・社会文化科学研究科・准教授

研究者番号：80456772