

修士論文

多様なサービスの連携による ローカリゼーションプロセスの効率化

指導教官 石田 亨 教授

京都大学大学院情報学研究科
修士課程社会情報学専攻

古白川 亮太

平成 23 年 2 月 9 日

多様なサービスの連携による ローカリゼーションプロセスの効率化

古白川 亮太

内容梗概

現在、ローカリゼーションは翻訳元言語と翻訳先言語の両方に精通したバイリンガルが行っている。また、一度翻訳を行ったフレーズを蓄積し、その内容を利用する翻訳メモリがローカリゼーションにおいて活用されている。しかし、翻訳作業は人間が行っているため、コストが問題となる。低コストで翻訳を行う方法として機械翻訳があるが、機械翻訳は未だに精度が低いため、高い翻訳精度が求められるローカリゼーションにおいては機械翻訳は活用されていない。そこで、本研究では、ローカリゼーションにかかるコストを下げるために、機械翻訳を活用する翻訳プロセスについて提案する。

機械翻訳を用いることにより、バイリンガルによる翻訳よりも低いコストで翻訳を行うことが出来る。しかし、機械翻訳の精度はローカリゼーションに用いることが出来るほど高くはないので、人間による翻訳を行う必要がある。先行研究から、機械翻訳の結果をモノリンガルが修正し、その結果をバイリンガルがチェックするというプロセスで翻訳を行うと、バイリンガルのみが翻訳を行うというプロセスでの翻訳に比べ、低いコストで翻訳を行うことが出来るということが知られている。しかし、先行研究では、実際のローカリゼーションの現場で行われている翻訳用の専門辞書作成などの作業や、翻訳元言語から一度他の言語に翻訳し、その翻訳結果を更に他の言語に翻訳するマルチホップ翻訳について考察されていないので、本研究では、実際のローカリゼーションに則した翻訳プロセスの設計を行う。

また、本研究では、実際のローカリゼーションで翻訳コストを下げる事が出来る翻訳プロセスの提案を目的としているので、設計した翻訳プロセスの有効性を調査する必要がある。設計した翻訳プロセスで実際にローカリゼーションを行い、翻訳元の文章の性質や、用いた機械翻訳の品質、モノリンガル作業者の作業品質と、翻訳のトータルコストの関係を調査し、既存の翻訳プロセスと比べて翻訳コストを下げる事が出来る条件について考察を行う。どのような状況において、提案した翻訳プロセスを用いれば既存のプロセスでの翻訳に比べ翻訳コストを下げる事が出来るかを提唱することを本研究の目標とする。本研究で取り組む課題は以下の二点である。

翻訳プロセスの構築 現在、ローカリゼーションではバイリンガルが翻訳メモリなどの翻訳支援ツールを用いて翻訳を行っているが、機械翻訳は活用されていない。そこで、機械翻訳を活用し、翻訳コストを下げる事が出来るような翻訳プロセスを構築する。機械翻訳結果をモノリンガルが修正し、その結果をバイリンガルが修正するというプロセスで翻訳を行うと、バイリンガルのみで翻訳を行うよりも低いコストで翻訳を行えるということが知られている。本研究では、それを元に、実際のローカリゼーションに則した、実際に翻訳作業のコストを削減することが出来るような翻訳プロセスの構築を行う。

翻訳プロセスの有効性の検証 本研究では、構築した翻訳プロセスで実際に翻訳コストを削減することを目標としているので、構築した翻訳プロセスの有効性を検証する必要がある。そのため、構築した翻訳プロセスでローカリゼーションを行い、翻訳元文章の性質や、機械翻訳の品質、モノリンガル作業者の作業品質と、翻訳にかかるトータルコストの関係を調べ、どのような条件において翻訳コストを削減出来るかということ調査する。

また、本研究では、上記のプロセスで翻訳を行う際に発生する問題を解決するためのツールの設計を行う。複数の作業者がデータを共有して作業を行う際、効率的に作業を行うために、Web上で翻訳作業を行うツールを設計する。

本研究の貢献は以下の二点である。

翻訳プロセスの有効性の検証 上記で構築した翻訳プロセスの有効性について検証を行った。ローカリゼーションコストの削減には、モノリンガル修正文の品質とモノリンガル修正率が大きく影響するという結果が得られた。また、このプロセスでローカリゼーションを行う際、機械翻訳の品質はコストにあまり影響しないという結果を示すことが出来た。

翻訳支援ツールの設計 上記のプロセスでローカリゼーションを行う際に作業を支援するツールの設計を行い、ローカリゼーションプロセスの効率化に貢献した。

Improvement of Localization Processes by Composing Various Services

Ryota Koshirakawa

Abstract

Localization is performed by bilingual people currently, and translation memory is used. Translation memory is a system that memorizes translation result and reuse that in localization. But translation is performed by human, so localization comes at a problem of cost. If we use machine translation, the cost of translation would be cheaper. But adequacy of machine translation is bad, so machine translation can't be used at localization. In this paper, I suggest localization processes using machine translation to reduce the cost of localization.

If we use machine translation, the cost of translation would be cheaper. But adequacy of machine translation is not enough to localization. So localization should be performed by bilingual people. It is known by the previous study that the cost of translation would be cheaper when localization is performed by machine translation, monolingual people and bilingual people. Monolingual people check the result of machine translation, and Bilingual people check the result of it. But in the previous study, the cost to make the dictionary, or the processes of multi-hop translation was not considered. In this paper, I consider translation processes to real localization.

The purpose of this paper is to suggest translation processes that can reduce the cost of localization, so I have to consider that the effectiveness of the translation processes. I collected the data in the actual localization, and considered the relationship between the tendency of original sentence, quality of MT, quality of working result of monolingual people and localization cost. And I considered the requirement to cut the localization cost. The purpose of this paper is to declare the requirement to cut the localization cost.

Solved problems by this research are as follows.

Suggestion of translation processes Localization is performed by bilingual people, but machine translation is not used in localization. So I suggested translation processes. In these processes, monolingual speakers

check the result of machine translation and bilingual speakers check the result of it. It is known by the previous study that the cost of translation would be cheaper in this process. And based on that, I proposed new translation processes to reduce translation cost.

Verification of effectiveness of localization processes The purpose of this research is to suggest of localization processes, so I have to verify effectiveness of localization processes. So I consider that relationship between the tendency of original sentence, quality of MT, quality of working result of monolingual people and localization cost.

And in this research, I designed a tool that help monolingual and bilingual localization. When a number of workers translate while sharing data, this tool make it possible to they share data on the web.

The contributions of this research are as follows.

Verification of effectiveness of localization processes I verified effectiveness of localization processes. I assert that localization cost is affected by quality of monolingual check and percentage of sentences that monolingual can fix. And I assert that localization cost is not affected by quality of machine translation.

Designing the localization support tool I designed a localization support tool and made it efficient localization possible.

多様なサービスの連携による ローカリゼーションプロセスの効率化

目次

第1章 はじめに	1
第2章 研究背景	4
2.1 既存のローカリゼーション手法	4
2.2 機械翻訳	5
2.3 言語グリッド	6
第3章 ローカリゼーションプロセスの提案	8
3.1 ローカリゼーションプロセス	8
3.1.1 辞書作成プロセス	8
3.1.2 翻訳プロセス	12
3.2 機械翻訳とモノリンガル修正	13
第4章 実験の手法	18
4.1 実験手法	18
4.2 評価基準	21
第5章 評価	22
5.1 機械翻訳の品質と翻訳コストの関係	22
5.2 翻訳の品質の評価	27
5.3 モノリンガル修正文の品質と翻訳コストの関係	29
5.4 モノリンガル修正率と翻訳コストの関係	31
5.5 ローカリゼーションの効率化のための条件	34
第6章 翻訳支援ツール	37
6.1 ローカリゼーションワークフロー	37
6.2 協調翻訳	38
第7章 おわりに	42
謝辞	44
参考文献	45

付録：翻訳文の実例

A-46

- A.1 商店街ドキュメント..A-エラー! ブックマークが定義されていません。
- A.2 大学ドキュメント.....A-エラー! ブックマークが定義されていません。
- A.3 寺院ドキュメント.....A-3

第1章 はじめに

近年のグローバル化に伴い、様々な場面でローカリゼーションの需要が増加している。特に、企業案内やプレスリリース、WEB コンテンツ、事業計画書、各種報告書などのローカリゼーション需要は、グローバル企業であればどのような業種でも発生し、今日では翻訳市場で大きな割合を占めている。そのため、ビジネスの現場におけるローカリゼーションをツールを用いて支援することは重要である。

ローカリゼーションの支援を行うツールとして、機械翻訳機がある。機械翻訳機は、大量の文章を翻訳するのに適しているため、ローカリゼーションに機械翻訳を活用出来れば、翻訳コスト及び翻訳時間の削減を行うことが可能になると期待出来る。しかし、機械翻訳の翻訳品質は現状ではまだ十分ではなく、流暢さ、正確さともにバイリンガルによる翻訳と同等の品質は期待できない。そのため、製品のマニュアルや契約書など、実際のビジネスの現場でのローカリゼーションにおいては、バイリンガルが翻訳を行っている。

二言語を流暢に操るバイリンガルの数は限られており、今後ますますローカリゼーションの需要が増していくと、バイリンガルの人的コストは増大していくと考えられる。そのため、機械翻訳をローカリゼーションの現場で活用することで、バイリンガルの作業量を減らし、人的コストの削減を行いたい。また、ローカリゼーションには専門的な分野も多く、二言語を流暢に扱えるとしても、専門用語の翻訳が難しい場合がある。

翻訳の品質を高めるための仕組みとして、言語グリッドがある。言語グリッドでは、機械翻訳機の品質の向上以外の方法で翻訳精度を向上させるために、いくつかの複合サービスが実装されている。その例として、辞書連携翻訳がある。辞書連携翻訳とは、ある語句に対する訳語を予め対訳辞書の形で用意しておき、機械翻訳と連携させることで、文章中の指定語句を強制的に意図通りの訳語に翻訳させるものである。

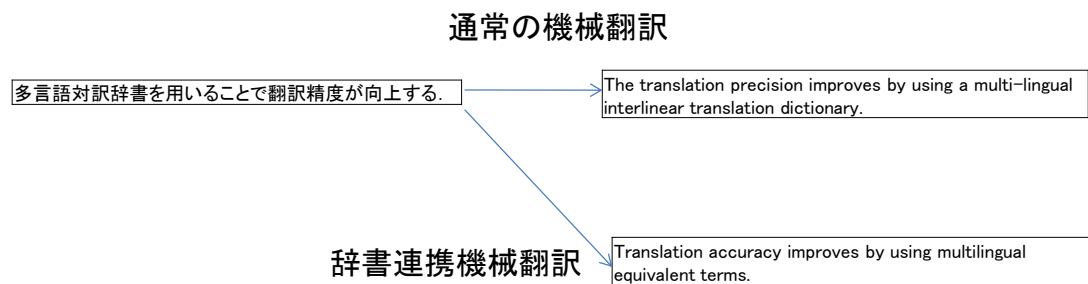


図 1：辞書連携機械翻訳の例

この機能は、ユーザの意図を反映した翻訳結果が得られることで、誤訳を減らすことが出来るという点で非常に有用なものである。また、言語グリッドは、折り返し翻訳機能や言語資源作成機能等により、人間とサービスの連携を可能とした。これらのサービスにより、様々な多言語コミュニケーションを支援することが可能となった。

本研究では、言語グリッドを用いたビジネス用のローカリゼーションプロセスの可能性について考察する。先行研究から、機械翻訳文をモノリンガル作業者が修正し、さらにそれをバイリンガル作業者が修正するというプロセスでローカリゼーションを行うことで、バイリンガル作業者のみが翻訳を行う場合と比較して安いコストでローカリゼーションを行える可能性があるということが知られている。

本研究では、先行研究で提示されたローカリゼーションプロセスの有効性を調査するために、先述のローカリゼーションプロセスでの翻訳実験を、先行研究よりも多様な、そして大量のドキュメントを用いて行った。また、複数の言語ペアについて、翻訳を行った。そこから得られたデータから、モノリンガル修正率、機械翻訳文の品質、モノリンガル修正文の品質、コスト削減率といった指標にどのような相関があるのかの分析を行った。

また、機械翻訳とモノリンガルとバイリンガルの共同でのローカリゼーションにおいて発生する問題の指摘と、その問題を解決するためのツールの設計を行った。

本論文では、以下のような構成で議論を進めていく。

まず、第 2 章において、現在実際の現場において行われているローカリゼーションの手法の紹介を行う。また、ローカリゼーションプロセスの改善におけるポイントとなる機械翻訳と言語グリッドについての紹介を行う。

第3章では、機械翻訳とモノリンガル修正とバイリンガル修正というプロセスでのローカリゼーションプロセスと、そのプロセスを用いた場合のコスト削減の可能性について、先行研究を交えての説明を行う。また、辞書連携翻訳のための専門用語辞書作成にかかるコストについての考察も行う。

第4章では、機械翻訳とモノリンガル修正とバイリンガル修正というプロセスでのローカリゼーション実験の手法についての説明を行う。また、計測するデータと分析のための指標についての解説を行う。

第5章では、実験から得られたデータを元に、上記のプロセスでローカリゼーションを行った場合の、コストや機械翻訳文の品質等の条件にどのような相関がみられるかの分析を行う。

第6章では、上記のプロセスでローカリゼーションを行う際に発生する問題点の指摘を行う。さらに、それを解決するためのツールの設計を行う。

最後に、第7章で本論文のまとめを行う。

第2章 研究背景

2.1 既存のローカリゼーション手法

実際のローカリゼーションの現場で用いられている手法について説明する。ローカリゼーションを受注して行っているS社の翻訳ワークフローは以下のようになっている。

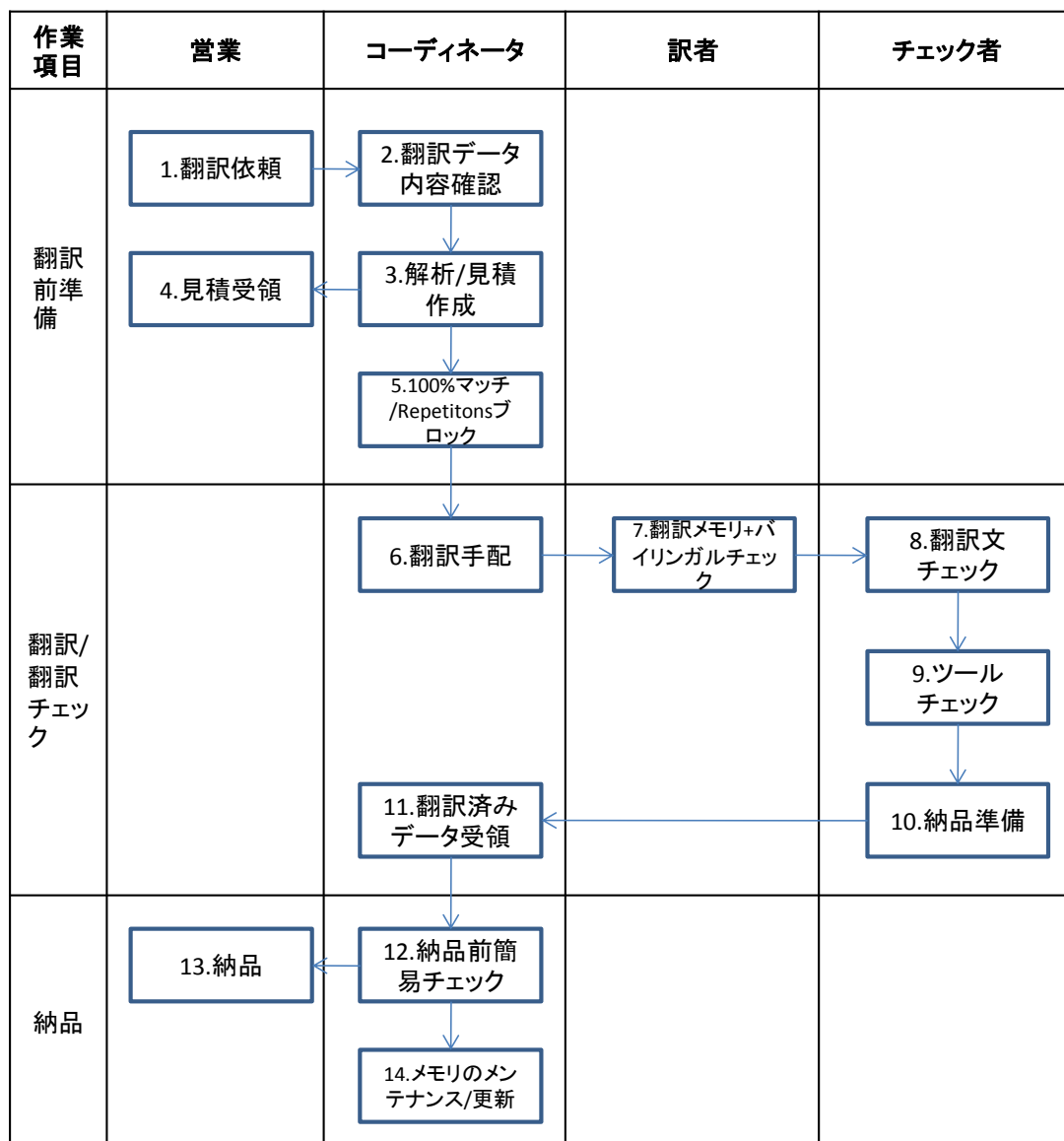


図 2 : S 社の翻訳ワークフロー

翻訳部分の作業においては、まず、5 の 100%マッチ/Repetitions ブロックで翻訳メモリを参照して、既に対訳が存在する文章を翻訳対象から除く。そして、翻訳メモリに対訳が存在しない文章については、7 でバイリンガル作業者が翻訳メモリを参照しながら翻訳を行う。そして、8 で日本人がその翻訳作業の結果のチェックを行い、さらに9 でツールを用いたチェックを行うというフローで作業が行われる。なお、翻訳メモリとは、既存の対訳データの集合を用いることで、類似する対訳文が存在する場合において精度の高い翻訳を行うシステムである。

機械翻訳は、翻訳メモリを用いたとしても、現状ではまだ精度が悪く、機械翻訳機によって翻訳された文章をそのまま対訳として用いることは出来ない場合がほとんどである。そのため、このワークフローにおいては、7 のバイリンガルによる翻訳作業に時間がかかっており、ボトルネックとなってしまう。

ローカリゼーションにおいて機械翻訳を用いてバイリンガル翻訳者の負担を減らすことは出来ていないため、この部分に改善の余地がある。よって、バイリンガル作業者の作業量を減らす手法の提案についての考察を第3章以降で行っていく。

2.2 機械翻訳

翻訳の支援を行うツールとして、機械翻訳がある。機械翻訳とは、ある言語で書かれた文章を別の言語に機械的に翻訳するサービスである。機械翻訳機の例として、Google Translate, WEB-Transter, J-Server, Translution, YakushiteNet 等がある。全ての言語の組み合わせに対して翻訳機を作るとは現実的でないので、非英語同士の機械翻訳は、英語をハブとしてマルチホップに行われることが多い。英語-非英語間の言語資源は、非英語-非英語間と比べ、豊富であるため、英語をハブとすることで比較的精度のよい翻訳が期待出来る。

機械翻訳では、翻訳作業を機械が自動的に行うので、大量のドキュメントを翻訳するのに適している。しかし、機械翻訳にはいくつかの問題がある。未知の単語は翻訳出来ないといった点や、訳語選択の非一貫性が存在するといった点により、機械翻訳の精度は現状では満足なものとは言えず、ビジネスにおけるローカリゼーションに通用するものではない。[1]

機械翻訳の精度を向上させるための研究として、用例ベース翻訳の研究がある。用例ベース翻訳とは、対訳コーパスを学習データとして用いて行う翻訳である。用例ベース翻訳では、対訳コーパスの量が増えるに従って、翻訳の精度もよくなっていく。そのため、インターネットを通じて大量のコーパスを入手できるようになった昨今では、用例ベース翻訳の精度の向上も期待でき、今後の実用化も期待されている。

また、機械翻訳機の精度の向上以外のアプローチで翻訳の精度の向上を行う試みとして、言語グリッドがある。本研究でも、機械翻訳の精度の向上のために、言語グリッドで実装されたサービスを用いている。次節では、言語グリッドについての説明を行う。

2.3 言語グリッド

言語グリッドプロジェクト¹では、言語・文化の壁を乗り越えた異文化コラボレーションのために、言語サービスの利便性と有用性を高めるための活動を行っている。

現在、インターネット上には、様々な辞書や機械翻訳機などの言語資源が蓄積され、有償、無償を問わず様々な形でサービスとして提供され始めている。しかし、これらの言語サービスは必ずしも使い勝手のいいものではない。そのため、インターネット上の既存のサービスを自由に組み合わせることで、新たな言語サービスを生み出し、利用するといったことや、ユーザ自らが新たな言語資源や言語処理機能を追加していくといったことが可能な枠組みとして、言語グリッドの開発が行われた。

言語グリッドでは、機械翻訳機の精度の向上以外の方法で翻訳精度を向上させるために、複数の言語資源を組み合わせた複合サービスがいくつか実装されている。そのうちの 하나가、辞書連携翻訳である。辞書連携翻訳とは、ある語句に対する訳語を予め対訳辞書の形であらかじめ用意しておき、機械翻訳と連携させることで、機械翻訳を行う際に文章中の指定語句を強制的に意図通りの訳語に翻訳させるものである。この機能は、ユーザの意図を反映した翻訳結果が得られることで、誤訳を減らすことが出来るという点で非常に有用なものである。また、機械翻訳機に登録されていない未知の用語の翻訳も可能となるので、ローカリゼーションのように専門用語が多数含まれるドキュメントの翻訳

¹ <http://langrid.nict.go.jp/>

において有効に機能すると言える。

本研究では、後述のローカリゼーション実験の機械翻訳の部分において、翻訳精度の向上のために、言語グリッドの辞書連携翻訳機能を用いた。

第3章 ローカリゼーションプロセスの提案

3.1 ローカリゼーションプロセス

本章では、先行研究を元に、既存のバイリンガルのみのローカリゼーションと比較して安いコストでローカリゼーションを行うことが出来る可能性のあるプロセスについての説明を行う。

また、言語グリッドで辞書連携翻訳を行うにおいて必要となる専門用語辞書の作成プロセス、及びそのコストについての考察も行う。

3.1.1 辞書作成プロセス

言語グリッドの辞書連携翻訳を用いることによる翻訳の精度の向上を可能とするために、辞書連携翻訳用の専門用語対訳辞書を作成するというプロセスが必要となる。そのため、実際のローカリゼーションワークフロー全体でのコスト及び時間について考察するためには、専門用語対訳辞書作成にかかるコスト及び時間についても考察する必要がある。

後述のローカリゼーション実験においては、通常翻訳時との専門用語辞書作成以外の翻訳部分での時間及びコストの比較を行ったため、辞書作成のコストは考察対象となっていないが、本研究では実験に伴い、専門用語対訳辞書作成のコストについての調査も行ったので、その結果についての説明を本節で行うこととする。

本研究では、翻訳対象となるドキュメントから単語を抽出し、それを多言語に翻訳し対訳辞書を作成するという作業を行った。この作業は、図 3 のようなプロセスで行われた。

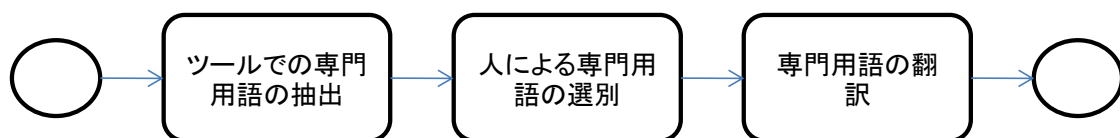


図 3：専門用語辞書作成プロセス

まずツールを用いて文章から単語のみを抽出する。ここでは、抽出に言選 Web¹を用いた。言選 Web では、入力された文章の形態素解析を行い、区切られた語句の中から名詞以外であると判定された単語を除くことで名詞と思われ

¹ 言選 Web <http://gensen.dl.itc.u-tokyo.ac.jp/gensenweb.html>

る単語のみを抽出し、さらに、そこから重複する単語を削除したものを出力として表示する。言選 Web を用いて得られる単語リストは下記の表 1 のようになる。

表 1：言選 Web を用いて抽出された単語リスト

研究経過報告書	修士課程	情報システム設計論演習II
研究	研究経過	社会情報学専攻基礎科目
研究経過報告	論文	学生
アドバイザー	研究指導	発表
指導教員	研究論文	情報学研究科
修士論文	教員	研究成果
専攻	情報システム分析論演習	社会情報学専攻外
社会情報学専攻	査読シート	研究活動
報告	研究室	研究科事務
社会情報学特殊研究 1	査読委員	社会情報学特殊研究 2
指導	経過報告	査読
修士	修士論文提出	研究発表会
単位	情報システム分析論	研究科開設科目
研究報告	研究科	社会情報学専攻学生
社会情報学特殊研究	研究科教務	アドバイザー制
科目	情報システム設計論演習I	情報学研究科教務
提出	先生方	研究報告会
公聴会	研究経過報告書提出	依頼書
報告書	修士論文査読	321 情報社会論全員社会情報 2

このように、ツールを用いて抽出した単語リストには、専門用語と思われる単語、一般用語と思われる単語、名詞でない語が含まれる。ここから、専門用語と思われる単語のみを抽出したいが、形態素解析ソフトでは専門用語か一般用語かの判別を行うことが出来ないため、人間が手動で専門用語のみの抽出を行う必要がある。これが、先述の専門用語辞書作成プロセスにおける、人による専門用語の選別というステップである。このようにして、まずツールを用いて単語を抽出した後、人間による選別を行って作成した単語リストを、翻訳するというステップで専門用語辞書は作成される。

この部分の作業の改善点として、ツールを用いて抽出された単語リストに対して、一般用語辞書との比較を行い、一般辞書に出現する単語は専門用語ではないと判定し自動的にリストから除くなどの手法が考えられる。しかし、一般辞書に登場するような単語であっても、翻訳の分野によっては、特殊な対訳を与えることがあり、専門用語辞書に含めることが望ましい場合もあるので、この手法でも専門用語辞書として望ましい単語リストが得られるとは限らない。専門用語リストの生成を機械的に行うには更なるプロトコルの検討が必要であ

と思われる。

日本語ドキュメントからツールを用いて単語を抽出する際の、ドキュメントの文字数と、抽出された単語数の関係を表 2 に示す。

表 2：日本語ドキュメントの文字数あたりの抽出単語数

	ドキュメント A	ドキュメント B
文字数	31223 文字	23869 文字
抽出された 単語数	1637 単語	2252 単語
文字数あたりの 単語数	52 単語/1000 文字	94 単語/1000 文字

ドキュメント A は不動産関連のドキュメント、ドキュメント B は観光関連のドキュメントである。ドキュメント B は、観光関連のドキュメントで、専門用語が多く含まれるので、ドキュメント A よりも文字数あたりの単語数がかなり多くなっている。ドキュメントの種類によって含まれる単語の数にばらつきがあることがわかる。

表 2 のデータから平均を求めると、ドキュメントからは、71 単語/1000 文字の単語が抽出されることになる。

抽出した単語リストから必要な単語の選別を行う際の、選別された単語数と、選別にかかった時間を表 3 に示す。

表 3：単語選別にかかる時間及び単語の適合率

	ドキュメント B	ドキュメント C
選別前の単語数	2252 単語	5100 単語
選別後の単語数	668 単語	1081 単語
適合率	30%	21%
選別にかかった時間	150 分	200 分
単語数あたりの選別時間	67 分/1000 単語	40 分/1000 単語

ドキュメント C は大学関連のドキュメントである。なお、ドキュメント B と C については、異なる作業者が作業を行っている。また、各作業者はいずれもモノリンガル作業者である。表 3 から、作業者や対象となるドキュメントの種類によって、選別にかかる時間にもばらつきがあるということがわかる。

単語の翻訳作業については、バイリンガルが行った。作業時間は、1 単語あたり 1.5 分である。

また、表 3 のデータから平均を求めると、ツールで抽出された単語の適合率は 24%、単語の選別にかかる時間は 48 分/1000 単語となる。

以上から、辞書作成に必要な時間とコストを表 4 にまとめる。

表 4：辞書作成に必要な時間及びコスト

作業内容	文字数あたりの単語数	文字数あたりの作業時間	文字数あたりのコスト
単語抽出	71 単語 /1000 文字	0 分	0 円
単語選別	17 単語 /1000 文字	1 分	17 円
単語翻訳	17 単語 /1000 文字	26 分	867 円

なお，ここでは，後述の翻訳作業の場合と同様に，バイリンガル作業者の時給は 2000 円，モノリンガル作業者の時給は 1000 円という設定を用いている。表 4 から，辞書作成には，ドキュメント 1000 文字につき，27 分，884 円がかかるということがいえる。

3.1.2 翻訳プロセス

既存の，バイリンガルによるローカリゼーションプロセスは，図 4 のようになる。

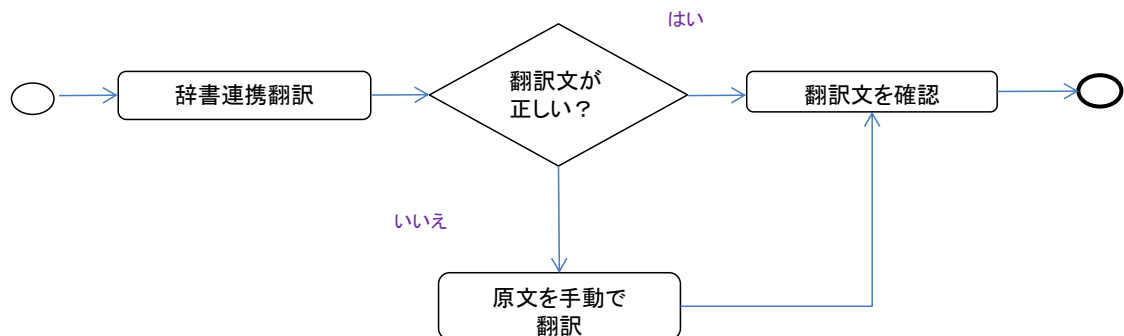


図 4：バイリンガルによるローカリゼーションプロセス

現状では上記のプロセスでローカリゼーションが行われているが，先述の通り，機械翻訳はローカリゼーションにおいて活用されているとは言えず，バイリンガル作業者の負担が大きくなっている。この問題を解決するための手法として，上記のローカリゼーションプロセスにモノリンガル作業者を組み込むことで，バイリンガル作業者の作業時間を減らし，ローカリゼーションに必要な

トータルのコストを削減出来る可能性があることを示した先行研究がある。[2]

機械翻訳を行った後の文章をモノリンガルが確認、修正し、それをバイリンガルがチェックするというプロセスでローカリゼーションを行う。このプロセスのワークフローは図 5 のようになる。

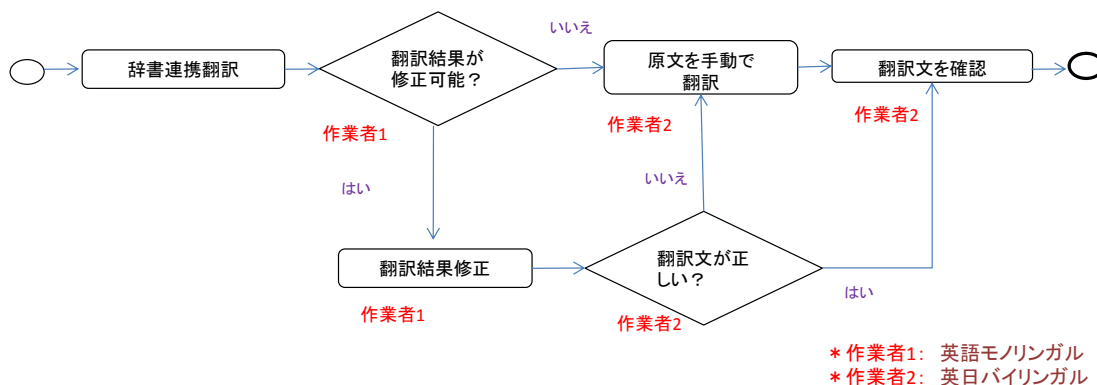


図 5: モノリンガル+バイリンガルによるローカリゼーションのプロセス

このプロセスでローカリゼーションを行った場合と、既存のプロセスでローカリゼーションを行った場合でのコストの比較を行った研究について、3.2 節以降で説明する。

3.2 機械翻訳とモノリンガル修正

人間とサービスの連携によるローカリゼーションプロセスに関する先行研究として、Lin らの行ったマニュアル翻訳実験がある。この実験では、日本語のドキュメントを中国語に翻訳するという作業を、日中バイリンガル作業員と、中国語モノリンガル作業員が行い、バイリンガル作業員のみで作業を行った場合と、バイリンガルとモノリンガルが共同で作業を行った場合のそれぞれの場合について、かかった作業時間を計測し、コストの比較を行った。バイリンガル作業員とモノリンガル作業員が共同で作業を行う場合は、まず、モノリンガル作業員が機械翻訳結果を評価し、そのうち一定以上の評価のものについては翻訳文を流暢に修正するという作業を行い、その後、バイリンガル作業員が、モノリンガル作業員が修正した文をチェックするという順序で作業が行われた。なお、この実験では、C 社の電子機器紹介 (17 文, 709 文字) を翻訳の対象とし、言語は日本語から中国語への翻訳を行った。

この実験において、翻訳文の適切さは、五段階評価で表される。まず、この

ドキュメントに対して通常の機械翻訳を行った場合の翻訳結果の適切さの評価と、言語グリッドを用いた辞書連携翻訳を行った場合の翻訳結果の適切さの評価を下記の表 5 に示す。

表 5：通常の機械翻訳文と辞書連携機械翻訳文の Adequacy 評価

Adequacy 評価	中国語訳文（専門辞書なし）		中国語訳文（専門辞書あり）	
	文字数	Percentage	文字数	Percentage
5 (All)	18	2%	35	5%
4 (Most)	217	31%	340	48%
3 (Much)	185	26%	235	33%
2 (Little)	164	23%	99	14%
1 (None)	125	18%	0	0%
Total	709	100%	709	100%

通常の機械翻訳では、Adequacy3 以上の文章は 59%だが、辞書連携翻訳を用いることで、Adequacy3 以上の文章が 86%に増えていることがわかる。辞書連携翻訳は機械翻訳文の翻訳精度を向上させているといえる。

次に、辞書連携機械翻訳の評価結果と、モノリンガル作業者が修正した文章の評価結果を表 6 に示す。

表 6：辞書連携機械翻訳文とモノリンガルによる修正文の Adequacy 評価

Adequacy 評価	辞書連携機械翻訳文	モノリンガルによる修正文
5(All)	5%	56%
4(Most)	48%	15%
3(Much)	33%	15%
2(Little)	14%	14%

1(None)	0%	0%
---------	----	----

モノリンガルは、適切さの評価が 3 以上の文章について修正を行った。表 6 から、モノリンガルが修正を行うことで、多くの文章の適切さが向上していることがわかる。

次に、翻訳形態ごとの、モノリンガル作業員及びバイリンガル作業員の作業時間、コストを示す。

表 7：翻訳形態ごとの、モノリンガル作業員及びバイリンガル作業員の作業時間及びコスト

翻訳形態	Resource	Time	Cost
通常の人手での翻訳作業	日中翻訳者 1 名	30 分 翻訳	1333 円
		10 分 校正	
翻訳メモリを利用した翻訳作業	日中翻訳者 1 名 (翻訳メモリ利用率 15% の場合)	25 分 翻訳	1167 円
		10 分 校正	
機械翻訳（専門辞書なし）を利用した翻訳作業	日中翻訳者 1 名	12 分 翻訳 (41% 文書)	1100 円
		15 分 校正	
	中国語修正者 1 名	12 分 修正 (59% 文書)	
機械翻訳（専門辞書あり）を利用した翻訳作業	日中翻訳者 1 名	4 分 翻訳 (14% 文書)	867 円
		12 分 校正	
	中国語修正者 1 名	20 分 修正 (86% 文書)	

ここでは、バイリンガル作業員の時給は 2000 円、モノリンガル作業員の時給は 1000 円と設定した。この設定は、翻訳を行っている企業から得られたデータである。表 7 から、バイリンガルとモノリンガルが共同で作業を行うことで、時間単価の高いバイリンガル作業員の作業時間を減らすことが可能となっており、翻訳全体でかかるコストを下げることに成功していることということ

が言える。

表 8 に、バイリンガルによる翻訳の速度と、モノリンガルとバイリンガルによる翻訳の速度の比較をまとめる。この表から、バイリンガルによる翻訳と、モノリンガルとバイリンガルによる翻訳の、翻訳速度の差は小さいということがわかる。

表 8：バイリンガルによる翻訳と、モノリンガルとバイリンガルによる翻訳の翻訳作業全体の時間

翻訳形態	モノリンガルによる修正作業	バイリンガルによる翻訳作業	合計
バイリンガルによる翻訳	なし	49.37 分 /1000 文字	49.37 分 /1000 文字
モノリンガルによる修正とバイリンガルによる確認, 修正	28.21 分 /1000 文字	22.57 分 /1000 文字	50.78 分 /1000 文字

先行研究では、日中翻訳のみについて考察していたが、本研究では、日本語->中国語のみではなく、日本語->英語、日本語->韓国語、英語->ドイツ語、英語->フランス語、英語->ポルトガル語、等の様々な言語間での翻訳、さらに、日本語->英語->フランス語といったマルチホップで翻訳を行う場合についての実験を行った。

また、先行研究では、709 文字 17 文の比較的短い文章に関して実験を行ったが、本研究では、様々な種類の大量のドキュメントに対して、先述のプロセスでローカリゼーションを行い、データを収集した。ローカリゼーションの対象となったドキュメントを下記の表 9 に示す。

表 9：本研究でローカリゼーションを行ったドキュメントの種類

ドキュメントの種類	文字数	文章数	言語
商店街	23827	551	英, 中, 台, 葡,

			仏, 独, 伊, 西
大学	21984	1084	英, 中, 韓
寺院	6263	179	英, 中, 台, 韓, 仏, 独, 伊, 西
貸貸	16035	1048	英, 中, 韓
医療	23880	1063	英, 中, 韓, 葡

第4章 実験の設計

4.1 実験手法

第3章で示した通り，機械翻訳結果をモノリンガルが修正し，それをバイリンガルが修正するというプロセスでローカリゼーションを行うと，既存のバイリンガルのみがローカリゼーションを行う場合と比べて，ローカリゼーションにかかるトータルのコストが削減出来る可能性があることが先行研究で述べられている。

本研究では，日中以外の言語ペアについても先行研究と同様のプロセスでローカリゼーションを行うことでコストの削減が可能かどうかの確認を行うということと，翻訳にかかる時間及びコストは翻訳対象となるドキュメントの翻訳のしやすさや難解さに影響を受けると言う前提のもとに，より多くのドキュメントに対して先行研究と同様のプロセスでローカリゼーションを行うことで，より一般性の高いデータを得ることを目的に，先行研究と比較して大量，多種のドキュメントに対して，複数の言語へのローカリゼーションを，機械翻訳+モノリンガル修正+バイリンガル修正というプロセスで行った。

本研究における実験では，図6のようなワークフローで，ドキュメントのローカリゼーションを行った。

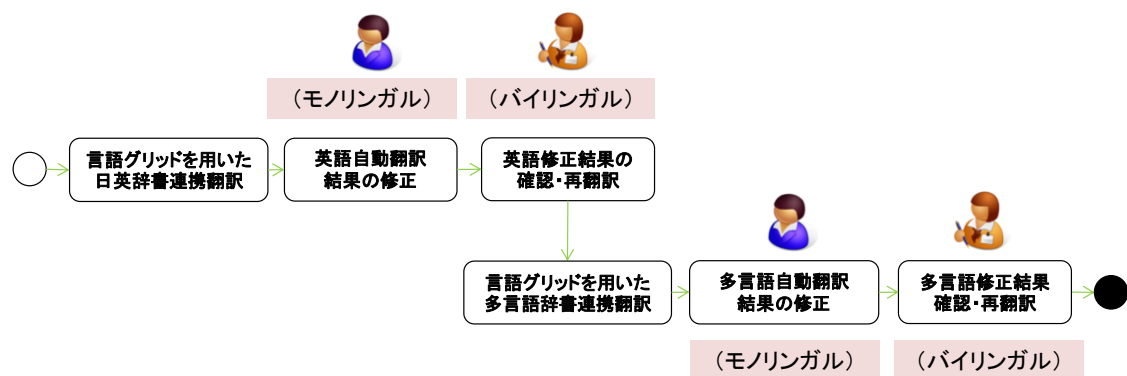


図6：実験で行ったローカリゼーションプロセス

各ドキュメントに対して，まず日本語から英語へのローカリゼーションを行い，その後英語からその他の言語へのローカリゼーションを行うというプロセスでローカリゼーションを行った。

英語をハブ言語として選択した理由は、英語と他言語間の言語資源が、英語以外の言語間同士の言語資源と比較して、量、質ともに優れており、辞書連携機械翻訳の精度がある程度期待出来るためである。また、英語と他の言語の二言語を使用出来るバイリンガルが、英語以外の二言語を操るバイリンガルと比べて数が多いという点も、ローカリゼーションにおいて一般的に英語がハブ言語として選択される理由の一つでもある。

辞書連携機械翻訳には、翻訳機は日本語から英語への翻訳には J-server を用い、英語からその他の言語への翻訳には Google Translate を用いた。J-server は日英翻訳において比較的精度の高い翻訳結果が得られるということが知られている。

また、各専門用語辞書は、こちらで予め単語の抽出及び多言語翻訳を行い多言語対訳辞書として作成しておいたものを用い、言語グリッドを用いて上記の機械翻訳機と連携させた。

モノリンガル作業者は、自分の母国語以外のドキュメントを見ずに、修正作業を行った。これは、作業者の母国語以外の言語（例えばドイツ語話者にとっての英語）の理解度の差によって作業時間や作業結果に差が生じるのを防ぐためである。また、モノリンガル作業者は、辞書連携によって翻訳された単語に関しては訳語の変更を行わないように制限した。これは、モノリンガル作業者が辞書登録されている単語の対訳を変更すると、辞書連携機械翻訳文と比較して翻訳の品質が低下する可能性があるためである。

機械翻訳+モノリンガル修正文の例を以下の図 7 に示す。

日本語	辞書連携機械翻訳文	モノリンガル修正文
この京都の優れた技術や美意識により育まれた着物は、この街のブランドイメージを込めて京呉服と呼ばれています。	This excellent technology in Kyoto and the kimono brought by a sense of beauty up include a brand image in this town, and are called Kyoto-style kimono .	The Kimono, which was developed thanks to the excellent technology and esthetic sense of Kyoto, and which is includes the brand image of this town, is called Kyoto-style kimono.

図 7：機械翻訳+モノリンガル修正の例

「京呉服（日本語）->Kyoto-style kimono（英語）」という対訳が専門用語辞書に含まれていたため、この部分の辞書連携翻訳結果を色つきで表示し、モノ

リンガルによる修正が行われないようにしている。

バイリンガル作業者は、辞書連携翻訳結果及びモノリンガルによる修正文を元に、対訳文の作成を行った。バイリンガル作業者は、他の言語のドキュメントも見ることが出来る状態で作業を行った。また、多言語対訳辞書も利用できるものとした。モノリンガル修正文をバイリンガルが修正する場合の例を以下の図 8 に示す。

日本語	辞書連携機械翻訳文	モノリンガル修正文	バイリンガル修正文
茶碗などの茶道具や茶室の床の間にかける掛け軸は各々が一級の美術品であり、茶事の時間全体を通して総合的な芸術とされます。	Each is an artwork of first kyuu and the hanging scroll which calls the alcove of tea ceremony utensils of a bowl and a tea-ceremony room is made overall art through whole time of tea ceremony .	All tea ceremony utensils are first-class artwork(the bowl, the scroll hanging in the alcove etc.). The tea ceremony room itself can be considered a piece of art.	Tea ceremony utensils such as tea bowls and hanging scrolls that are hung in the alcove of a tea ceremony room are all high-quality pieces of art, and these contribute to the overall art of the tea ceremony during the time over which a tea ceremony is held.

図 8 : モノリンガル修正文をバイリンガルが修正する場合の例

モノリンガルが辞書連携機械翻訳文を修正できない場合、モノリンガルは修正を行わず、バイリンガルが直接翻訳を行う。このような場合の例を以下の図 9 に示す。

日本語	辞書連携機械翻訳文	モノリンガル修正文	バイリンガル修正文
現在では優雅な伝統工芸品として、主に印籠紐や根付の紐など古美術関係、茶入や茶碗の仕覆の緒など、茶道具関係の紐に使われています。	String of medicine case string and netsuke (ornament on end of cord) is used for string of an antique relation and a tea ceremony utensils relation such as tea caddy and a cord of pouch for tea utensils of a bowl mainly as graceful traditional arts and crafts object present.		Now, as elegant traditional arts and crafts objects, with antiques they are mainly used as medicine case strings and netsuke (ornament on end of cord) strings, and with tea ceremony utensils they are mainly used as cords for tea caddies or tea utensil pouches.

図 9：モノリンガルが修正を行わずバイリンガルが翻訳を行う場合の例

4.2 評価基準

実験では、モノリンガル作業時間、モノリンガルの作業結果、バイリンガル作業時間、バイリンガルの作業結果、機械翻訳の品質、モノリンガル修正文の品質をデータとして得ることが出来た。機械翻訳文の品質及びモノリンガル修正文の品質の測定方法については、第 5 章にて後述する。また、モノリンガルの作業結果から、モノリンガル修正率を求めた。モノリンガル修正率とは、翻訳対象ドキュメント全体に対する、辞書連携機械翻訳後の文章をモノリンガル作業者が修正出来る文の割合である。

また、モノリンガル作業時間とバイリンガル作業時間の合計を合計作業時間として求めた。これを言語グリッド利用時の作業時間と定義した。

既存のプロセスでローカリゼーションを行った場合のバイリンガルの作業時間に関しては、既知のデータから計算して求めた。これを、通常翻訳時の作業時間と定義した。

バイリンガル作業時間と通常翻訳時の作業時間から、機械翻訳+モノリンガル修正+バイリンガル修正でのプロセスでのローカリゼーションの場合のバイリンガル作業時間削減率を求めた。

また、言語グリッド利用時の作業時間と通常翻訳時の作業時間から、合計作業時間削減率を求めた。

モノリンガル作業者及びバイリンガル作業者の時間単価は定まっているので、それとモノリンガル作業時間、バイリンガル作業時間から、言語グリッド

利用時のコストを計算した。

また、通常翻訳時のコストも同様に計算した。

言語グリッド利用時のコストと、通常翻訳時のコストから、機械翻訳+モノリンガル修正+バイリンガル修正でのプロセスでローカリゼーションを行った場合の、バイリンガルのみが行うプロセスと比較してのコスト削減率を求めた。

それぞれの指標がどのような関係にあるか、またそこから、どのような条件を満たせばコストの削減が可能になるかについての分析を第5章で行う。

第5章 評価

本章では、実験によって得られたデータの評価と分析を行う。5.1 節では、モノリンガル修正率と作業時間及びコストの関係について考察していく。5.3 節以降では、機械翻訳文及びモノリンガル修正文の品質も評価の対象とした。そのための翻訳文の品質評価手法についての説明を 5.2 節で行う。5.3 節では、機械翻訳文の品質と、作業時間及びコストの関係に焦点を置いて評価を行う。5.4 節では、モノリンガル修正文の品質と、作業時間及びコストの関係を中心に評価を行う。そして、5.5 節でこれらの評価のまとめと考察を行う。

5.1 モノリンガル修正率と翻訳コストの関係

モノリンガル修正率と作業時間及びコスト削減率の関係は表 10 のようになった。

表 10：モノリンガル修正率と作業時間及びコスト削減率

言語ペア	コンテンツ	モノリンガル修正率	モノリンガル作業時間 (1 ページあたり)	バイリンガル作業時間 (1 ページあたり)	コスト削減率
日本語→英語	ドキュメント A	47.77%	21.43	102.86	30.03%
	ドキュメント B	52.79%	18.91	129.09	-8.59%
	ドキュメント C	45.59%	19.44	213.33	-35.36%
英語→中国語	ドキュメント A	46.04%	23.6	128.57	-19.23%
	ドキュメント B	71.79%	17.03	101.82	-36.69%
	ドキュメント C	40.08%	14.44	120	4.53%
英語→韓国語	ドキュメント A	94.01%	25.71	16.29	76.29%
	ドキュメント B	98.70%	22.5	56.25	19.50%
英語→ドイツ語	ドキュメント A	100%	27.6	53.14	43.96%
	ドキュメント C	75.28%	22.78	60	46.37%
英語→中国語(繁体字)	ドキュメント A	78.89%	35.14	137.14	-30.96%
英語→ポルトガル語	ドキュメント A	99.63%	34.86	53.14	41.21%
英語→フランス語	ドキュメント C	95.01%	19.44	60	47.64%
英語→ス	ドキュメント C	90.82%	26.67	60	44.90%

ペイン語					
------	--	--	--	--	--

ドキュメント A は、京都の商店街のコンテンツ，ドキュメント B は、大学の事務関係のコンテンツ，ドキュメント C は、寺院関係のコンテンツである。

このデータから、各要素の関係について分析を行う。

モノリンガル修正率と、1 ページあたりのモノリンガル作業時間の関係を図 10 図 10 に示す。

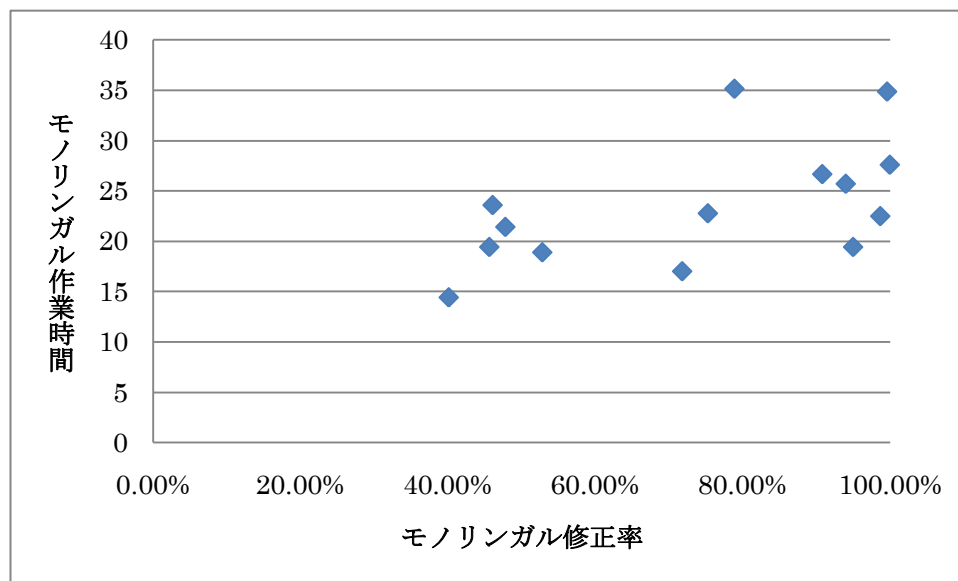


図 10 : モノリンガル修正率とモノリンガル作業時間の関係

モノリンガル修正率とモノリンガル作業時間の相関係数は 0.5612 となり、これらの間には中程度の相関がみられる。モノリンガル修正率が高いほど、モノリンガル作業者の作業対象となるドキュメントの量は増えるため、モノリンガル修正率とモノリンガル作業時間に正の相関があるというのは妥当な結果であると言える。

次に、モノリンガル修正率と、1 ページあたりのバイリンガル作業時間の関係を図 11 に示す。

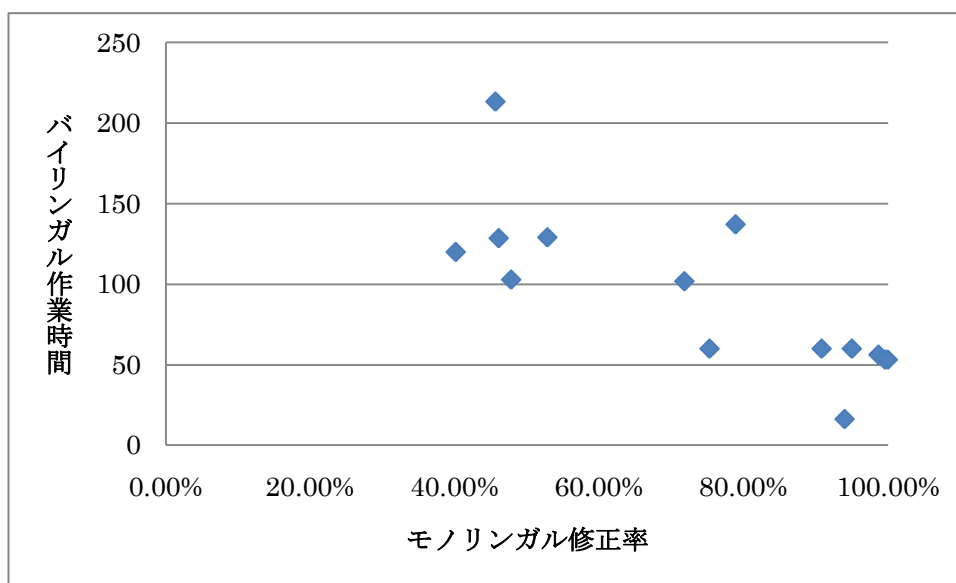


図 11 : モノリンガル修正率とバイリンガル作業時間の関係

モノリンガル修正率とバイリンガル作業時間の相関係数は -0.7807 となり，これらには強い負の相関があると言える．モノリンガル修正率が高いと，バイリンガルの作業時間を大幅に減らすことが可能になると言える．モノリンガルが修正を行うことでバイリンガルの作業量を減らすことが出来るという結果は，先行研究の実験で得られた結果と同様の結果を示すものであり，先行研究を裏付ける結果になったと言える．

次に，モノリンガル修正率を 40%-60%と 60%-80%、80%-100%の三つの範囲で分け，それぞれのモノリンガル作業時間及びバイリンガル作業時間をグラフにしたものを，図 12 に示す．

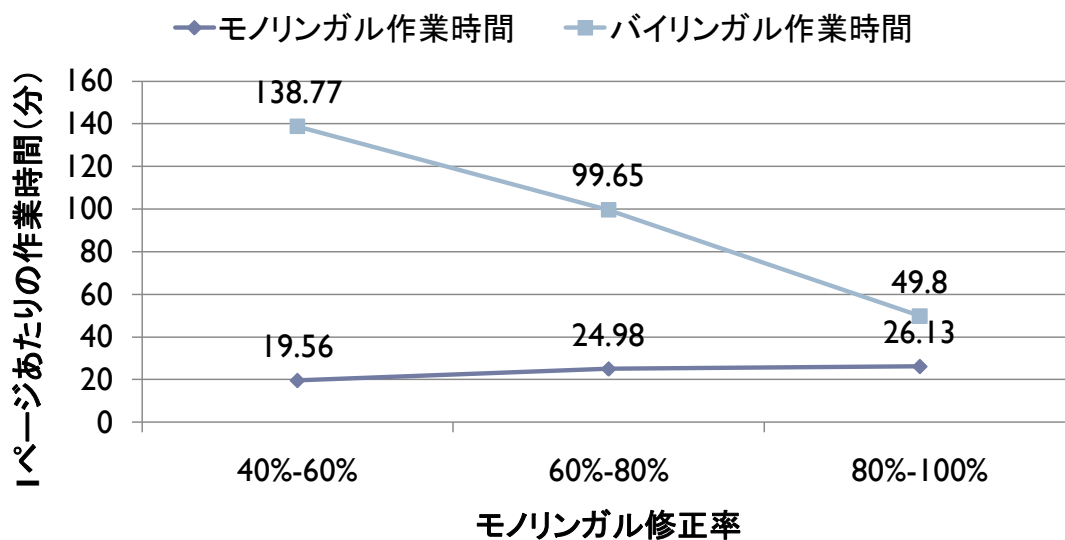


図 12:モノリンガル修正率とモノリンガル作業時間及びバイリンガル作業時間

このグラフからは、モノリンガル修正率が上昇した場合、モノリンガル作業時間の増加と比較して、バイリンガル作業時間の削減が大きいという結果が読み取れる。また、モノリンガルの時間単価はバイリンガルの時間単価に比較して小さいということが知られているため、モノリンガル作業時間の増加に比べてバイリンガル作業時間の削減が大きくなるということは、トータルコストの削減が期待できるということになる。

次に、モノリンガル修正率とコスト削減率の関係をグラフにしたものを図 13 に示す。

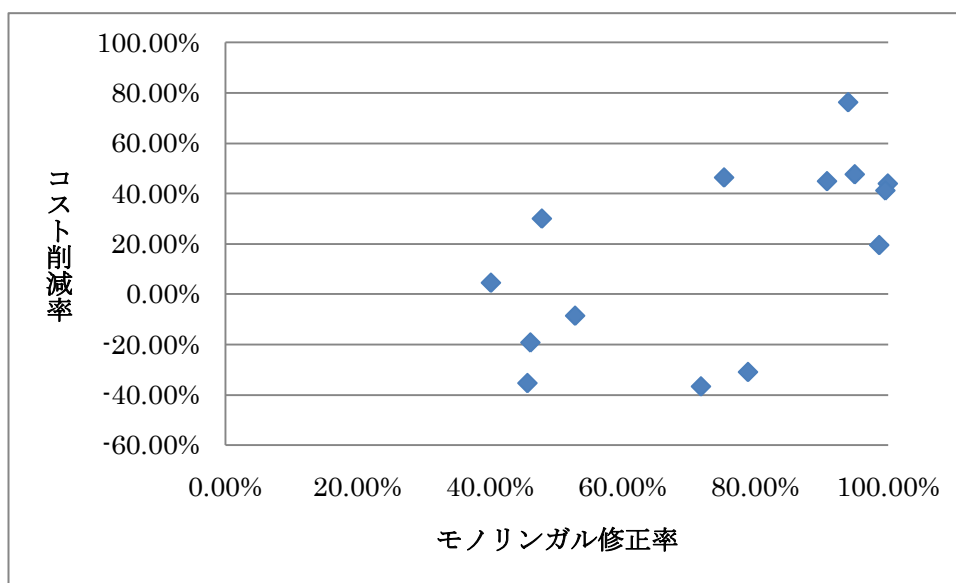


図 13：モノリンガル修正率とコスト削減率の関係

モノリンガル修正率とコスト削減率の相関係数は **0.6070** となり、中程度の相関があると言える。

5.2 翻訳の品質の評価

5.1 節では、モノリンガル修正率とコスト削減率の関係について調べたが、更に機械翻訳の品質、及びモノリンガル修正文の品質とコスト削減率の関係についてを以下で考察する。

機械翻訳文及びモノリンガル修正文の品質を求めるにおいて、BLEU を用いた。本節では、BLEU の仕組みについて説明する。

BLEU は、翻訳文の品質を自動的に評価するツールである。翻訳サービスの品質を比較するために、人手で大量の翻訳結果を評価するのはコストが大きい。ため、翻訳結果自動評価手法を使って評価を行うという発想は自然なものである。

BLEU の基本的なアイデアは翻訳結果と正解文の **n-gram** の一致度を見ることである。図 14 に **1-gram** による評価値の計算の例を挙げる。

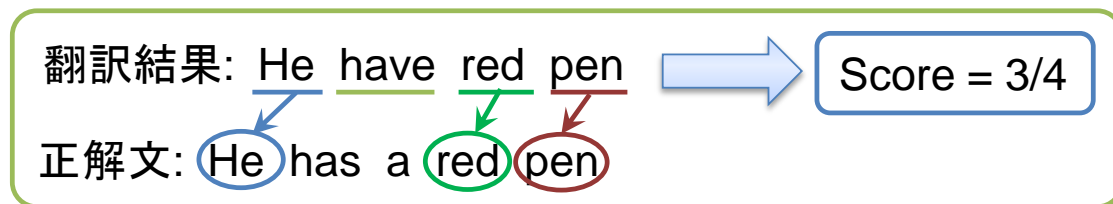


図 14 : BLEU を用いた 1-gram による評価値の計算の例

ここでは簡単のために正解文を 1 文だけ用いている。図 14 の翻訳結果から 4 つの 1-gram (" He ", " have ", " red ", " pen ") を得ることができる。その内、正解文に出現する 1-gram は " He ", " red ", " pen " の 3 つなのである。翻訳結果が含む 4 つの 1-gram のうち、3 つが正解文にも含まれるため、評価値は 3/4 となる。もう 1 つの例として、2-gram による評価値の計算を図 15 示す。

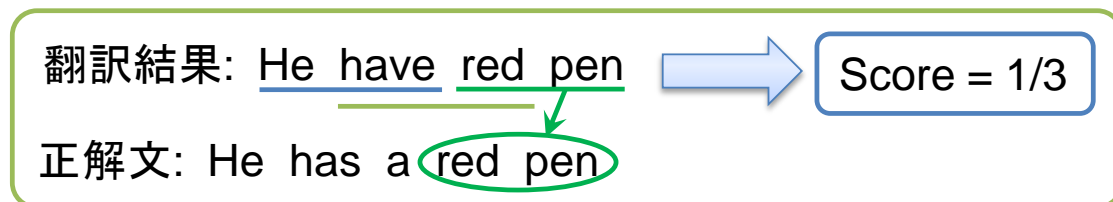


図 15 : BLEU を用いた 2-gram による評価値の計算の例

図 15 の翻訳結果から得ることができる 2-gram は (" He have ", " have red ", " red pen ") の 3 つである。その内、正解文に出現するのは " red pen " の 1 つだけである。翻訳結果から得られた 3 つの 2-gram のうち 1 つが正解文中にも出現するため評価値は 3/4 となる。BLEU では n-gram 以下の評価値を計算して、評価値の平均値を計算する。今回の例では、2-gram の BLEU スコアは 1/3 と 3/4 の平均である 13/24 である。最後に、翻訳結果が短すぎることで評価値が大きくなることを防ぐために Brief Penalty を評価値に乗算して最終的な評価値を得る。翻訳結果が短いことにより評価値が大きくなる理由は、翻訳結果が " He " であった場合を考えればよい。

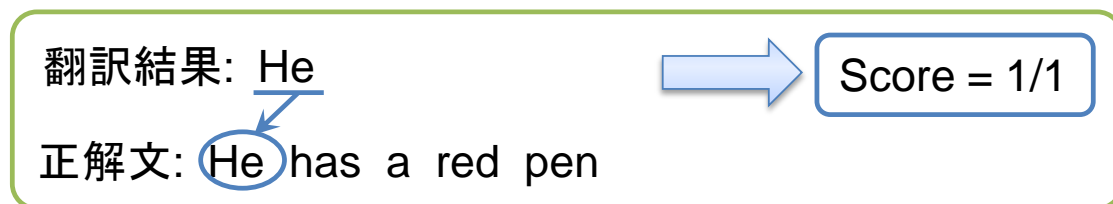


図 16: 翻訳結果が短い場合に評価値が大きくなる例

図 16 のように翻訳結果が” He ” だけのとき，1-gram による評価値が最高値である 1 になることが分かる．しかし，明らかに翻訳結果の品質は悪く，正しく評価できているとは言えない．このような短い翻訳結果の評価値を是正するために，Brief Penalty を評価値に乗算する．翻訳結果の長さが正解文の長さより短くなるほどペナルティは大きくなる．つまり Brief Penalty は 0 に近づく．最後に，BLEU の評価手法を以下の図 17 に示す．

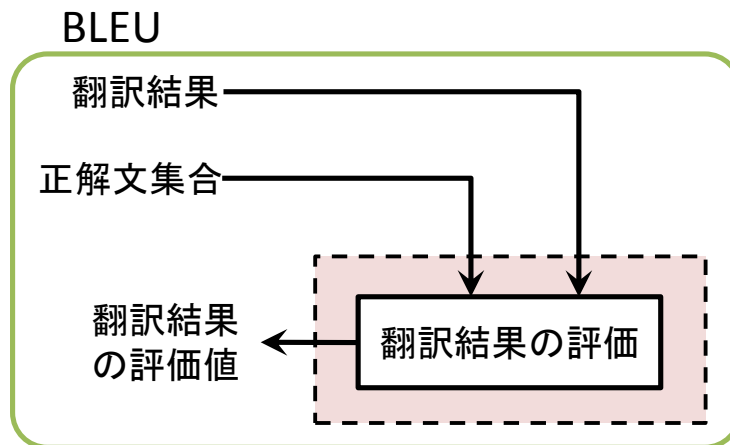


図 17 : BLEU の評価手法

本実験において，翻訳結果として，機械翻訳結果及びモノリンガル修正文を入力し，正解文集合として，バイリンガル修正文を入力することで，評価値を得た．

5.3 機械翻訳の品質と翻訳コストの関係

本節では，機械翻訳の品質と各要素の関係についての考察を中心に行っていく．

機械翻訳の品質とコスト削減率は下記の表 11 ようになった．

表 11 : 機械翻訳の品質とコスト削減率

言語ペア	コンテンツ	機械翻訳の品質	コスト削減率
日本語→英語	ドキュメント A	0.261	30.03%
	ドキュメント B	0.1142	-8.59%
	ドキュメント C	0.1558	-35.36%
英語→中国語	ドキュメント A	0.2281	-19.23%
	ドキュメント B	0.2254	-36.69%

	ドキュメント C	0.2163	4.53%
英語→韓国語	ドキュメント A	0.1487	76.29%
	ドキュメント B	0.0678	19.50%
英語→ドイツ語	ドキュメント A	0.057	43.96%
	ドキュメント C	0.2313	46.37%
英語→中国語（繁体字）	ドキュメント A	0.18	-30.96%
英語→ポルトガル語	ドキュメント A	0.3063	41.21%
英語→フランス語	ドキュメント C	0.5824	47.64%
英語→スペイン語	ドキュメント C	0.4311	44.90%

また、機械翻訳の品質と、コスト削減率の関係をグラフにすると、下記の図 18 のようになる。

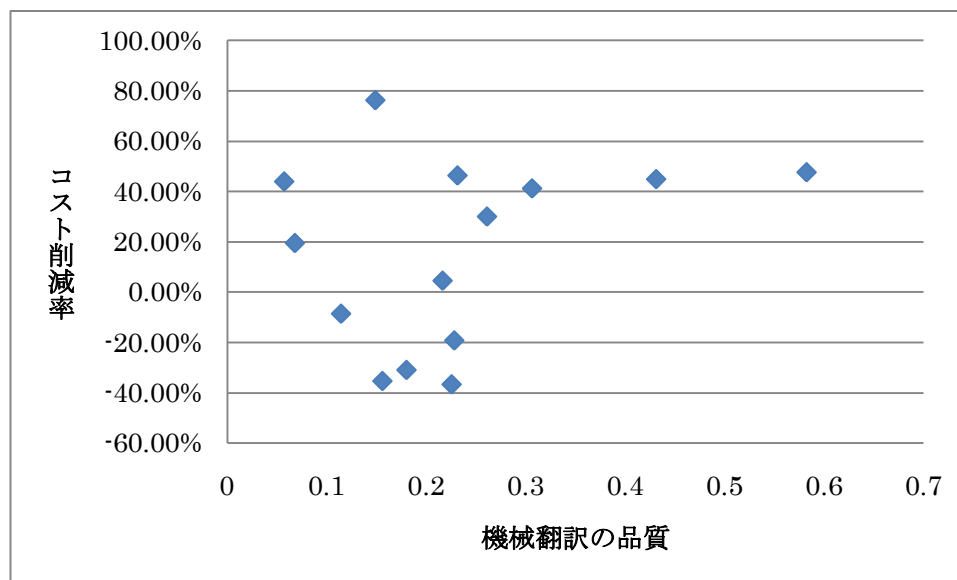


図 18：機械翻訳の品質とコスト削減率の関係

機械翻訳の品質とコスト削減率の相関係数は、0.2759 となり、これらの間には弱い相関があると言える。機械翻訳の品質が 0.3 以上のケースであれば、コストの削減には成功しているが、0.3 以下の場合では、逆にコスト削減率がマイナス、すなわちコストが多くかかっているケースも見受けられる。

機械翻訳の品質と、コスト削減率に正の相関があるというのは、妥当な結果であるといえるが、その相関は比較的弱いという結果が得られた。この点については、5.5 節でも考察を行いたい。

5.4 モノリンガル修正文の品質と翻訳コストの関係

各ドキュメントについて BLEU を用いて求めたモノリンガル修正文の評価値の平均とモノリンガル作業時間、及びコスト削減率の関係は以下の表 12 のようになる。

表 12：モノリンガル修正文の品質とモノリンガル作業時間及びコスト削減率

言語ペア	コンテンツ	モノリンガル修正文の品質	モノリンガル作業時間（1ページあたり）	コスト削減率
日本語→英語	ドキュメント A	0.3407	21.43	30.03%
	ドキュメント B	0.486	18.91	-8.59%
	ドキュメント C	0.1492	19.44	-35.36%
英語→中国語	ドキュメント A	0.2529	23.6	-19.23%
	ドキュメント B	0.3328	17.03	-36.69%
	ドキュメント C	0.3556	14.44	4.53%
英語→韓国語	ドキュメント A	0.4297	25.71	76.29%
	ドキュメント B	0.4241	22.5	19.50%
英語→ドイツ語	ドキュメント A	0.5901	27.6	43.96%
	ドキュメント C	0.587	22.78	46.37%
英語→中国語（繁体字）	ドキュメント A	0.268	35.14	-30.96%
英語→ポルトガル語	ドキュメント A	0.3747	34.86	41.21%
英語→フランス語	ドキュメント C	0.8345	19.44	47.64%
英語→スペイン語	ドキュメント C	0.6135	26.67	44.90%

モノリンガル修正文の品質とモノリンガル作業時間の関係をグラフにすると、下記の図 19 のようになる。

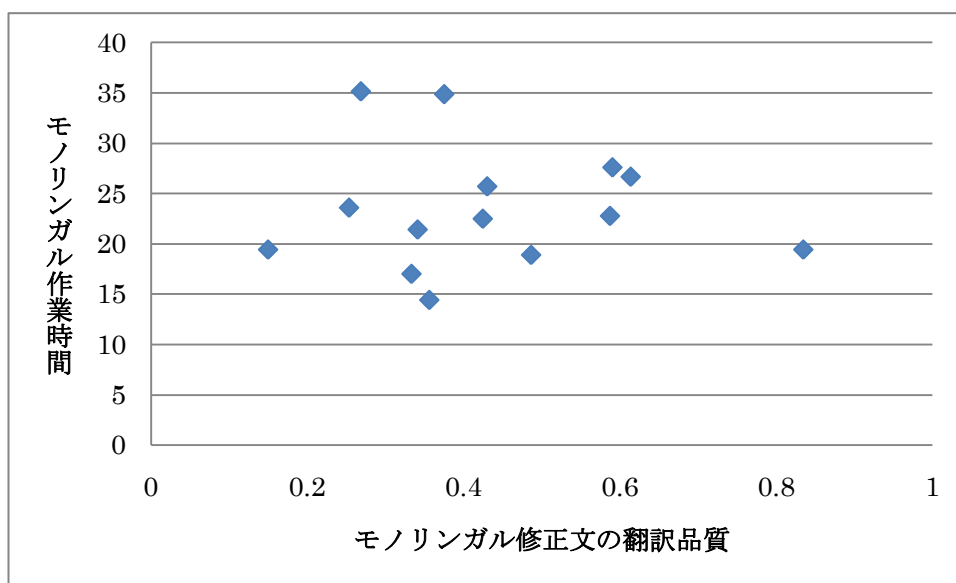


図 19：モノリンガル作業時間とモノリンガル修正文の翻訳品質の関係

モノリンガル作業時間とモノリンガル修正文の翻訳品質の相関係数は、**-0.0478** となり、相関は無いと言える。すなわち、モノリンガルが長時間修正作業を行ったからといって、モノリンガル修正文の品質は向上することはないという結果が得られた。

また、モノリンガル修正文の品質とコスト削減率の関係をグラフにすると、以下の図 20 のようになる。

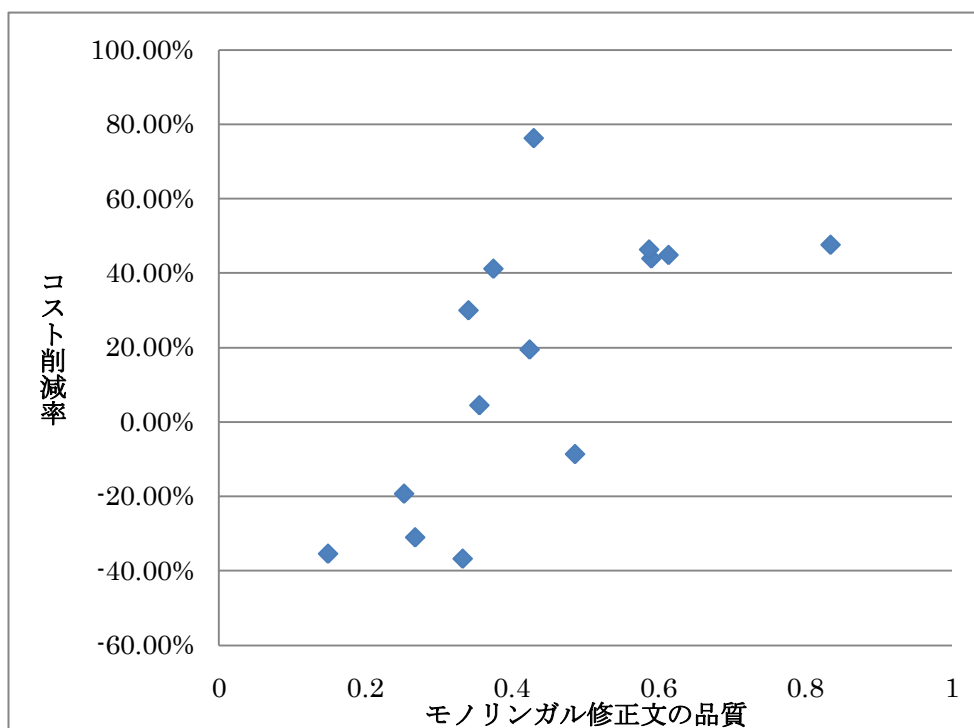


図 20 : モノリンガル修正文の品質とコスト削減率の関係

モノリンガル修正文の品質とコスト削減率の相関係数は 0.6742 となり、これらの間には中程度の相関があると言える。

モノリンガル修正文の品質とコスト削減率に正の相関があるのは、妥当な結果と言える。また、機械翻訳の品質とコスト削減率の相関と比較して、モノリンガル修正文の品質とコスト削減率の相関が強い点については注目すべき点であると考えられる。このあたりも含めた考察を 5.5 節で行う。

更に、モノリンガル修正文の品質を、0-0.3 (低), 0.3-0.5 (中), 0.5 以上 (高) の三つに分けてコスト削減率を求めた。それをグラフ化したものが、下記の図 21 である。

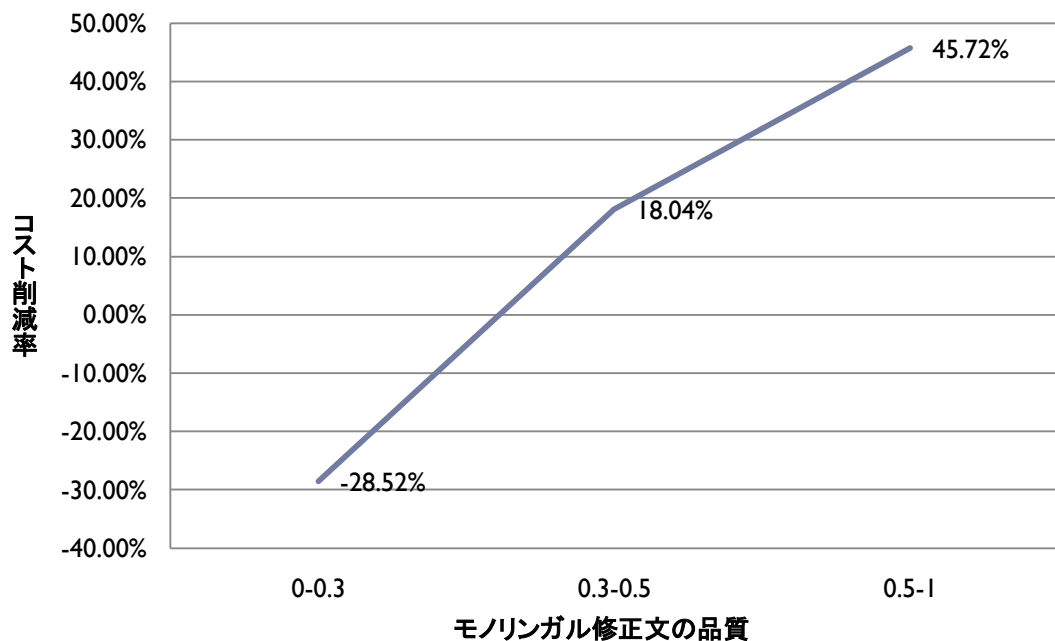


図 21：モノリンガル修正文の品質とコスト削減率

上記のように、モノリンガル修正文の品質とコスト削減率には明確な正の相関があり、モノリンガル修正文の品質を向上させるとコストの削減が可能となるという結果が言える。

5.5 ローカリゼーションの効率化のための条件

以上の結果から、それぞれの要素の相関をまとめると、下記の表 13 のようになる。

表 13：各要素の相関関係のまとめ

	相関係数
モノリンガル修正文の品質とコスト削減率	0.6742
機械翻訳の品質とコスト削減率	0.2759
モノリンガル修正率とコスト削減率	0.607
モノリンガル修正率とバイリンガル作業時間	-0.7807
モノリンガル修正率とモノリンガル作業時間	0.5612

これらの結果から言えることを以下にまとめる。

まず、モノリンガル修正率が高いと、モノリンガル作業時間は増加し、バイリンガル作業時間は減少する。モノリンガル作業時間の増加分と比べて、バイリンガル作業時間の減少分が大きい。また、モノリンガル作業者の時間単価は、バイリンガル作業者の時間単価よりも安い。これらから、モノリンガル修正率が高くなると、機械翻訳+モノリンガル修正+バイリンガル修正でローカリゼーションを行う場合のトータルでのコストは、バイリンガルのみでローカリゼーションを行う場合に比較して、小さくなる傾向にある。

また、機械翻訳の品質とコスト削減率には正の相関がある。これは妥当な結果であると言えるが、その相関は比較的小さい。

モノリンガル修正文の品質とコスト削減率には正の相関がある。これも妥当な結果であると言える。また、モノリンガル修正文の品質とコストの相関は、機械翻訳文とコストの相関と比べて大きく、よりコストの削減に関係するのは、モノリンガル修正文の品質であるという結果が得られた。

これらの分析結果から、以下で考察を行う。

機械翻訳文の品質に比べ、モノリンガル修正文の品質及びモノリンガル修正率がコストの削減に大きな影響を与えるという結果が得られたが、これは、ローカリゼーションのプロセスにおいて、機械翻訳文、モノリンガル修正文、バイリンガル修正文という順で作業が行われ、作業プロセスでより後ろにあるモノリンガル修正文の出来がバイリンガル作業者の作業時間に強く影響を及ぼすためであると考えられる。

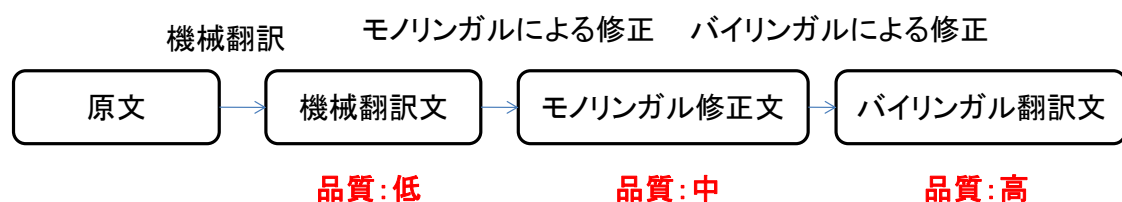


図 22 : 機械翻訳とモノリンガル修正とバイリンガル修正による作業のステップ

機械翻訳よりもモノリンガルによる修正が後にあるため、例え機械翻訳文の品質が低くても、モノリンガル作業者が修正を行うことによって、その品質はある程度改善され、バイリンガル作業者の作業量の軽減を行うことが可能となっ

ているものと考えられる。

機械翻訳の品質よりもむしろモノリンガル修正文の品質のほうがコストの削減に大きく影響するという結果となった。モノリンガル修正文の品質は、モノリンガル作業者の品質に影響を受けるので、サービス連携を用いたプロセスでローカリゼーションを行う場合、能力の高いモノリンガル作業者を雇うことが重要であるという予測が可能である。

また、機械翻訳の品質は低い場合でも、モノリンガル作業者の働きによって、コストの削減が可能である可能性があると言えるので、言語資源が少なく機械翻訳の品質もあまりよくない言語ペアに関してもこのプロセスを用いることでローカリゼーションコストを抑えることが可能であるという可能性があることを示した。

第6章 翻訳支援ツール

6.1 ローカリゼーションワークフロー

機械翻訳+モノリンガル修正+バイリンガル修正というプロセスでローカリゼーションを行えば、ローカリゼーションにかかるコストを削減出来る可能性があることを第5章までで示してきた。しかし、本実験でこのプロセスでローカリゼーションを行ったところ、いくつかの問題が発生した。実際のローカリゼーションの現場でこのプロセスでのローカリゼーションを行う場合でも、同様の問題が発生する可能性が高い。そのため、本章では、その問題の解決を行うためのツールの検討と設計を行う。

実際のローカリゼーションでは、翻訳作業以外の部分に、タスクの割り振り、モノリンガル作業者とバイリンガル作業者でのドキュメントの共有等の作業が発生する。実際のローカリゼーションのプロセスは図23のようになる。

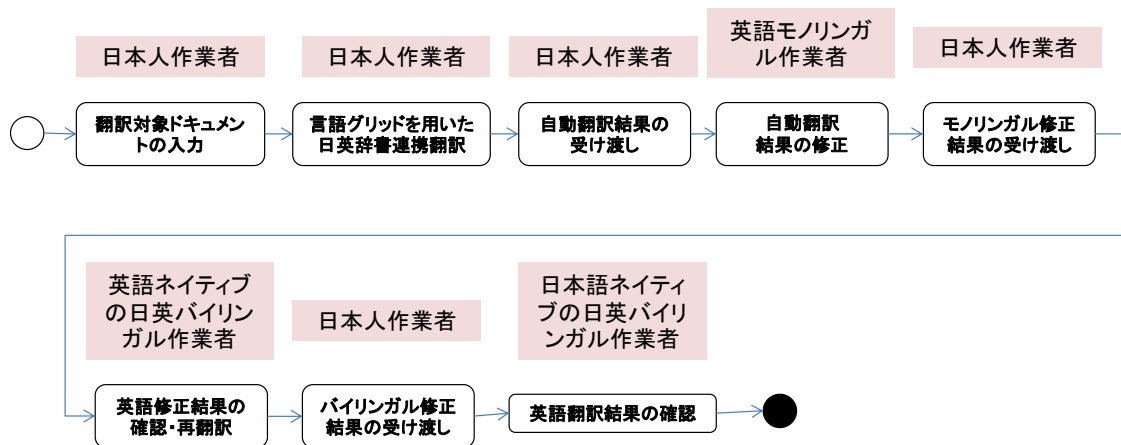


図 23 : 実際のローカリゼーションのプロセス

機械翻訳+モノリンガル修正+バイリンガル修正のローカリゼーションプロセスでは、既存のバイリンガルのみでのローカリゼーションプロセスと比較し、モノリンガル作業者をプロセスに組み込んでいる分、タスクの割り振りや作業結果の受け渡し等の作業に必要なコストが増大するため、バイリンガルのみでのプロセスと比較してコスト及び時間が多く必要となる可能性があると考えられる。このコストを減らすために、タスクを Web 上で管理するツールの設計を行った。Web 上でドキュメントを管理することにより、翻訳文の受け渡しが必要

なくなり、ローカリゼーションのプロセスは図 24 のように改善される。

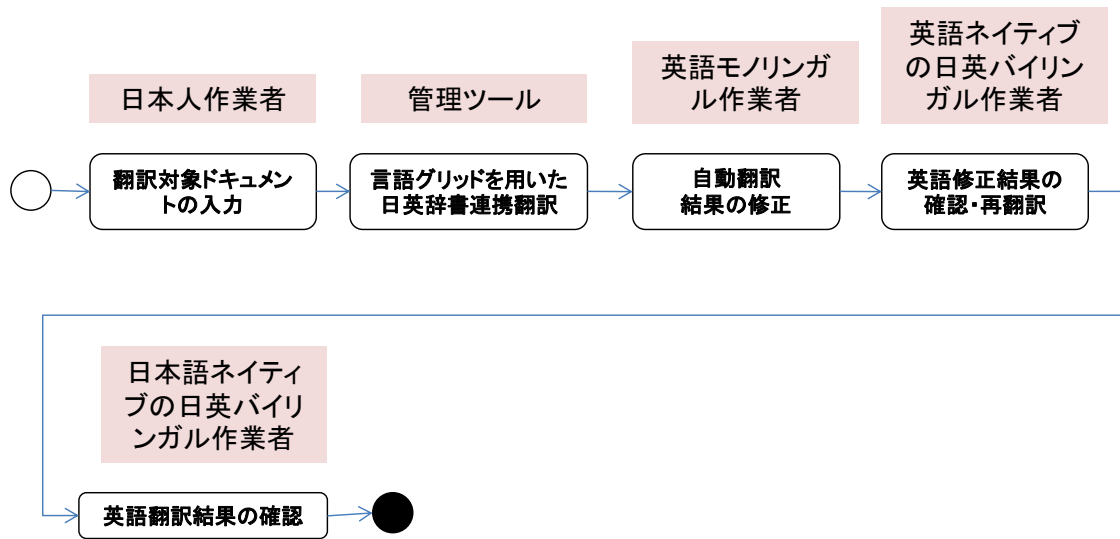


図 24 : ローカリゼーション支援ツールを用いた場合のローカリゼーションプロセス

以下で実際に設計したローカリゼーション支援ツールについての詳細な説明を行う。

6.2 協調翻訳

上記のローカリゼーションを支援するためのツールとして、Language Grid Toolbox¹上で協調翻訳機能の設計を行った。

Language Grid Toolbox とは、情報通信研究機構 (NICT) が開発した、言語グリッドを用いた多言語コミュニケーション支援ツールである。Toolbox の特徴として、オープンソースの CMS である XOOPS Cube をベースに開発されているという点があり、XOOPS のモジュールを開発することにより、用途に合わせて Toolbox の機能を拡張することが可能である。機械翻訳+モノリンガル修正+バイリンガル修正でのローカリゼーションプロセスでは、言語グリッド辞書連携翻訳を用いることを前提としていたため、言語グリッドの言語資源を用いたサービスを構築出来る Language Grid Toolbox 上でローカリゼーション支援ツールの設計を行った。

ローカリゼーション支援ツールを設計するにおいて、必要とされる要件は以

¹ <http://langrid-tool.nict.go.jp/toolbox/>

下のようになった。

- 作業対象ドキュメントの Web 上での共有
- 機械翻訳前，機械翻訳済モノリンガル修正前，機械翻訳済モノリンガル修正不可，モノリンガル修正後バイリンガル修正前，バイリンガル修正後といった作業状態の明示
- 言語の選択
- 翻訳対象ドキュメントの一括での辞書連携機械翻訳
- 翻訳対象ドキュメントの文ごとの辞書連携機械翻訳
- 複数の機械翻訳機の翻訳結果の比較
- 翻訳対象ドキュメントの自動での文ごとの分割
- 折り返し翻訳結果の表示
- 専門用語対訳辞書の参照
- 専門用語対訳辞書への単語，対訳の追加
- 用例対訳集の参照
- 用例対訳集への用例，対訳の追加
- 修正履歴の閲覧

これらの要件を満たすツールのデザイン及び設計を Language Grid Toolbox 上で実装出来るように行った

協調翻訳機能の作業画面は図 25 のようになる。



図 25：協調翻訳機能の作業画面

協調翻訳機能を用いた翻訳作業では、まず、作業管理者が翻訳対象となるドキュメントの入力を行う。そして、翻訳元言語と翻訳先言語の設定を行い、辞書連携機械翻訳を行う。ここでは更に、用いる機械翻訳機の選択や専門用語対訳辞書、用例対訳辞書の選択なども拡張として行うことが出来る。また、折り返し翻訳結果の参照を可能としているので、これを用いることで最適と思われる機械翻訳機を選択も可能となっている。

折り返し翻訳とは、ある言語から他の言語に機械翻訳を行い、その機械翻訳結果に対しさらに翻訳先言語から翻訳元言語への機械翻訳を再度行い、元の言語に翻訳し直す翻訳である。こうして元の言語に再翻訳された結果が、原文と類似した文章となっていれば、翻訳先言語の翻訳結果もある程度信頼出来るものとなっていると考えられる。このような機能を用いることで、翻訳先の言語が理解出来ない場合でも、ある程度機械翻訳文の品質の確認を行うことが可能となっている。

次に、モノリンガル作業者は、辞書連携機械翻訳が行われたドキュメントに対し、修正作業を行う。この際、作業対象ドキュメント、及び作業状態を Web 上に保存することが出来、他の作業者と作業内容を共有することが可能となっ

ている。このため、バイリンガル作業者は、モノリンガル作業者が作業を終えると即座に修正作業に入ることが可能となっている。

次に、バイリンガル作業者は、モノリンガル修正が行われたドキュメントを対象に修正作業を行う。この際、モノリンガル修正文だけでなく、機械翻訳文や専門用語対訳辞書、用例対訳辞書等も参照しつつ訳文の作成を行うことが可能である。

以上のように、翻訳対象となるドキュメントとその状態を Web 上で管理、保存することで、作業者は、自分の担当箇所を適宜確認、修正し、保存することが出来るようになった。複数の作業者が共同で作業を行うにおいて、Web 上でドキュメント及び進捗を管理することによって、ドキュメントの受け渡し等に必要な時間及びコストを削減することが可能になると思われる。

第7章 おわりに

本論文では、機械翻訳及びモノリンガル作業者を組み込んだローカリゼーションプロセスについて考察を行ってきた。

先行研究から、機械翻訳文をモノリンガル作業者が修正し、それをさらにバイリンガル作業者が修正するというプロセスでローカリゼーションを行うとコストが削減出来る可能性があるということが知られており、さらに本研究では、コスト削減が可能となるための条件を調べるために、先行研究と比較して多様な種類の、そして大量のドキュメントを対象に、日英、英中等の言語ペアで、機械翻訳及びモノリンガル作業者を組み込んだプロセスでのローカリゼーション実験を行い、データの収集を行った。得られたデータから、モノリンガル修正率、機械翻訳文の品質、モノリンガル修正文の品質、コスト削減率といった指標にどのような相関があるのかの分析を行った。そこから、コスト削減率と強い相関があるのは、モノリンガル修正率、及びモノリンガル修正文の品質であるという結果を得た。また、機械翻訳の品質とコスト削減率の相関は小さいという結果も得られた。機械翻訳の品質があまり良くない言語ペアに関しても先述のプロセスでのローカリゼーションは用いることが出来る可能性があるということを示した。

さらに、先述のプロセスでローカリゼーションを行う際に、作業の待ち時間やデータの受け渡しの時間等が発生し、バイリンガルのみでのローカリゼーションと比較して時間がかかってしまう可能性があるという問題の指摘を行った。そして、その問題を解決するために Web 上で翻訳作業の経過を共有するシステムの設計を行い、ローカリゼーションプロセスの効率化に貢献した。

今後の検討課題として、今回はモノリンガル作業者及びバイリンガル作業者の時間単価を日本におけるローカリゼーションの相場で計算したが、他国でのローカリゼーションを想定した場合、各作業者の時間単価は異なってくるものと考えられるので、その点も考慮に入れた考察が必要となる。

また、今回は機械翻訳文及びモノリンガル修正文の品質を評価するにおいて、翻訳文の品質を自動で評価するツールである BLEU を用いたが、バイリンガル作業者による品質評価を行うことで、BLEU より精密な翻訳文の品質を調査することが可能である。

さらに、今回設計を行ったローカリゼーション支援ツールを用いた場合に、

支援ツールを用いなかった場合と比較し、どの程度作業を効率化出来ているかの検討も考察の余地がある。

今後、用例対訳ベース翻訳等の発展により、さらに効率的なローカリゼーションが可能になることが期待される。

謝辞

本研究における実験データ収集の際にご協力いただいた情報通信研究機構の皆様、石田大成社の皆様、ナビックスの皆様に深く感謝いたします。

また、研究を進める中で、アドバイザーとして様々なお助言を頂きました、荒井修亮教授、馬強准教授、研究の方針についてご指導をいただいた **Donghui Lin** 氏にも感謝の意を表します。

そして、研究中、活発に議論に参加していただき、多くのご意見を頂きました、松原繁夫准教授、**David KINNY** 准教授、稲葉利江子特定講師、服部宏充助教、中島悠特定助教、石田・松原研究室の諸先輩方にも御礼申し上げます。

最後に、本論文を執筆するにあたり、一貫してご指導いただいた石田亨教授に心よりの感謝を申し上げます。

参考文献

- [1] Yamashita, N. and Ishida, T.: Effects of Machine Translation on Collaborative Work, In Proceedings of International Conference on Computer Supported Cooperative Work. CSCW-06, pp 515-523(2006).
- [2] Donghui, L. Murakami, Y. Ishida, T and Tanaka, M.: Composing Human and Machine Translation Services: Language Grid for Improving Localization Processes, 7th International Conference on Language Resources and Evaluation (LREC 2010), pp. 500-506, (2010).
- [3] Donghui, L. Murakami, Y. Ishida, T. and Tanaka, M.: Lessons Learned from Composing Web Services and Human Activities, 7th International Joint Conference on Service Oriented Computing (ICSOC2009), Industry Track, November, (2009).
- [4] Morita, D. and Ishida, T.: Collaborative Translation by Monolinguals with Machine Translators, Proceedings of ACM Conference on Intelligent User Interface (IUI'09), pp. 361-365,(2009).

付録：翻訳文の実例

本実験で実際に翻訳された文章の例を以下に示す.

A.1 商店街ドキュメント

Japanese	English	Chinese (Simplified)	Chinese (Traditional)	Korean	Portuguese	German
1. 京呉服	1. Kyoto-style gofuku (kimonos)	1. 京都式和服	1. 京都和服	1. 교토식 기모노	1. Gofuku ao estilo de Kyoto (Kimonos)	1. Gofuku nach Kyoto-Art (Kimonos)
アジア大陸の東に位置する日本は、古くから大陸文化の影響を強く受けてきました。	Japan, located to the east of the Asian mainland, has long been heavily influenced by Asian mainland culture.	日本, 位于亚洲大陆东边, 很久以前就受到了大陆文化的深远影响。	日本位於亞洲大陸東部, 長期深受亞洲大陸文化影響。	아시아 대륙의 동쪽에 위치한 일본은 오랫동안 대륙 문화의 영향을 많이 받아 왔습니다.	O Japão, situado a leste do continente asiático, há muito que é fortemente influenciado pela cultura do continente asiático.	Japan, östlich des asiatischen Festlandes gelegen, wurde lange Zeit stark von der Kultur des asiatischen Festlandes beeinflusst.
日本の伝統的な民族衣装である着物も、その例外ではありません。	The kimono, a traditional Japanese garment, is no exception.	作为日本传统服装的和服, 也不例外。	作為傳統日本服裝的和服也不例外。	일본의 전통적인 민족 의상인 기모노도 예외가 아닙니다.	O kimono, uma peça de vestuário tradicional japonesa, não é exceção.	Der Kimono, ein traditionelles japanisches Kleidungsstück, bildet da keine Ausnahme.
着物を意味する呉服と言う名称は、中国の古い王朝名に由来します。	The term gofuku, which means kimono, derives from an old Chinese dynasty name.	“呉服”即和服, 是源于中国古代王朝的一个名字。	「呉服」即和服, 是源于中國古代王朝的一個名字。	기모노를 의미하는 오복(吳服)이라는 명칭은 중국의 옛 왕조 이름에서 유래하였습니다.	O termo gofuku, que significa kimono, deriva do nome de uma antiga dinastia chinesa.	Der Begriff Gofuku bedeutet Kimono und wurde vom Namen einer alten chinesischen Dynastie abgeleitet.

A.2 大学ドキュメント

Japanese	English	Chinese (Simplified)	Korean
はじめに	Introduction	简介	처음에
留学生ハンドブック (生活編) について	About the International Student Handbook (Living Guide)	留学生手册(生活指南)	유학생 안내 책자 (생활 편) 소개
この「留学生ハンドブック (生活編)」には、みなさんが充実した留学生活を送るための大切な情報を掲載しています。	This International Student Handbook (Living Guide) contains important information to help students get the most from international student life.	本《留学生手册(生活指南)》中的重要信息可以帮助学生尽快适应留学生生活。	이 유학생 안내 책자(생활 편)는 유학 생활에 필요한 중요한 정보를 제공합니다.
「留学生ハンドブック (履修編)」(学部生対象)と併せて、学生生活に積極的に活用してください。	Please actively utilize this guide and the International Student Handbook (Course Guide) (for undergraduate students) to help make your student life easier and more enjoyable.	请灵活运用本指南和《留学生手册(课程指南)(本科生用)》以使自己更加轻松地融入留学生活之中,并充实自己的留学生活。	유학생 안내 책자(학사 과정 편)(학부생용)를 적극 활용하여 더욱 편안하고 즐거운 학교 생활을 하시기 바랍니다.
みなさんはこれから立命館大学で日本人学生といっしょに授業を受け、同じように学生生活を送ります。	At Ritsumeikan University you will take classes with Japanese students, and enjoy the same student life as Japanese students.	在立命馆大学您将与日本学生一同上课,并体验与日本学生一样的学生生活。	여러분은 이제부터 리츠메이칸 대학교에서 일본인 학생들과 함께 수업하고 학교 생활을 하게 될 것입니다.

A.3 寺院ドキュメント

Japanese	English	Chinese (Simplified)	Spanish	German	French
高台寺月真院座禅体験	Zen Meditation Experience at Kodai-ji Gesshin-in Temple	高台寺月真院坐禅体験	Experiencia de meditación Zen en el Templo Kodai-ji Gesshin-in	Erleben der Zen-Meditation im Kodai-ji Gesshin-in-Tempel	Expérience de méditation Zen au temple Kodai-ji Gesshin-in
高台寺月真院は臨済宗建仁寺派の寺院です。	Kodai-ji Gesshin-in Temple is a temple of the Kennin-ji school of the Rinzai sect.	高台寺月真院属于临济宗建仁寺派的寺院。	El Templo Kodai-ji Gesshin-in es un templo de la escuela Kennin-ji de la secta Rinzaï.	Der Kodai-ji Gesshin-in-Tempel ist ein Tempel der Kennin-ji-Linie der Rinzaï-Schule.	Le temple Kodai-ji Gesshin-in est un temple du temple Kennin-ji d'obédience de l'école Rinzaï.
1617年の創建当初は高台寺と同様に曹洞宗でしたが、1624年に本寺である高台寺が曹洞宗から臨済宗に変わったときに必然的に変わりました。	Gesshin-in Temple is actually a sub-temple of Kodai-ji Temple. Upon its founding in 1617, Gesshin-in Temple was part of the Soto sect just like Kodai-ji Temple. However, when the head temple (Kodai-ji Temple) converted to the Rinzaï sect in 1624, Gesshin-in Temple also naturally converted to the Rinzaï sect.	1617年建造之时，月真院和高台寺同属曹洞宗，1624年高台寺改为临济宗后，月真院也随之改为临济宗。	El templo Gesshin-in es un templo secundario del Templo Kodai-ji. Cuando se fundó en 1617, el templo Gesshin-in era parte de la Secta Soto, así como el Templo Kodai-ji. Sin embargo, cuando el templo principal (Templo Kodai-ji) fue convertido a la Secta Rinzaï en 1624, el Templo Gesshin-in también fue convertido a la Secta Rinzaï.	Der Gesshin-in-Tempel ist in ein Zweigtempel des Kodai-ji-Tempels. Nach seiner Gründung im Jahr 1617 gehörte der Gesshin-in-Tempel zur Soto-Schule, genau wie der Kodai-ji-Tempel. Als sich der Kodai-ji-Tempel im Jahr 1624 jedoch der Rinzaï-Schule anschloss, schloss sich der Gesshin-in-Tempel natürlich ebenfalls der Rinzaï-Schule an.	Le temple Gesshin-in est en fait un temple secondaire du temple Kodai-ji. Lors de la fondation en 1617, le temple Gesshin-in faisait partie de la secte Sôtô tout comme le temple Kodai-ji. Cependant, quand le temple principal (temple Kodai-ji) fut converti par la secte Rinzaï en 1624, Le temple Gesshin-in fut naturellement également converti en secte Rinzaï.