社団法人 電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE

協調型機械翻訳システムのための予測入力インタフェース

岸田 章† 北村 泰彦‡

†関西学院大学大学院理工学研究科 〒669-1337 三田市学園 2-1
‡関西学院大学理工学部 〒669-1337 三田市学園 2-1
E-mail: †czy62628@ksc.kwansei.ac.jp, ‡kitamura-lab@ksc.kwansei.ac.jp

あらまし 既存の機械翻訳システムにおける翻訳の質は入力文に大きく依存する．協調型機械翻訳システムでは，システムが翻訳結果を入力言語に折り返し翻訳し，ユーザがそれをもとに入力文を修正することで協調する．しかし，機械翻訳の初心者が翻訳しやすい入力文に修正することは必ずしも容易ではない．そこでその支援のために，正しい翻訳を得られた入力ログを解析し，翻訳しやすい入力文を提示する予測入力インタフェースを提案する．

キーワード 言語グリッド，機械翻訳，折り返し翻訳，予測入力

Predictive input interface for collaborative machine translation system

Akira KISHIDA† Yasuhiko KITAMURA‡

†Graduate School of Science and Technology, Kwansei Gakuin University
2-1 Gakuen, Sanda-shi, Hyogo, 669-1337 Japan
‡School of Science and Technology, Kwansei Gakuin University 2-1 Gakuen, Sanda-shi, Hyogo, 669-1337 Japan
E-mail: †czy62628@ksc.kwansei.ac.jp, ‡ykitamura@ksc.kwansei.ac.jp

Abstract The quality of translation by machine translation systems greatly depends on the input sentence. In a collaborative machine translation system, the system translates the translated result into the input language again in a reverse way and the user modifies the input sentence referring to the back translation. However, it is not easy for a non-expert on machine translation to modify the input sentence to an appropriate one. In this paper, we propose a predictive input interface for collaborative machine translation system and it proposes input phrases which are easy to be translated by analyzing the log of correct inputs.

Keyword Language Grid, machine translation, back translation, predictive input method

1. はじめに

言語グリッドプロジェクトは，インターネット上の言語資源（対訳辞書など）や言語処理機能(機械翻訳など)を自由に組み合わせて使うことによって多言語翻訳サービスの実現を目的とする[1]．例えば，日本語→ロシア語の翻訳サービスが利用できない場合でも，日本語→英語と英語→ロシア語の翻訳サービスを組み合わせることによって，日本語からロシア語への翻訳を行うことができる．

プロジェクトには，様々な組織団体がパートナーとして活動しており，その１つが JEARN である．JEARN は世界最大の国際教育ネットワーク iEARN の日本センターとして，国際協働プロジェクトを推進する教育 NPO（特定非営利活動法人）である．JEARN 主催の防災世界こども会議では，電子掲示板上で英語でのやり取りを行っている．しかしながら，英語を母国語としないこども達は思うように発言できないことが多い．そこで，英語で発言することへの苦手意識を取り除くことができれば，こども達が積極的にディスカッションに参加することができ，お互いの意見を交換することができる．

現在 JEARN アクティビティでは，折り返し翻訳機能を持つ言語グリッドシステムが導入されている．折り返し翻訳機能は，翻訳結果をもう一度入力言語に翻訳しなおす機能であり，翻訳言語を理解できないユーザでも，翻訳結果の良し悪しの確認を行うことができる．ユーザは折り返し翻訳を参照しながら，入力文を修正することで,適切な翻訳結果を得ることができる．しかしながら，機械翻訳に十分な知識を持っていないこども達には，どういう文が機械翻訳しやすい文であるかが分からないという問題がある.そこで本研究は，ユーザが機械翻訳しやすい文の入力を支援する予測入力インタフェースを提案する．

2. 協調型機械翻訳システム

現在 Web 上にある Excite や Yahoo! JAPAN 翻訳など

の機械翻訳システムにおける翻訳の質は入力文に大きく依存する．例えば，主語の有無だけで翻訳文の内容は変化する．「英語の授業を楽しんでいますか.」と主語のない文を入力した場合，”Does it enjoy the class of English?”という翻訳結果を得る．一方，「あなたは英語の授業を楽しんでいますか.」と主語を付けて入力した場合，”Are you enjoying the class of English?”という翻訳結果を得る．このように入力文の主語があるかないかで翻訳の質は大きく変化する。

しかし，全く英語を理解できない人にとっては，翻訳結果が正しいかどうかの判定ができないので，入力文を修正することは難しい．その対策として，折り返し翻訳（back translation）を使うことが考えられる[2]. 折り返し翻訳とは，翻訳結果を再度入力言語に翻訳することである．折り返し翻訳を用いることによって，ユーザは翻訳結果の内容を母国語で確認することができる．例えば，日本語で入力し，英語に翻訳する場合には，日本語→英語→日本語と翻訳を行う．このように折り返し翻訳結果の日本文を見ることで翻訳文の正誤を推察することができ，入力文の修正を行うことによって翻訳結果を改善することができる[3].

言語グリッドプロジェクトでは，このような折り返し翻訳機能を実装した Langrid Input システム（図１）を既に開発している．システムは入力文に対する翻訳結果を表示し，折り返し翻訳結果の提供を行い，ユーザは折り返し翻訳結果を確認しながら入力文の修正を行う．そして，ユーザが入力文と折り返し翻訳結果を比べて，おおよその意味が同じになったと判断したときに，翻訳文は完成する．

図１：Langrid Input

図１のように，Langrid Input は下段に入力スペースがあり，中央のスペースに翻訳結果が表示され，上段のスペースに折り返し翻訳結果が表示される.例では，「フォーラムへの書き込みは英語です。」という文が入力されており,折り返し翻訳を確認すると,「フォーラムに投稿することはイギリス風である。」と表示されている．この場合，入力文と折り返し翻訳結果の意味が異なるので，翻訳結果である”Writing in to electronic forum is English.”は正しくないと判断できる．そこで，折り返し翻訳結果を確認しながら入力文の修正作業を行う．

本稿では，このようにしてシステムとユーザが協調して正しい翻訳文を作り上げるシステムを協調型機械翻訳システムと呼ぶ．協調型機械翻訳システムは，JEARN アクティビティの中でこども達に導入されようとしている．こども達が，防災世界こども会議で使用されている電子掲示板への書き込みを行う際，日本語で書き込みたい内容を考え，協調型機械翻訳システムを用いることで，折り返し翻訳結果を確認しながら書き込む英文を作成することができる．しかし，こどもにとっては入力文と折り返し翻訳結果を比べて，翻訳結果がおかしいと分かっても，機械翻訳しやすい入力文に修正することは容易ではない．こども達が入力を行う際に機械翻訳しやすい文を入力できるように誘導する入力支援が可能であれば，こどもでも質の高い翻訳結果を得ることができると考えられる．そこで，正しい翻訳結果を得たログを利用することで，機械翻訳しやすい入力文をユーザに提示する予測入力インタフェースを提案する．

3. 予測入力

予測入力とは単語辞書の情報やユーザの入力履歴などに基づいて，ユーザが入力した単語の部分的な読みなどから入力単語を予測し，複数の候補をユーザに提示して選択させることにより，少ないキー入力で効率的な文書作成を実現する文字入力手法である．現在は携帯電話などの文字入力に利用されている．

例えば図２に示す予測入力は「私達」という単語の入力を,「わた」という先頭の文字の入力と，単語の選択によって行っている．

これまでに予測入力は次に入力されるべき単語の予測をするものであった．例えば，携帯電話などに搭載されている予測入力システム PoBox[4]では，単語の一般的な出現頻度の情報やユーザの操作履歴などが予測に使用される．そして，日本語動的単語補完手法として開発された Nanashiki では編集中の文書から単語を抽出し，予測候補に加えるという機能を持つ[5]．さらに,文書蓄積システム Kukura を用いた予測入力では，

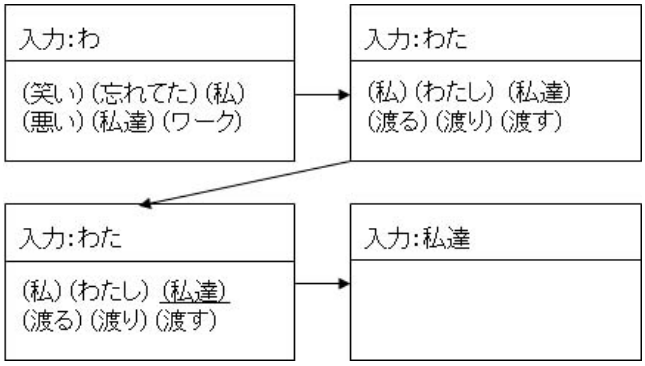

図２：予測入力

ウェブページや閲覧中の文書を予測に用いている[6]. 日本語動的単語補完手法 Nanashiki と文書蓄積システム Kukura を用いた予測入力は PoBox と併用することでより質の高い予測入力を行うことが可能である.

これまでの予測入力機能は，ユーザのキー入力の回数を減らすことで，効率よく入力できるようにすることが目的であった.

4. 協調型機械翻訳のための予測入力インタフェース

協調型機械翻訳システムのための予測入力インタフェース実装の目的は，ユーザが機械翻訳しやすい文を入力できるようにするための支援を行うことである. 機械翻訳しやすい文になるように，前節で述べた予測入力という形で，候補単語をユーザへ提示していくことによって，支援を行う.

今回提案する文章の予測入力インタフェースの利用例を図３に示す．このインタフェースの特徴は，英文翻訳において主語が重要であるとの観点から，①に示すようにユーザが何も入力されていない状態からでも候補の提示が始まる点である.

この予測入力インタフェースは，正しい翻訳結果を得た入力ログを利用することで，機械翻訳しやすい入力文をユーザに提示する．そのためには，入力ログを，解析し，文章形態と単語を別々に保存する．そして，文章形態のログと単語のログを連携させて候補を提示する.

以下，文章形態ログと単語ログの保存方法，予測候補の提示の手法について述べる.

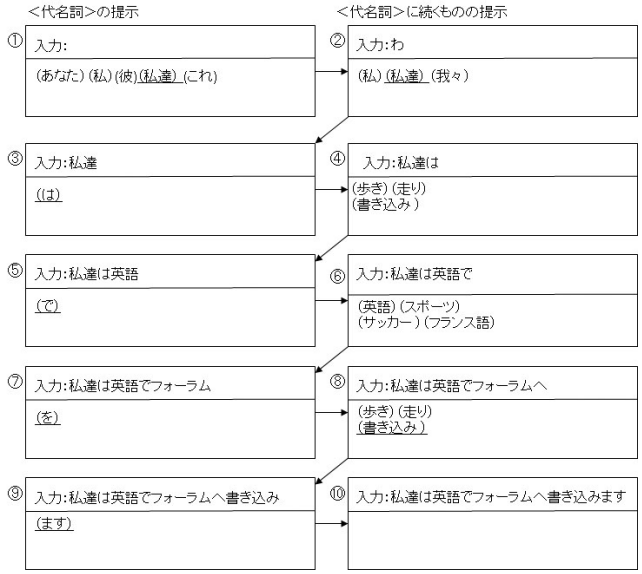

図３：予測入力インタフェースの利用例

4.1. 文章形態ログの保存

ユーザに対して予測候補として推薦する文章形態ログの保存について述べる．まず，入力文の形態素解析を行う．そして，助詞，助動詞以外の品詞である名詞，代名詞，動詞，形容詞，形容動詞，連体詞，副詞，接続詞，感動詞をそれぞれ＜名詞＞，＜代名詞＞，＜動詞＞，＜形容詞＞，＜形容動詞＞，＜連体詞＞，＜副詞＞，＜接続詞＞，＜感動詞＞に置き換え，文章形態ログとして保存する．＜名詞＞など，抽象的な形に置き換えることにより，予測候補を提示する際に柔軟性ができる.

例えば，正しい入力文が「私達は英語でフォーラムへ書き込みます」であれば，「＜代名詞＞は＜名詞＞で＜名詞＞へ＜動詞：自立＞ます」を文章形態ログとして保存する.

4.2. 単語ログの保存

ユーザによって入力された文の中の単語を予測候補として保存する.まず,入力文の形態素解析を行う. 次に助詞，助動詞以外の品詞である単語を＜名詞＞，＜代名詞＞，＜動詞＞，＜形容詞＞，＜形容動詞＞，＜連体詞＞，＜副詞＞，＜接続詞＞，＜感動詞＞の品詞別に分け，品詞情報と合せて単語ログとして保存する．品詞別に保存することにより，品詞を絞った単語の予測候補を提示することが可能になる.

例えば，正しい入力文が「私達は英語でフォーラムへ書き込みます」であれば，「英語」「フォーラム」を＜名詞＞，「私達」を＜代名詞＞，「書き込み」を＜動詞:自立＞の予測候補となる単語ログとして保存する.

4.3. 予測提示

過去の入力ログを元に生成された単語と文章形態のログを用い，予測候補の提示を行う．編集中の入力文を形態素解析，置き換えを行い，前方一致する文章形態のログを検索する．前方一致するものが存在すれば，次に入力すると予測される品詞による予測単語の絞込みを行い，その品詞の単語のログを予測候補として提示する．また，前方一致する文章形態が存在しなければ，何も推薦しない．

4.4. 具体例

図３をもとに，予測入力インタフェースを用いた具体的な入力の例を挙げる．Ⅰ:「＜代名詞＞は＜動詞：自立＞ます」とⅡ:「＜代名詞＞は＜名詞＞で＜名詞＞を＜動詞：自立＞ます」とⅢ:「＜代名詞＞は＜名詞＞で＜名詞＞へ＜動詞：自立＞ます」という３つの文章形態ログが存在する場合で考える．優先順位は高いものからⅠ，Ⅱ，Ⅲという順とする．

まず，まだ何も入力していない①の場面では，一番優先順位の高い文章形態ログⅠ:「＜代名詞＞は＜動詞：自立＞ます」の先頭の＜代名詞＞が推薦される．＜代名詞＞の単語ログの中では，「あなた」，「私」，「彼」，「私達」，「これ」などの単語が存在するので，それらの単語を予測候補として提示する．そこで，②のようにユーザが「わ」と入力することによって「私」，「私達」，「我々」の単語に絞り込まれる．そこで「私達」を選ぶことによって，「私達」を入力として決定する．次に，③のように「＜代名詞＞」の次の「は」が予測候補として提示される．④では，「＜代名詞＞は」の次に来るものとして＜動詞：自立＞が推薦される．＜動詞：自立＞の単語ログの中では「歩き」，「走り」，「書き込み」などの単語が存在するので，それらの単語を予測単語として提示する．そこで，今回は「英語」という単語を入力する．「英語」という単語は＜名詞＞に区分される．すると，「＜代名詞＞は＜名詞＞」という文章形態になり，Ⅰのログとは前方一致しなくなる．この場合，他のログの中で優先順位の高い文章形態ログから順に，入力文の文章形態と前方一致するものを検索する．次に優先度の高い文章形態ログⅡ:「＜代名詞＞は＜名詞＞で＜動詞：自立＞ます」と比べと，前方一致するので，今度はⅡが入力文の文章形態の予測候補となる．よって，⑤のように「＜代名詞＞は＜名詞＞」に続く「で」が推薦される．「で」を選択すると，⑥のように「＜代名詞＞は＜名詞＞で」に続く＜名詞＞が推薦される．＜名詞＞の単語ログは，「英語」，「スポーツ」，「サッカー」，「フランス語」が存在するが，今回は「フォーラム」と単語ログには存在しない単語を入力する．しかしながら，「フォーラム」は＜名詞＞に属するので，「＜代名詞＞は＜名詞＞で＜名詞＞」という文章形態ログと同じ構造を維持する形となる．よって，「フォーラム」という単語の入力後には⑦のように，「を」が推薦される．そこで，「を」を選ばず「へ」を入力する．すると入力文の文章形態は文章形態ログⅡの文章形態と異なる形となる．そこで，先程と同じように他のログの中で，入力文と文章形態が前方一致する文章形態ログを検索する．次に優先順位の高い文章形態ログⅢ:「＜代名詞＞は＜名詞＞で＜名詞＞へ＜動詞：自立＞ます」と前方一致することから，⑧のように「＜代名詞＞は＜名詞＞で＜名詞＞へ」の後の＜動詞：自立＞が推薦される．＜動詞：自立＞の単語ログは「歩き」「走り」「書き込み」が存在し，「書き込み」を選ぶ．すると⑨のように「＜代名詞＞は＜名詞＞で＜名詞＞へ＜動詞：自立＞」に続いて入力すると予測される「ます」が予測候補として推薦される．「ます」を選択すると⑩のような文「私達は英語でフォーラムへ書き込みます」が完成する．

5. まとめ

協調型機械翻訳システムはユーザと機械翻訳システムの協調により，ユーザが入力文を修正しながら正しい機械翻訳を行うシステムである．しかし，ユーザが機械翻訳しやすい文を入力することが容易ではない．そこで，その問題の解決法として予測入力インタフェースを提案した．今後は，JEARN アクティビティの中のこども達に利用されることを目標に，予測入力インタフェースを実装していく．

文　献

[1] 言語グリッドホームページ http://langrid.nict.go.jp/indexj.htm

[2] 小倉 健太郎，林 良彦，野村 早恵子，石田 亨．機械翻訳を介したコミュニケーションにおけるユーザの機械翻訳システム適応の言語依存性，自然言語処理，Vol.12，No. 3，pp. 183-202，2005.

[3] 石田 亨．異文化コラボレーション研究の構想，異文化コラボレーション研究グループ, 2006.

[4] T.Masui. An efficient text input method for penbased computers. In *Proceedings of the ACMConference on Human Factors in Computing Systems (CHI '98)*, pp. 328–335, 1998.

[5] 小松 弘幸，高林 哲，増井 俊之．動的略語展開を利用した文脈をとらえた予測入力，情報処理学会論文誌，Vol.44，No.11，2003.

[6] 小松 弘幸，高林 哲，増井 俊之．文書蓄積システム Kukura を用いた予測入力，WISS，2002.