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ガイド入力インタフェースの開発

岸田章†　　北村泰彦††

† 関西学院大学大学院理工学研究科
†† 関西学院大学理工学部
〒 669-1337 三田市学園 2-1
E-mail: †{a-kishida,ykitamura}@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，その結果をユーザが確認しながら入力文を修正することで協調する．しかし，機械翻訳システム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，正しく翻訳された過去の入力文を解析し，翻訳しやすい入力文を提示するガイド入力インタフェースを開発する．
キーワード　言語グリッド，機械翻訳，折り返し翻訳，ガイド入力

Development of guided input interface for collaborative machine translation system

Akira KISHIDA† and Yasuhiko KITAMURA††

† Graduate School of Science and Technology, Kwansei Gakuin University
†† School of Science and Technology, Kwansei Gakuin University
E-mail: †{a-kishida,ykitamura}@ksc.kwansei.ac.jp

Abstract　The quality of translation by machine translation systems greatly depends on the input sentence. Collaborative machine translation systems back-translate the translated result into the input language in a reverse way and the user modifies the input sentence referring to the back translation. However, it is not easy for non-experts of machine translation to modify the input sentence to an appropriate one. In this paper, we develop a guided input interface for collaborative machine translation systems, which suggests input phrases which are easy to be translated by analyzing the log of correct inputs.
Key words　Language Grid，machine translation，back translation，guided input method

1. はじめに

近年，インターネットが世界中に普及したことによって，国際的なコラボレーションが可能になった．それに伴い，コミュニケーションの多言語化が進行し，機械翻訳システムの必要性が高まっている．しかし現状では，言語によっては適切な機械翻訳システムが見つけられない，機械翻訳システムの翻訳の質が低いなどの問題がある．そこで，既に存在するさまざまな言語サービスを連携させることによって言語の壁を越えることを目的として，NICT，大学，NTT などの研究グループは産官学民協同の体制で言語グリッドプロジェクトを立ち上げた [1].

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，日本語からロシア語への翻訳サービスが利用できない場合でも，図 1 のように日本語から英語への翻訳サービスと英語からロシア語への翻訳サービスの 2 つ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．さらに，コミュニティ辞書を連携させることで，そのコミュニティの中で使われる単語を適切な形で翻訳することが可能になり，翻訳の質をより向上させることができる．

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（Japan Education and Resource Network）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）であり [2]，こども達の国際交流を目的として活動している．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

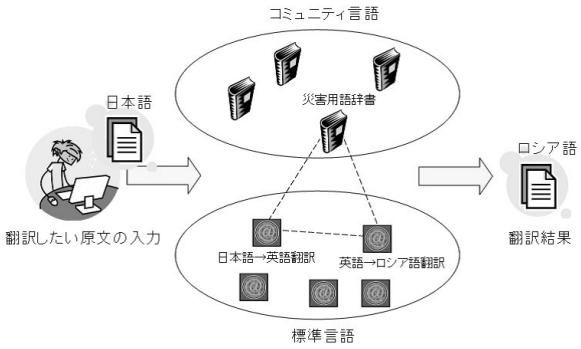

図 1　日本語→ロシア語の災害についての翻訳の例（日本語→ロシア語の翻訳サービスが利用できない場合）
Fig. 1　An example of translation from Japanese to Russian about disaster.

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言語の壁を意識してしまうことによって，思うように翻訳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 [3]．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 [4]，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，折り返し翻訳の結果から入力文を修正する必要があるということが確認できても，どのような文が機械翻訳しやすい文であるかが分からないので，結局思うように発言でき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ガイド入力インタフェースを開発する [5]．

本稿では，２章で JEARN アクティビティで導入されている協調型機械翻訳システムとその課題について述べる．３章では予測入力について，４章では協調型機械翻訳システムのためのガイド入力インタフェースについて述べる．５章ではガイド入力インタフェースの実装について，６章ではまとめと今後の課題について述べる．

2.　協調型機械翻訳システム

現在 Web 上にある Excite. 翻訳や 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.」という文を入力すると，”It enjoys the class of English.”という正しくない翻訳結果を得る．一方，「私は英語の授業を楽しんでいます.」という入力文に対しては，”I am enjoying the class of English.”という正しい翻訳結果が得られる．このように，入力文に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 [4]．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 [6]．

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 2）を開発している [7]．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 2　Langrid Input
Fig. 2　Langrid Input.

図 2 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，こども達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，こども達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ガイド入力インタフェースを提案する．

3. 予測入力

ガイド入力の関連する研究として，予測入力が存在する．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 3 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．

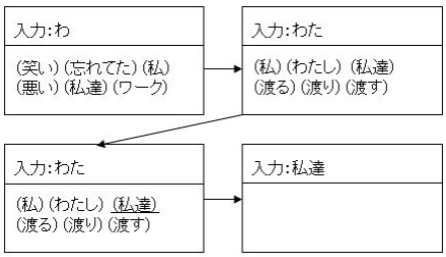

図 3　予測入力
Fig. 3　Predictive Input.

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 [8] 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．

4. 協調型機械翻訳システムのためのガイド入力インタフェース

協調型機械翻訳システムのためのガイド入力インタフェースの目的は，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，前章で述べた予測入力と似た形で，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．以下でガイド入力インタフェースのシステム構成，入力文の解析方法，ガイド候補の提示方法と具体例について述べる．

4.1 ガイド入力インタフェースのシステム構成

ガイド入力インタフェースのシステム構成は図 4 のようなクライアント・サーバ型になる．

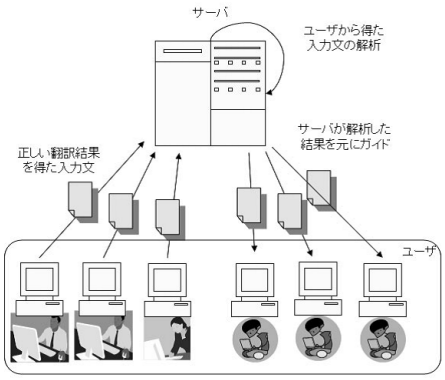

図 4　ガイド入力インタフェースのシステム構成
Fig. 4　System components of Guided input interface.

このシステムの特徴は，利用ログを蓄積していくことで，ガイドされる文章構造のレパートリーが増えていくところである．利用者の誰かがこのシステムを利用して，ある正しい翻訳結果を得られる入力文を完成させたとする．すると，サーバでその入力文を解析し，ログとして保存する．その後，他の人がこのシステムを利用して翻訳文を作成しようとしたときに，以前に取り込んだログを元にガイドする．この仕組みによって，他の人が作成した入力文を元に，英語に対する知識のない機械翻訳初心者でも正しい翻訳文を作成する支援ができる．

4.2 入力文の解析方法

正しい翻訳結果を得ることができた入力文を，その後行われる入力のガイドに利用するため，解析し，文章構造ログ，単語ログとしてそれぞれ保存する．解析し，ログの保存までのアルゴリズムを以下に示す．

（1） 入力文の形態素解析を行う．

（2） 単語をそれぞれの品詞別に分け，品詞情報と合せて単語ログとして保存する．

（3） 入力文の出現単語をそれぞれ品詞に置き換え，文章構造ログとして保存する．

解析時に利用される形態素解析システムとは，単語辞書や各品詞の接続確率などをもとに，文の構造を解析するものである．文章構造ログを保存する際に，＜名詞-代名詞-一般＞，＜名詞-一般＞，＜動詞-自立＞などのように抽象的な形に置き換えて保存し，単語を品詞別に保存しておくことによって，文章構造を保ちながら，その文章構造によって推薦される品詞に含まれる単語を柔軟にガイドしていくことが可能となる．具体的な解析の例を図 5 に示す．

正しい翻訳結果を得ることができた入力文が「私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-一般＞,「私」を＜名詞-代名詞-一般＞,「書き込み」を＜動詞-自立＞,「は」を＜助詞-係助詞＞,「で」,「へ」を＜助詞-格助詞-一般＞,「ます」を＜助動詞＞,「.」を＜記号-句点＞というように，品詞別に単語ログとして保存し,「＜名詞-代名詞-一

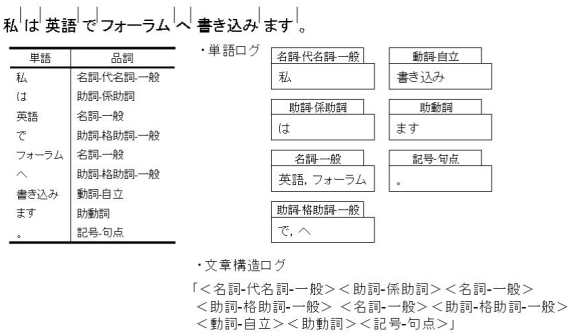

図 5　解析の例
Fig. 5　An example of analysis.

般＞＜助詞-係助詞＞＜名詞-一般＞＜助詞-格助詞-一般＞＜名詞-一般＞＜助詞-格助詞-一般＞＜動詞-自立＞＜助動詞＞＜記号-句点＞」を文章構造ログとして保存する．

4.3　ガイド候補の提示方法

保存した単語ログと文章構造ログを用いてガイドを行う．ガイドのためのアルゴリズムを以下に示す．

（1）　入力欄が空白の場合，文章構造ログのそれぞれ最初に来る品詞に含まれる単語を単語ログからガイド候補として提示する．

（2）　キーボードによる文字入力が行われた場合，ガイド候補として提示されている単語の中から読み方が前方一致するものに絞込みを行う．

（3）　キーボード入力やガイド候補の選択によって入力する単語が決定されると，入力スペースにある文字列を形態素解析し，出現単語を品詞へ置き換える．

（4）　3の置き換えの結果と前方一致する文章構造を文章構造ログから検索する．

（5）　4の結果，前方一致する文章構造が存在すれば，その中で前方一致する部分の次に来る品詞を読み込み，その品詞に含まれる単語を単語ログからガイド候補として提示する．前方一致する文章構造が存在しなければ，何も提示しない．

（6）　2～5を繰り返し，一つの文章が完成すれば終了する．

ユーザの単語入力から新たなガイド候補となる単語の提示までの流れは図 6 のようになる．

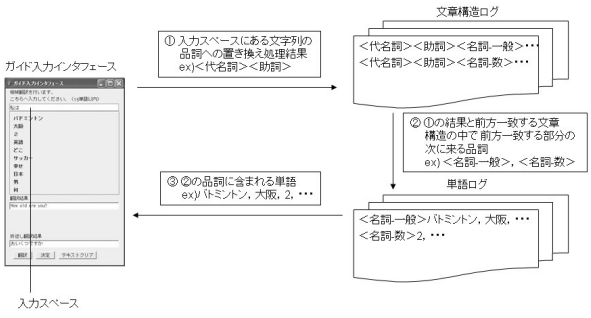

図 6　ガイドの流れ
Fig. 6　Flow of guide.

ガイド入力インタフェースの入力スペースにある文字列を形態素解析し，品詞への置き換えを行う．図 6 の場合,「＜代名詞＞＜助詞＞」とする．次に，その置き換え結果と前方一致する文章構造の前方一致する部分のすぐ後ろに来る品詞を読み込む．図 6 の場合,「＜代名詞＞＜助詞＞＜名詞-一般＞ … 」,「＜代名詞＞＜助詞＞＜名詞-数＞ … 」という文章構造が存在するので，＜名詞-一般＞，＜名詞-数＞が読み込む対象となる．そして，読み込んだ品詞に含まれる単語を単語ログから読み込む．図 6 の場合，＜名詞-一般＞に含まれる「バドミントン」,「大阪」など，＜名詞-数＞に含まれる「2」などが読み込む対象となる．最後に，その読み込んだ単語をガイド候補として提示する．

4.4　ガイド入力の具体例

図 7 をもとに，ガイド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．文章構造ログの中に，文章構造 1 :「＜名詞-代名詞-一般＞＜助詞-係助詞＞＜動詞-自立＞＜助動詞＞＜記号-句点＞」と文章構造 2 :「＜名詞-代名詞-一般＞＜助詞-係助詞＞＜名詞-一般＞＜助詞-連体化＞＜名詞-サ変接続＞＜助詞-格助詞-一般＞＜動詞-自立＞＜助詞-接続助詞＞＜動詞-非自立＞＜助動詞＞＜記号-句点＞」文章構造 3 :「＜名詞-代名詞-一般＞＜助詞-係助詞＞＜名詞-一般＞＜助詞-連体化＞＜名詞-サ変接続＞＜助詞-格助詞-一般＞＜動詞-自立＞＜助詞-接続助詞＞＜動詞-非自立＞＜助動詞＞＜助詞-副助詞／並立助詞／終助詞＞＜記号-句点＞」という 3 つの文章構造が存在する場合で考える．

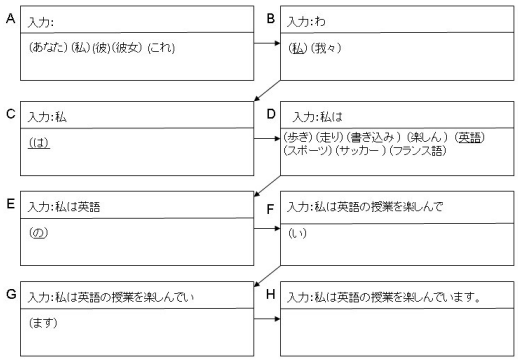

図 7　協調型機械翻訳システムのためのガイド入力
Fig. 7　Guided input for collaborative machine translation system.

A に示すように，ユーザが何も入力していない状態からでも候補の提示は始まる．文章構造 1，2，3 の先頭にある＜名詞-代名詞-一般＞が推薦され，単語ログの中で，＜名詞-代名詞-一般＞に含まれる単語は「あなた」,「私」,「彼」,「彼女」,「これ」などが存在する．よって，それらの単語をガイド候補として提示する．そこで，B のようにユーザが「わ」と入力することによって「わ」の読みで始まる「私」,「我々」の単語に絞り込まれる.「私」を選ぶことによって,「私」を入力として決定する．

次に,「私」を形態素解析，品詞への置き換えを行った結果,「＜名詞-代名詞-一般＞」となり，文章構造 1，2，3 全てと前方一致する．よって，C のように「＜名詞-代名詞-一般＞」の次の＜助詞-係助詞＞に含まれる単語「は」が予測候補として提示される．

D では,「私は」を解析した結果である「＜名詞-代名詞-一般＞＜助詞-係助詞＞」に続く文章構造１の＜動詞-自立＞，文章構造２，３の＜名詞-一般＞が推薦される．＜動詞-自立＞に含まれる「歩き」,「走り」,「書き込み」「楽しん」などの単語＜名詞-一般＞に含まれる「英語」,「スポーツ」,「サッカー」,「フランス語」などの単語が存在するので，それらの単語をガイド候補として提示する．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-一般＞に区分される．すると,「私は英語」を解析した結果,「＜名詞-代名詞-一般＞＜助詞-係助詞＞＜名詞-一般＞」という文章構造になり，文章構造１とは前方一致しなくなる．この場合,「＜名詞-代名詞-一般＞＜助詞-係助詞＞＜名詞-一般＞」という文章構造と前方一致する文章構造２，３のみがガイド候補となる（図 8）．よって，E のように「＜名詞-代名詞-一般＞＜助詞-係助詞＞＜名詞-一般＞」に続く＜助詞-連体化＞に含まれる単語「の」が提示される．

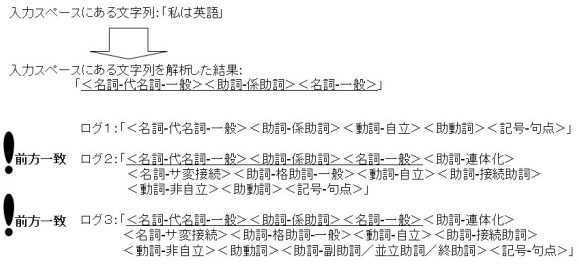

図 8　ログ検索の例
Fig. 8　An example of searching sentence log.

その後入力を続けていくと，F では，＜動詞-非自立＞に含まれる単語「い」が提示されている．このように,「い」が提示されることで,「楽しんでます。」のような「い」の抜けた正しく機械翻訳できない文の入力を避けることができる．

このようにしてガイドに従って入力していくと，H のように「私は英語の授業を楽しんでいます.」という機械翻訳しやすい入力文が完成する．入力スペースにある文字列の品詞への置き換え処理の結果,「＜名詞-代名詞-一般＞＜助詞-係助詞＞＜名詞-一般＞＜助詞-連体化＞＜名詞-サ変接続＞＜助詞-格助詞-一般＞＜動詞-自立＞＜助詞-接続助詞＞＜動詞-非自立＞＜助動詞＞＜記号-句点＞」という文章構造になり，前方一致し，なおかつ後ろに候補が続く文章構造は存在しなくなるとガイドを終了する．

5. 実　　装

前章で述べた方法を用いて，図 9 の左にあるガイド入力インタフェース実装した．実装は java で行った．

5.1 システム構成

協調型機械翻訳のためのガイド入力インタフェースのシステム構成図を図 10 に示す．

システムは正しい翻訳結果を得ることができた入力文から，前章で述べた方法で単語ログ，文章構造ログを保存する．そして，保存したログ，ユーザの入力を元にガイドを行う．また，

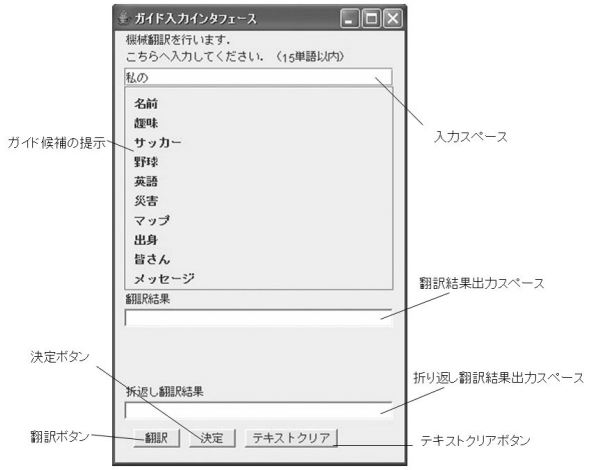

図 9　ガイド入力インタフェース
Fig. 9　Guided input interface.

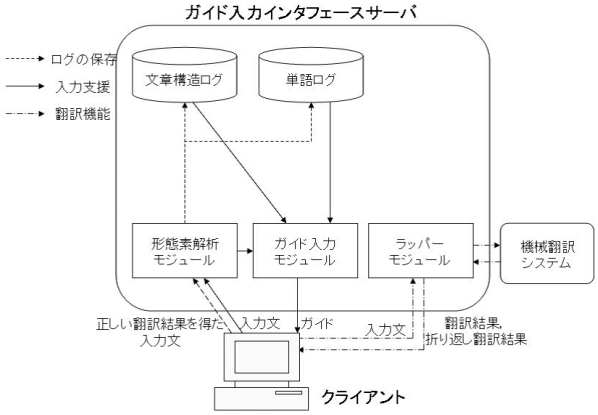

図 10　ガイド入力インタフェースのシステム構成図
Fig. 10　System component of guided input interface.

機械翻訳システムによって翻訳結果，折り返し翻訳結果を提供する．形態素解析システムには，java で実装されている形態素解析ツール Sen [9] を用いる．翻訳システムには,「Excite. 翻訳（http://www.excite.co.jp/world/）」を利用している．

5.2 ガイド入力

図 11 に示されているように，入力スペースに何も入力されていない状態から，ガイド候補となる単語が提示される．

プログラムの起動と同時に文章構造ログ，単語ログを読み込み，それぞれの文章構造で最初にくる品詞を調べ，その品詞に含まれる単語を単語ログから取得し，ガイド候補となる単語として提示している．なお，ガイド候補として提示されている単語は，３章で紹介した予測入力と同様に，ユーザが入力スペースに文字列を入力することによって，その入力文字列と読み方が前方一致する単語だけに絞り込まれる（図 12）．さらに，その中で表示される単語は，使用頻度の多い順に 10 個までとしている．

入力する単語を決定し，”Enter キー ”を押すと，ガイド候補であった単語が更新され，新たな単語がガイド候補として提示される．前方一致する文章構造が存在し，なおかつその文章構造において後ろに候補が続く限り，ガイドは継続される．

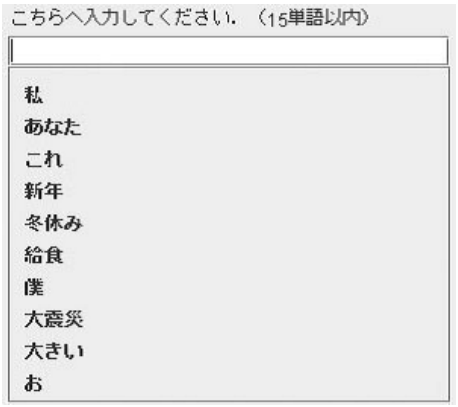

図 11　ガイド候補の提示
Fig. 11　The presentation of the candidates.

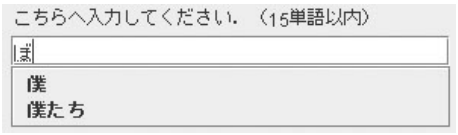

図 12　ガイド候補絞込みの例
Fig. 12　An example of narrowing down candidates.

5.3　翻訳，折り返し翻訳機能

入力文が完成し，”翻訳ボタン”を押すと，図 12 のように翻訳結果，折り返し翻訳結果がそれぞれのスペースに出力される．翻訳，折り返し翻訳機能には「Excite. 翻訳 (http://www.excite.co.jp/world/)」を利用している．「Excite. 翻訳」のホームページをラッパープログラムで情報抽出することによって，翻訳機能，折り返し翻訳機能を本システムに導入する．ラッッパープログラムとは，HTML 文書から部分的に情報を抽出するプログラムのことである．入力文，入力言語，翻訳言語をパラメータとして「Excite. 翻訳」のホームページに送信することで，Excite. 翻訳の翻訳結果を取得し，また，その翻訳結果を再び入力文として送信することで，折り返し翻訳結果を取得し，それぞれの結果をラッパープログラムで抽出している．２章で紹介した Langrid Input と同様に，入力文と折り返し翻訳結果を比較することで，翻訳結果の質を確認することができる．

5.4　ログの取得

折り返し翻訳結果の参照などを経て，正しく翻訳されたと確認すると，”決定ボタン ”を押す．すると，そのとき入力されている入力文を前章で述べた方法を用いて解析し，文章構造ログ，単語ログをそれぞれ保存する．

”テキストクリア ”ボタンを押すと，入力スペース，翻訳結果スペース，折り返し翻訳結果スペースの内容を全てクリアする．再び入力を行う際，”決定ボタン ”を押して取り込んだログは，ガイド候補として利用される．

6.　まとめと今後の課題

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を実現しようとしている．その中で開発された協調型機械翻訳システム Langrid Input 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，本研究では，その問題の解決法としてガイド入力インタフェースを提案し，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，実装に取り組んだ．

図 13　翻訳結果，折り返し翻訳結果の出力
Fig. 13　Output of translation and back translation.

今後は，実装したガイド入力インタフェースの評価実験に取り組む．ガイド入力インタフェースを利用する場合と，利用しない場合とで機械翻訳可能な文章入力成功者の割合の変化，成功までの所要時間の変化を観察する．

文　　献

[1] 言語グリッドホームページ. http://langrid.nict.go.jp/indexj.htm
[2] JEARN ホームページ. http://www.jearn.jp/
[3] 納谷淑恵．ICT を活用した実践的コミュニケーション能力の育成―防災をテーマとした英語での情報発信を通して―.
[4] 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.
[5] 岸田章，北村泰彦．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．電子情報通信学会技術研究報告，AI2006-72,2007.
[6] 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.
[7] 重信智宏，藤井薫和，吉野孝，灘本明代．言語グリッド: 異文化コラボレーション支援環境の構築．The 21st Annual Conference of the Japanese Society for Artificial Intelligence, 2007.
[8] 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.
[9] Sen Project http://ultimania.org/sen/
[10] 林田 尚子，石田 亨．翻訳エージェントによる自己主導型リペア支援の性能予測．電子情報通信学会論文誌，Vol.J88-D-I，pp.8，2005.
[11] 坂本ひろ子．マレーシア，フィリピン進出日系企業における異文化間コミュニケーション摩擦．多賀出版，2002.