

インスタントメッセージ会話文翻訳における主語補完による翻訳品質の改良

片江伸之[†]、鄭育昌[†]、長瀬友樹[†]、出内将夫[‡]、岡田伊策[‡]

[†] (株)富士通研究所 メディア処理研究所

[‡] 富士通(株) S I 技術本部 システム技術統括部

{katae, cheng.yuchang, nagase.tomoki, ideuchi.masao, isaac-okada}@jp.fujitsu.com

1 はじめに

企業のグローバル化が進むとともに、業務において外国語話者との言語の壁を越えたコミュニケーションの必要性が増大している。富士通グループでは、全社員共通のグループウェアであるグローバルコミュニケーション基盤を運用しており、そのアプリケーションの一つに、LyncIM(インスタントメッセージング)がある。弊社のLyncIMでは、機械翻訳を利用して外国語話者とのコミュニケーションを支援する「LyncIM翻訳」機能が利用できる。これはLyncIMのインターフェースの中で、社内の機械翻訳クラウドサービスの日英/英日翻訳、日中/中日翻訳を利用できるようにしたものである。

機械翻訳は外国語話者とのコミュニケーションを効率化する強力なツールであるが、翻訳品質が不十分な場合には、伝達内容の齟齬が生じ、結果として不要な会話が増加して効率が損なわれる。日本語を入力とする機械翻訳における品質低下の要因として、主語・目的語欠落の問題は古くから議論されているが、インスタントメッ

セージでも同様である。今回は、意味構造解析に基づく簡易なルールに基づく主語補完の導入により翻訳品質を改良するとともに、開発した主語補完機能を LyncIM のアプリケーションから利用できる日本語チェックサーバに実装した。

2 主語補完機能の開発

本研究の対象は、LyncIM を用いた日本語話者と外国語話者の業務連絡の会話である。先行研究^[1]では会話ログにおける発話の特徴の分析結果が報告されている。この報告では、会話ログでの主語や目的語の省略の出現確率は比較的低いが、翻訳品質への影響は他の要因と比べて大きく¹、主語補完は優先して取り組むべき課題であることが示唆されている。

図1に、今回開発する主語補完機能を適用する LyncIM の運用イメージを示す。①日本語話者が日本語によるメッセージを発信すると、システムは省略された主語を推定し、主語補完の必要性の有無を判断する。②補完が必

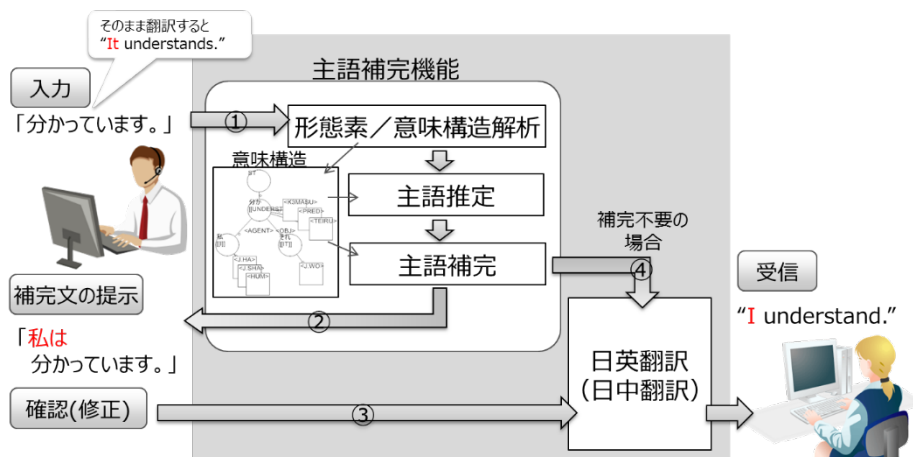


図1 主語補完機能を適用したインスタントメッセージングの運用イメージ

¹ 主語・目的語の省略を含む文が会話ログ全体に占める割合は約3%と少ないが、翻訳品質の指標である Acceptability^[4]は、全体平均が3.42のとこ
ろ、主語・目的語省略を含む文は3.0と低く、品質低下の要因となっている。

要な場合には、主語を補完した文が生成され発信者に提示される。③発信者は補完文を確認し、問題があれば適宜修正した後に再度発信を行う。補完文は翻訳されて相手に送信される。④補完が不要な場合には、最初のメッセージのまま翻訳されて相手に送信される。

システムによる主語補完では誤りを生じる可能性があるが、この運用方法によれば、発信者が確認、修正することで相手に誤った訳文を送ってしまう心配がなく、同時に、機械翻訳に適した日本語文の書き方を発信者に教示できると期待される。

2. 1 開発方針

今回は、機械翻訳向けに開発した意味構造解析技術④を利用し、ルールベース方式による主語補完機能を開発する。一般的な主語推定では、照応解析によって、会話履歴や文脈から候補となる単語を抽出し、処理対象テキスト中のゼロ代名詞に相当する単語を特定する方法が採られる。しかし、LyncIM 翻訳向けの機能として考えると、外国語話者側のメッセージが外国語のままであったり、機械翻訳結果であったりすることが多く、照応解析の候補となる単語が適切に抽出できるとは限らない。また、機械翻訳への適用を目的とすると、文の主語が、一人称、二人称、三人称のいずれであるかを特定するだけで事足りることが多い。それには会話履歴等を利用せずとも、処理対象の文の意味構造解析の結果を利用することで十分な性能が達成できると考え、ヒューリスティックな主語推定ルールおよび主語補完ルールを設計することとした。

2. 2 主語推定ルール

図2に例示するように、一文の意味構造は、意味の構成要素である「ノード」を、各ノードの関係を示す「関係アーク」で接続した有向グラフで表される。また、各ノードは文法的／意味的属性を示す「付加アーク」を持つ場合がある。

主語推定は、基本的に「述語」の属性（付加アーク）を持つ全てのノードに対して実施する。ただし、原文において既に「動作主」となるノードが存在する述語ノードや慣用句は処理対象から除外する。処理対象の述語ノードに対し以下の順にルールを適用して主語推定を行う。

- (1) 当該主語ノードに、「疑問」または「丁寧(PLEASE)」の属性がある場合

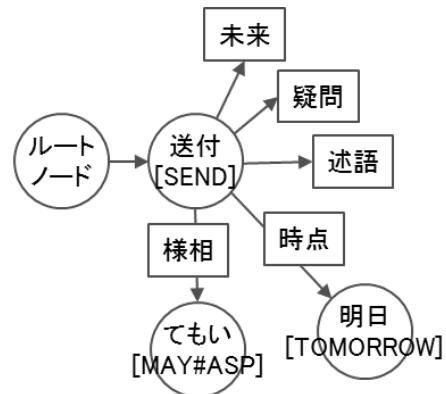


図2 意味構造の例（明日送付してもいいですか）

- ・「許可」または「申し出」を意味する関係ノードある場合、主語を**一人称**とする
例) 「○○してもいいですか」
「○○しなくてもいいですか」
「○○しましょうか」
- ・その他の場合は主語を**二人称**とする。
- (2) 当該述語ノードに「依頼」の属性がある場合、主語を**二人称**とする。
例) 「○○していただけますか」
「○○してもらえますか」
- (3) 当該述語ノードに「もらう」の属性がある場合、主語を**三人称**とする。
例) 「○○してもらいます」
- (4) 上記以外の場合、主語を**一人称**とする。

例えば図2に示した文では、ノード[SEND]は「述語」属性を持ち、「動作主」ノードが存在しないので主語推定の対象となる。さらに「疑問」の属性があり、ノード[MAY#ASP]と「様相」の関係アークで連結している(=「許可」)のため、上述のルール(1)により、主語は一人称と推定される。このような意味構造によるルールを記述すれば、一つのルールで複数の日本語表現のバリエーションをカバーすることができる。例えば、本ルールは、「～してもいいですか」「～してもよい?」「～してもいいかい?」などの表現に対応することができる。

2. 3 主語補完ルール

■ 主語補完の実施の可否の判定

原則として、前節で説明した主語推定により主語が特定された場合には、入力文に対する主語補完を行う。た

だし、以下の場合は主語補完を行わない。

- (1) 当該述語ノードに対して、関係アーク「並列」で連結する別の述語ノードがあり、かつ、それらが入力テキストで連続している場合

例) 「ちゃんと開けず確認できませんでした。」

※「確認できませんでした」が本ルールに該当するためその前には主語補完を行わず、「開けず」の前だけに主語「私は」が補完される。

- (2) 当該述語ノードに対して、関係アーク「並列」で連結する別の述語ノードがあり、かつ、その述語ノードに付加アーク「命令」がある場合

例) 「環境を作成して、テストを消化して下さい。」

※「作成して」が本ルールに該当する。「消化してください」には後述のルール(4)に従って主語補完を行わないが、それと同様に「作成して」にも主語補完を行わない。

- (3) 当該述語ノードが、ガ格のノードと連結する場合

例) 「今日のテストが終わりました。」

※「終わりました」には「テストが」と連結するため主語を補完しない。

- (4) 当該述語ノードが「命令」の場合

例) 「ちょっと待ってください。」

■ 主語を挿入する位置

原則として、述語ノードの直前に主語を挿入する。ただし、以下の場合は挿入位置をより前方に変更する。

- 述語が「する」であり、直前のサ変名詞が目的語である場合にはサ変名詞の前に挿入する。

例) 「マウントします」⇒「私はマウントします」

- 述語ノードから前方ノードをたどり、直前ノードと「連語」または「並列」の関係にある間、挿入位置を前方に移動する。

例) 「開発やテストを行います」

⇒「私は開発やテストを行います」

■ 挿入する文字列

- 挿入文字列は、主語推定結果が一人称の場合には「私」、二人称の場合は「あなた」、三人称の場合には「彼」とする。

- 主語を挿入する対象の述語が連体修飾節の中にある場合は、助詞「が」、その他の場合は助詞「は」を伴う。

例) 先ほどいただいたメールにリプライします。

⇒先ほど私がいただいたメールに私はリプライします。

2. 4 翻訳品質予測に基づく主語補完の抑制

前節で説明したルールを適用したところ、過剰に主語が補完され、かえって翻訳品質が低下するケースが、主語補完が行われた文の約2割を占めた。そこで、以下の方法によって主語補完を抑制した。

(1) 翻訳品質予測に基づく抑制

形態素解析や意味構造解析の誤りにより原文自体の翻訳品質が期待できない場合は、誤った解析結果に基づいて主語補完を行うことでかえって分かり難い訳文になる恐れがある。そこで、そもそもの翻訳品質が期待できない文の主語補完を回避する。翻訳品質予測には、折り返し翻訳(日英翻訳ののち英日翻訳を行って日本語に戻す)から得られる、入力文と折り返し翻訳文の意味構造類似度スコアを使用する。意味構造類似度スコアの算出方法は先行研究^[1]を参照していただきたい。

意味構造類似度スコアは0.0~1.0の範囲の値をとり、数値が大きいほど入力文と折り返し翻訳文が類似している、つまり、翻訳品質が高いと予測される。いくつかの閾値を検討した結果、今回は意味構造類似度スコアが「0.3未満」場合には主語補完を行わない設定とした。

(2) 能動態・受動態の差異のみの場合の抑制

機械翻訳では、日本語の主語省略により能動態での翻訳が困難な場合に、受動態で翻訳するケースが多い。英語の受動態は、動作主が分からない場合や、動作主を曖昧にしたい場合などに一般的に使われる表現である。そこで受動態で翻訳できるケースは、無理に主語補完を行わないこととした。

2. 5 主語補完の例

主語補完の処理例とその翻訳結果を示す。翻訳には弊社の機械翻訳エンジン^[2]を使用した。日本語文字列中の下線部が補完された文字列である。従来の補完なしの場合には、不確定な主語を「it」で訳出していたが、主語補完により正しく「I」や「you」が使われている。

例1

■ 補完なし

はい、さっき、実機で確認しました

Yes, it confirmed it with a real machine a

little while ago.

■ 補完あり

はい、さっき、実機で私は確認しました

Yes, I confirmed it with a real machine a little while ago.

例2

■ 補完なし

どのように削減するかの相談については、こちらに連絡してください。

Please contact the consultation how it reduces here.

■ 補完あり

どのようにあなたは削減するかの相談については、こちらに連絡してください。

Please contact the consultation how you reduce here.

3 評価

3.1 評価方法

■ 評価尺度

弊社独自のシンプルな評価尺度を用意して評価を行った。評価者は、日本語文とその英訳文の組を参照し、英訳文の1文ごとに、以下の5個の設問に回答する。

- ① 理解できる (Yes/No)
- ② 大意が日本語文と一致している (Yes/No)
- ③ 文法が正しい (Yes/No)
- ④ 自然な英語である (Yes/No)
- ⑤ ネイティブと同等である (Yes/No)

以上の回答から、図3のフローによって0～5の6段階の評価点を求めた。

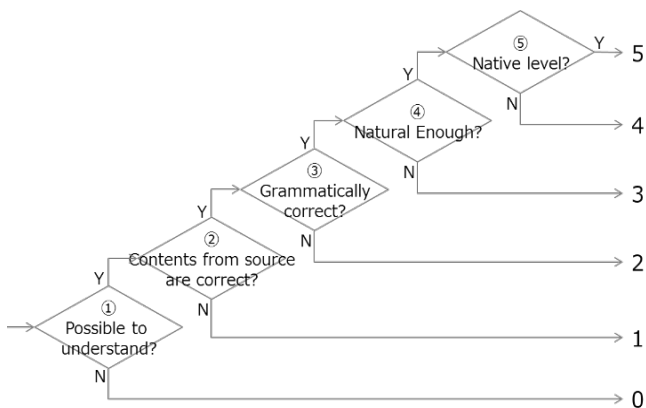


図3 評価点算出のフロー

■ 評価対象

以下の4種類の条件で生成した訳文を評価対象とする。

- (1) 補完なし
- (2) 補完あり
2.2, 2.3節のルールにより主語補完を行った場合
- (3) 補完あり (抑制あり)
(2)に対し、2.4節の主語補完の抑制を行った場合
- (4) 手修正
(2)に対して、主語補完の誤りを著者がマニュアルで修正した場合

LyncIM 翻訳に適用する際には、自動補完したテキストを発信者が確認、修正したのちに翻訳して送信する運用のため、「手修正」も評価対象とした。

■ 評価テキスト

社内 SNS での、日本語話者と外国語話者との日本語によるインスタントメッセージのログから抽出したテキストを評価に利用した。日本語として不自然なメッセージや、外国語のメッセージは評価データから除外した。メッセージの1行を1データとした。文の途中で区切られた節程度の長さのメッセージも含まれる。1行のメッセージに複数の文が含まれている場合には、文に分割してそれぞれを1個の評価データとした。結果として評価データ数は以下の個数になった。

- ・クローズデータ： 108 個
- ・オープンデータ： 99 個

ここでクローズデータとは、主語推定や主語補完のルールを検討する際に参考にしたデータセットであり、オープンデータとはそれとは別に選出したデータセットである。1データあたりの文字数は、日本語テキストで平均20文字である。評価データの例を以下に示す。

高木さん、今日のテストは終わりました。
どれくらいやりましたか？
今は6回テストしました
明日も1回テストします
合計10回くらいですけど、19日までにファームダウンも含めて完了しそうですか？
OK です
順調ですね。

■ 評価者

評価者は、翻訳を業務としている英語ネイティブ話者3名である。

3. 2 評価結果

図4に評価点平均を示す。また、評価点2以上が実用に使える訳文と考え、その割合を図5に示した。図4によれば、クローズデータ、オープンデータともに、補完なし、補完あり、補完あり（抑制あり）の順に平均点が上昇している。また、図5に示すように、評価点2以上の割合は、クローズデータでは従来の約58%から補完あり（抑制あり）の約63%に、オープンデータでは従来の約69%から補完あり（抑制あり）の約74%に増加した。

「手修正」では概ね良好な結果が出ているが、必要以上に主語補完を行うと、かえって翻訳結果に悪影響を及ぼす場合もあるので、この点は注意して利用してもらう必要がある。

以上の評価では、開発機能によって翻訳品質が改善する傾向にあった。しかし、個別の評価データを確認すると、補完しない場合よりも評価が低くなる「改悪」が少なからず存在した。個々のデータの評価値の変化を相対評価値により評価した。

相対評価値-2~+2は、評価者による評価点から以下のように換算した。ここでは、評価点2以上/未満の区別、つまり「実用に使える訳文」であるか否かを重視した。

相対評価値：

- 2： 評価点2以上が評価点1以下に下降
 - 1： 上記以外の評価点の下降
 - 0： 訳文に変化があるが評価点は変化せず
 - +1： 下記以外の評価点の上昇
 - +2： 評価点1以下が評価点2以上に上昇
- 訳文変化なし

図6に相対評価値の分布を示す。主語補完の抑制なしの場合には、クローズデータで約42%、オープンデータで約45%のメッセージで主語補完が発生しており、過剰に主語補完を行うことで改悪（-2または-1）が多数発生していると考えられる。これに対して、主語補完を抑制した場合には、クローズデータで約18%、オープンデータで約11%のメッセージでの主語補完に抑えられており、その結果、改悪が抑制されている。特に、評価に用いたオープンデータでは、改悪をゼロに抑えながら、7.6%の翻訳品質の改善（+2と+1の合計）であった。

4 日本語チェックサーバの開発

開発した主語補完機能を搭載した日本語チェックサーバのプロトタイプを開発した。なお、本サーバでは、先

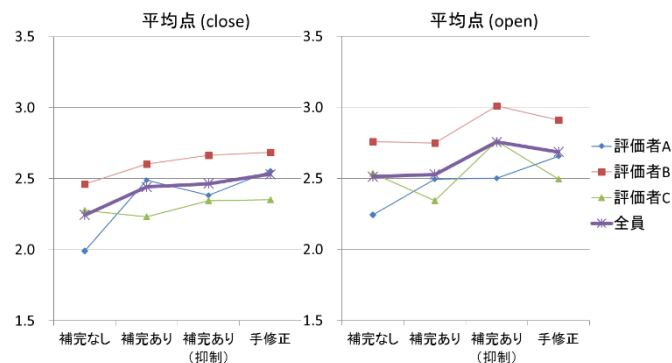


図4 評価点平均
(左：クローズデータ、右：オープンデータ)

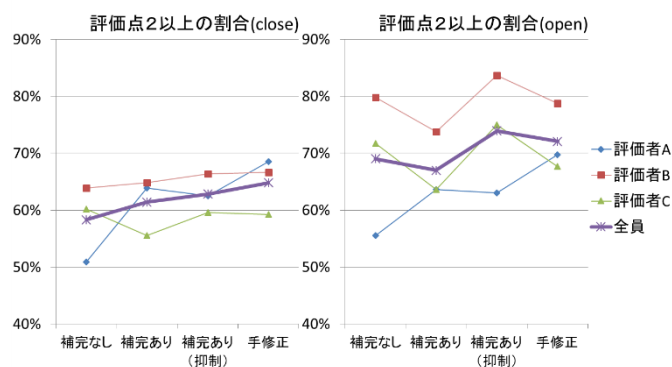


図5 評価点2以上の割合
(左：クローズデータ、右：オープンデータ)

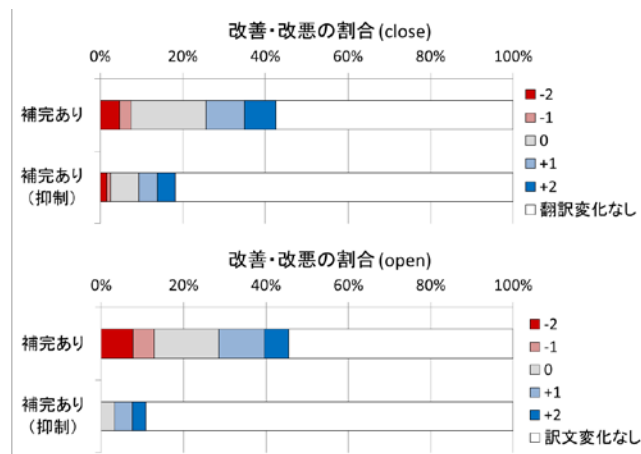


図6 従来からの改善・改悪の割合
(上：クローズデータ、下：オープンデータ)

に開発した語尾崩れ検出機能^[1]も同様に利用することができる。日本語チェックサーバの構成を図7に示す。

日本語チェックサーバでは、翻訳品質予測モジュールによって主語補完前の入力テキストの翻訳品質の予測を行い、翻訳品質予測値が閾値以上の場合のみ、主語補完モジュールを呼び、主語補完の処理を行う。各モジュール

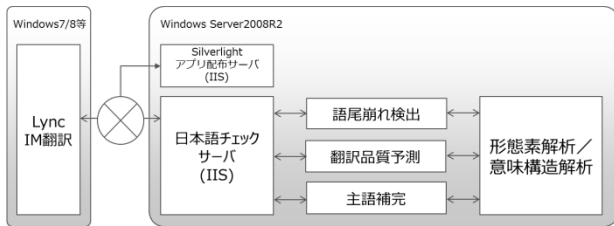


図7 日本語チェックサーバの構成

ルは、形態素解析／意味構造解析モジュールを呼び、必要な解析を実施する。

日本語チェック機能を利用する LyncIM 翻訳アプリケーションの GUI を図8に示す。画面右に翻訳元言語（本例では日本語）でのメッセージの履歴、画面左に翻訳先言語（本例では英語）でのメッセージの履歴が表示されている。また、右下部に3つのテキストボックスがあり、上から順に、テキスト入力ボックス、日本語チェック確認ボックス、訳文確認ボックスとなっている。テキスト入力ボックスに翻訳元言語でメッセージを入力し右のボタンを押下すると、日本語チェックが実行され、結果が日本語チェック確認ボックスに表示される。ここで、補完した主語は青字で表示される。チェック結果を確認し、必要に応じて修正を加えたのち、右ボタンを押下して翻訳を行う。さらに、訳文確認ボックスで翻訳結果を確認して送信する。

本開発によって、シンプルで分かり易いインターフェースで主語補完機能を利用できるようになった。

5 おわりに

インスタントメッセージ会話文の機械翻訳において、入力となる日本語文で省略された主語を補完する前処理により翻訳品質の改良を行った。簡易なルールを用いて主語推定および主語補完を行うことにより、社内 SNS のインスタントメッセージの 7.6%にあたる文で翻訳品質の改善を実現した。また、今回開発した主語補完機能を搭載した、日本語チェックサーバのプロトタイプを開発した。

今後は、開発機能を搭載した LyncIM 翻訳を業務に適用し、本機能の適用による会話効率の向上の程度を評価する予定である。また、口語表現に起因する翻訳品質劣化への対処などの残された課題に取り組む。

参考文献

- [1] 鄭 他「多言語リアルタイム会話における日本語動詞語尾崩れの検出」, 言語処理学会第21回年次大会, 2015. 3.
- [2] 富士通 英日・日英翻訳ソフト ATLAS
http://software.fujitsu.com/jp/atlas/
- [3] 大倉 他「複雑な文に対応した意味構造検索システムの開発」, 言語処理学会第20回年次大会, 2014. 3.
- [4] I. Goto, et al, “Overview of the patent machine translation task at the NTCIR-10 workshop”, 2013.



図8 LyncIM 翻訳の GUI の例