3ZB-3

翻訳前編集を用いた多段翻訳プロセスによるベトナム農業支援

北川 大輔† 中島 悠† 菱山 玲子‡ 稲葉 利江子† 林 冬惠† 石田 亨†

京都大学大学院 情報学研究科† 早稲田大学 理工学術院‡

1. はじめに

国際的な活動を行う NPO やプロジェクトでは多言語によるコミュニケーションが行われ，そこでは機械翻訳が利用されるケースも多い[1]．例えば，NPO パンゲア [1)]らは，2010 年より，日本の農業専門家の農業知識をベトナムの農家に伝達する YMC-Viet と呼ばれるプロジェクトを実施している．このプロジェクトでは，日本語で書かれた農業の専門知識が，機械翻訳と人間による翻訳により越語(越はベトナムの意)に翻訳される．日本語から越語への翻訳は直接行われるのではなく，日本語から英語，英語から越語へと翻訳されている．これは，日本語と英語，英語と越語を理解する人に比べて，日本語と越語を理解する人が少ないことに起因している．

このような多段の翻訳において機械翻訳を用いるだけでは翻訳精度が十分でないことが知られている[2]．人手による翻訳は機械翻訳よりも高い精度の翻訳ができるがコストが高い．そこで，本研究では，機械翻訳と翻訳前編集を組み合わせた翻訳プロセスを検討する．

このプロセスにおいて，人間は機械翻訳に入力させる文の編集（以下，前編集と呼ぶ）のみを担当する．入力文に対して前編集を行うことで，機械翻訳の出力文を改善できることが知られている[3]．一般の人間が担当可能な前編集により，機械翻訳結果の品質を改善することが狙いである．

2. 機械翻訳と翻訳前編集を用いた多段翻訳プロセス

機械翻訳の入力文が正確な文章でない場合などに，機械翻訳の出力文の品質が非常に低いことがある．多段翻訳プロセスの場合，ある翻訳の出力文は次の翻訳の入力文となるため，低い品質の翻訳出力文は，次の翻訳の品質も下げてしまうことになりうる．そこで，我々は，機械翻訳と翻訳前編集を併用した多段翻訳プロセスを検討する（図 1）．

このプロセスでは，機械翻訳への入力文を，適宜，人間が機械翻訳へ適した入力文へと書き換えることで，機械翻訳結果の品質を高めることを狙う．また，その書き換えルールを簡単なものとすることで，人手による翻訳よりも低コストな翻訳プロセスとする．

Multistage translation process with machine translation and pre-edit

Daisuke KITAGAWA†, Yuu NAKAJIMA†, Reiko HISHIYAMA‡, Rieko INABA†, Donghui LIN†, and Toru ISHIDA†

† Graduate School of Informatics, Kyoto University

‡ Faculty of Science and Engineering, Waseda University

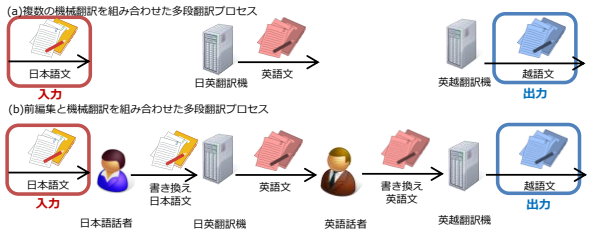

図 1: (a)既存の多段翻訳プロセスと(b)改善された多段翻訳プロセス

3. 多段翻訳プロセスの品質に関する実験

3.1. 設定

前編集と機械翻訳を組み合わせた多段プロセスを以下の通り実施した．

1. 日本語を母語とする大学生に書き換えルールに基づき日本語文を編集させる．書き換えルールは先行研究[4]から特に使用頻度の高そうなものを抜粋した．例として，(1)なるべく漢字を使う，(2) 略語は略さずに使う，(3) 文章を短く切る，(4) 係り受けを明確にする，(5) 複合名詞・複合動詞を避ける，などを使用した．
2. 前編集された日本語文を日英翻訳機に入力し，英文を取得する．
3. 英語を母語とする大学生に書き換えルールに基づき英文を編集させる．書き換えルールは，日本語用の書き換えルールのうち，英語に適用できそうなものを抜粋，修正し，英訳したものを使用した．例として，(1) 綴りの間違いを正す，(2) 省略形や頭字語を書き下す，(3) “who, whom that, which, whoever, whomever, whichever” などの関係代名詞を書き加える，(4) 複合名詞や複合動詞を避ける，(5) 文法の間違いを正す，などを使用した．
4. 前編集された英文を英越翻訳機に入力し，越文を取得する．

機械翻訳のみからなる多段プロセスとして，以下を実施した．

1. 日本語文を日英翻訳機に入力し英文を取得する．
2. 前ステップで得られた英文を英越翻訳機に入力し，越文を取得する．

3.2. 結果

機械翻訳の品質を評価する指標として Adequacy（適切さ）[5]を用いた．この指標は，原文の意味がどれだけ翻訳文に反映できているかを人間が評価し，

1) http://www.pangaean.org/

表 1: 各多段翻訳プロセスでの生成された越文と原文日本語間での Adequacy の比較

	機械翻訳のみの多段翻訳	前編集+機械翻訳による多段翻訳
Adequacy 5 の文章割合（5 以上の割合）	0.00% (0.00%)	3.70% (3.70%)
Adequacy 4 の文章割合（4 以上の割合）	9.26% (9.26%)	14.81% (18.52%)
Adequacy 3 の文章割合（3 以上の割合）	16.67% (25.93%)	35.19% (53.70%)
Adequacy 2 の文章割合（2 以上の割合）	53.70% (79.63%)	35.19% (88.89%)
Adequacy 1 の文章割合（1 以上の割合）	20.37% (100.00%)	11.11% (100.00%)

5 段階付けするものである．“How much of the meaning expressed in the gold-standard translation is also expressed in the target translation?”という問いに対し，それぞれ 1:None, 2:Little, 3:Much, 4:Most, 5:All で評価する．つまり，全く意味が伝わらないことを示す 1 から完全に意味が伝わったことを示す 5 までで表される．

実験には，2011 年の YMC-Viet にて日本人の農業専門家がコミュニケーションシステムに書き込んだ 54 文を用いた．機械翻訳のみで日英/英越翻訳を行った場合，前編集と機械翻訳を使って日英/英越翻訳を行った場合でそれぞれ前編集を行ったのと別で，越語が堪能な日本人被験者に Adequacy を評価してもらった．その結果を表 1 に示す．前編集と機械翻訳を組み合わせた多段翻訳プロセスと機械翻訳のみからなる多段翻訳プロセスとでは，日越翻訳の結果について，前者の品質がより高かったことがわかる．

前編集と機械翻訳を組み合わせた多段翻訳プロセスでは，個々の文章に着目すると 50%以上の文章において Adequacy が 3 以上となっている．Adequacy の 3 は「およその意味がわかる」という評価を表したものである．前編集はあるルールに従って与えられた文章をその文章が書かれた言語で書き換えるだけであり，翻訳に比べてコストの低い作業である．高いコストを支払い翻訳技能を持った人を使って高い翻訳精度を得るか，低いコストを支払い前編集と機械翻訳を用いて「およその意味がわかる」程度の翻訳精度を得るかがトレードオフとなる．

3.3. 考察

提案手法での翻訳において，多義語の翻訳がうまくいかず，英越機械翻訳に失敗し，Adequacy が低く評価された文章について考察する．「肥料やけが発生した時は、大量の水を与えて土から過剰な栄養分を流しましょう。」という日本語文を前編集し，日英機械翻訳した結果，“In cases of excessive fertilization, wash with water abundantly and remove excess fertilizer.”という英文を得た．これを英越機械翻訳した“Trong trường hợp thụ tinh quá mức, rửa lại bằng nước dồi dào và loại bỏ Phân bón dư thừa.”という文章について，付与された Adequacy は 2 であった．確認のため，越英機械翻訳で再度英語に戻すと，“In the case of excessive fertilization, rinse with abundant water and remove excess fertilizer.”となり，問題ないように見える．しかし，英語の“fertilization”には「土地の肥沃化」のほかに「受精」という意味も持つため，越語文で“thụ tinh（受精）”と訳出され，Adequacy が低く評価されてしまった．

4. おわりに

本研究では，機械翻訳と翻訳前編集を用いた多段翻訳プロセスを検討した．機械翻訳を単に連結するだけでは十分な翻訳品質が得られないため，機械翻訳に入力する文を適宜，人間が機械翻訳に適した文章に修正するステップを導入した．

前編集と機械翻訳を組み合わせた多段翻訳プロセスは，機械翻訳を組み合わせただけの多段翻訳プロセスに比べて，品質の高い翻訳文を生成することができた．前編集と機械翻訳を組み合わせた多段翻訳プロセスにより，半分以上の文章において「およその意味がわかる」程度の品質の翻訳文を得ることが出来たが，同時に多義語翻訳の課題も見つかった．

日越翻訳のように，アジア言語間で翻訳を行う際，英語を介在させる手法がよく用いられる．その際には今回示したように，翻訳元言語と英語を書き換えることで品質が上がる．書き換え作業はモノリンガルでも可能であり，英語の書き換え作業であれば，全世界で行える人が多い．今後は書き換え作業にクラウドソーシングを用いて，書き換え作業の金銭的，時間的コストを減らすことができないかを調べたい．

謝辞

この研究は科学研究費基盤研究(S)（24220002, 平成 24 年度～28 年度）の助成を受けた．

参考文献

[1] Ishida, T. ed., *The Language Grid*, Springer (2011).
[2] Kita, K., Takasaki, T., et al.: Case Study on Analyzing Multi-Language Knowledge Communication, *International Conference on Culture and Computing* (2012).
[3] 宮部真衣, 吉野孝, 重信智宏：折返し翻訳を用いた翻訳リペアの効果，電子情報通信学会論文誌. D, Vol. 90, No. 12, pp. 3141-3150 (2007).
[4] 山下直美, 坂本知子, 野村早恵子, 石田亨, 林良彦, 小倉健太郎, 井佐原均：機械翻訳へのユーザの適応と書き換えへの教示効果に関する分析，情報処理学会論文誌, Vol. 47, No. 4, pp. 1276-1286 (2006).
[5] Walker K., Bamba M., Miller D., Ma X., Cieri C., and Doddington G., Multiple-Translation Arabic (MTA) Part 1, Linguistic Data Consortium (LDC) (2003).