2C−9

ユーザ協調型日本語生成システム

熊野 明　野上 宏康　吉村裕美子　天野真家
(株)東芝 総合研究所

1. はじめに

英日機械翻訳システムの性能向上は，ハードウェアの進歩による処理速度の高速化で実現されている．しかし，システムの出力である訳文をユーザの希望に近付けるためには，辞書や文法に記述されている知識を拡張することによって，より正確な翻訳を行う必要がある．

しかし，ユーザの要求は多様である．その内容が翻訳処理の立場から些細なことでも，その要求が実現されればユーザにとって質の良いものになる．そのためには個々のユーザに対応した環境が必要である．これに対応するために，従来はユーザ辞書の構築を行ってきた．しかし，ユーザ辞書構築による効果が現れるのは，辞書に登録した語が直接関与する場合のみであり，翻訳文書全体に対する割合は数パーセントに過ぎない．その結果，出力された訳文が要求を満たさないものである場合，後編集作業が大きくなってしまう．日本人が英日機械翻訳システムを使用する場合，オペレータの作業の大部分は後編集作業であり，この作業を軽減することは機械翻訳全体の効率を上げることにつながる．

ユーザの意向を反映させる手段として，対話的な学習も考えられる[1]．しかし，これは一般的に手間がかかる．機械翻訳処理の主流は「一括翻訳」による大量翻訳である．そこで，翻訳処理前にユーザの希望をパラメータとして入力し，必要に応じてその値を参照して，すなわちユーザの意向と協調して訳文生成処理を行うことにした．

2. 対象

翻訳結果の訳文としてゆれのない部分は，ユーザ協調の対象にはなりえない．対象となるのはあくまで，「決めかねる」部分，すなわちfloatingな部分である．それは以下の2種類に分けることができる．

ひとつはhow to sayに関わるものであり，訳文候補が複数あり，いずれも誤りではない場合である．日本語の区切り記号に句読点を用いるかカンマ・ピリオドを用いるかは，単にユーザの好みの違いであり，一方が正しく一方が誤りとは言えない要素である．カタカナの複合語を表記する際，構成語の間を中黒(・)で区切るか，連続して表現するかの違いは，ユーザの文章基準に左右されるものである．

その他，英語の命令文に対して日本語でどういう文末表現に生成するか，英語の受動態の文に対して日本語で受動態で生成するか能動態で生成するかの選択などがこの種類に含まれる．

もうひとつはむしろwhat to sayに関わるものであり，機械翻訳の訳文生成を助けるためのものである．例えば分詞構文には，理由，同時進行など複数の解釈の候補があり，翻訳処理の上でそのいずれかに解釈する必要がある．しかし大規模な知識辞書をもってしても解決できない例はまだ多い．このような場合は，「わからなければこの解釈にする」という基準が必要である．

英語の法助動詞には，根源的意味と陳述緩和的意味をもつものがある[2]．例えばmayは，許可を示す根源的意味と可能性を示す陳述緩和的意味をもっている．翻訳処理の過程でそのいずれの意味かを判断し，相当する日本語のモーダル表現を生成する必要があるが，必ずしも解釈をひとつに絞りこめない場合がある．この場合も「わからない場合はこの表現で出力する」という基準が必要である．

3. 実現手段

3.1 ユーザ協調変数

翻訳処理に先立って，ユーザ協調変数をセットする．このユーザ協調変数は，訳文に対するユーザの希望を項目ごとにパラメータ化したもので，専用のツールを用いて入力する．システムはあらかじめすべての項目に対してデフォルト値をもっているので，設定作業を行わなければすべて標準設定値で処理が行われる．

今回英日機械翻訳システムの生成機構に実現したユーザ協調変数の例を表1に示す。

変数の意味	変数のとりうる値
受動態	通常受動態 / 能動態変換
主語you	省略する / 省略しない
分詞構文	して / するので / しながら
文体	常体 / 敬体
区切り記号	句読点 / ピリオド・カンマ
命令文	しなさい / してください / すること
助動詞may	してもよい / するかもしれない
助動詞must	に違いない / なければならない
助動詞should	べきである / したほうがよい
省略可能な送りがな	出力する / 省略する

表1 ユーザ協調変数

User-cooperative Japanese Sentence Generation

Akira KUMANO, Hiroyasu NOGAMI, Yumiko YOSHIMURA and Shin-ya AMANO
TOSHIBA Corporation

ユーザ協調変数の値は翻訳処理中に各文法から参照される．文法をユーザ協調変数の値の数だけ用意するのではなく，単一の文法であらゆる変数値に対応した処理ができるようにした．つまり，文法にはユーザ協調変数を参照する記述が含まれており，参照値をもとにして各処理の制御を行う．

3.2 処理の流れ

ユーザ協調変数を用いた日本語生成処理は，以下の規則を順次適用して行われる．

[規則1]
翻訳処理による解釈を必要としないものは，ユーザ協調変数の指示通り生成する．
[規則2]
翻訳処理の過程で解釈ができ，適切な生成手段が判断できた場合は，その結果に従った生成処理を行う．ここで「解釈ができた」「適切な生成手段が判断できた」という基準は，解釈または生成の確信度が十分高いことを示している．この時点で確信度が十分高くないものは，解釈を保留している．
[規則3]
解釈または生成の確信度の低い解釈に対しては，ユーザ協調変数を参照する．ユーザによる変数設定があれば，その指示に従って生成する．
[規則4]
ユーザによる変数設定がなければ，システムが用意したデフォルト値に従って生成する．この値には，一般的なマニュアル翻訳に適した設定値が用意してある．

4. 生成処理例

例えば，日本語の区切り記号を句読点にするかカンマ・ピリオドにするかは，翻訳処理の解釈を必要としない．[規則1]により，ユーザ協調変数の指示に従って出力する．命令文の日本語文末表現なども，解釈の問題ではなく生成だけの問題であり，[規則1]に適合する．

受動態英文を日本語で生成する際の表現は，受動態でも能動態でも意味する内容には本質的に差はないが，訳文の理解し易さに差が生じることが多い．

次の受動態の英文(a)に対して，受動態の訳文と能動態の訳文例を示す．

(a) The remote files cannot be removed.
(a-1) リモートファイルは削除されることができない。
(a-2) リモートファイルは削除することができない。

能動態に変換した(a-2)が，日本語として明らかに自然な文である．これは助動詞canの意味が関与しており，日本語における可能表現の性質によるものである．この場合は[規則2]に従い，ユーザ協調変数が「受動態で生成する」を指示していても，生成の確信度の高い「能動態変換」で訳文生成を行う．

助動詞を含んだ次の英文(b)に対して，助動詞の2通りの解釈による訳文例を示す．

(b) The program name may appear on the list.
(b-1) プログラム名はリストに現れてもよい。
(b-2) プログラム名はリストに現れるかもしれない。

自動詞 appearの意味と無生物主語の性質から，(b-1)の意味は不自然で，(b-2)の訳文が正しいと思われる．理由は，appearは意志性をもたないので，その動作を許可することは意味をもたないし，無生物に対して許可を与えることも意味がないからである．この場合も[規則2]に従い，ユーザ協調変数が許可を示す根源的意味の「してもよい」を指示していても，生成の確信度の高い可能性を示す陳述緩和的意味の「するかもしれない」の表現を用いる．このように，特徴的な性質をもつ語に関しては，語彙的な例外処理が可能である．

5. 結果

ユーザ協調変数を生成処理に導入することにより，10種類以上の項目に対して多様な日本語訳文を実現することができた．結果的にユーザの希望する翻訳結果に近付くことができる多くの出力を得た．

ユーザ協調変数の利用には，確信度を考慮した．翻訳システムが確信をもっている部分はその情報を優先して生かし，確信度の低い部分に対してユーザ協調変数を参照させた．その結果，システムの知識とユーザの嗜好が協調する訳文生成を実現することができた．

6. おわりに

翻訳処理の前に，訳文に対するユーザの嗜好を補助的に入力することによって，文書全体にわたって従来よりユーザにとって質のよい翻訳結果を得ることができた．これまで原則として一通りの結果しか出力しなかった機械翻訳の日本語生成システムが，ユーザと協調して訳文を出力するものになった．

その結果，後編集に要する時間は確実に軽減され，翻訳システム全体のパフォーマンスが向上した．

参考文献

[1] 増山顕成，他: 機械翻訳システムにおけるチューニング機構について，情報処理学会第39回全国大会，3G-6 (1989).
[2] 熊野明，他: 英日機械翻訳システムの訳文生成について，情報処理学会自然言語処理研究会資料，NL40-6 (1983).