

日中機械翻訳における日本語可能表現の翻訳方法について

徐 金安 長田 誠也 山端 潔
NEC 共通基盤ソフトウェア研究所

{j-xu@ap, s-osada@cd, k-yamabana@ct}.jp.nec.com

1. はじめに

我々は、旅行会話向け日中翻訳システム[1]を開発しており、日本語可能表現をいかに翻訳するかの課題があった。中国語で可能を表す方法は、大別して能願動詞(助動詞)を用いたやり方と可能補語で表現するやり方の二つある。日本語では、可能は専ら助動詞または助動詞相当句により表現されるため、日中翻訳では、日本語の可能助動詞を中国語の能願動詞に対応させて翻訳すると、翻訳規則が単純化できる。しかし、可能補語でしか表せない意味が多数存在するため、能願動詞を使うだけでは、正しく中国語へ翻訳することができないケースが多数ある。また、中国語可能表現以外の言語表現でしか日本語可能表現を正しく翻訳できないケースも少なくない。本稿では、日本語可能表現を、能願動詞、可能補語及び可能でない表現へと訳し分けできる翻訳手法を考案し、日中旅行会話コーパスを対象とした予備的な評価結果について述べる。

2. 課題と翻訳方法

日中翻訳における日本語可能表現の翻訳は、次の課題が挙げられる。

①. 日本語可能表現文に対して、結果補語、方向補語などの可能でないしかでしか正しく翻訳できない部分の文法現象と、その潜在的な規則を解明することが必要である。

A. 日: チケットが確認できた。 中: 票确认好了。

説明: 「确认好了」の意味が「確認済み」で、「動詞+結果補語」のため、能願動詞及び可能補語が共に使えなくなる。

②. 能願動詞と可能補語との意味上での共通・差異部分を求め、使い分けの翻訳規則を解明することが必要である。

B. 日: 日本円で払えますか

中: 能付日元吗?(能願動詞) 中: 日元付得了吗?(可能補語)

C. 日: この本を読めません。

中: 不能读这本书。 能願動詞(誤訳)

中: 读不懂这本书。 可能補語(正解)

説明: B文の二つの訳文は意味が同じだが、C文の「不能读这本书」は、「この本を読んではいけない」との禁止的な意味を示す。C文には、能願動詞が使えなくなる。

③. 用言が同時に複数の可能補語を持つ場合は、その訳し分けの翻訳規則を解明することが必要である。

D. 日: 高く食べられません。 中: 太贵吃不起。

日: 多くて食べられません。 中: 太多吃不完。

説明: 「吃不起」が「値段が高い」または「お金がない」という意味を含む。「吃不完」が「食べきれない」と相当、「量が多い」ことを示す。例のように、「食べられる」の訳し分けが必要になる。

本稿では、前記課題の①と②を中心に、アスペクト計算結果、否定概念、文スタイル、依頼勧誘、尊敬表現、仮定について分析した結果を用いた翻訳方法を報告する。

3. 日中可能表現

日本語可能表現は、「れる(られる)」と「ことができる(できる)」の二種類の構成に大別される。中国語可能表現は、能願動詞と可能補語が存在する。能願動詞は一般的に「能、可以、会」などを用いる。主に以下の意味を示す。

- ① 主観条件や客観条件による判断結果を示す。
- ② 許可を示す。
- ③ 可能性や見込みなどを示す。

可能補語は、一般的にA型、B型、C型に分類される[2][3]。本稿では、離合詞を含む可能補語を独立させてD型と定義する。4種の可能補語パターンを表1に示す。

表1 中国語可能補語の種類

可能補語型	可能補語パターン
A型肯定	訳語+得+可能補語要素 ¹
A型否定	訳語+不+可能補語要素
B型肯定	訳語+得了
B型否定	訳語+不了
C型肯定	訳語+得
C型否定	訳語+不得
D型肯定	訳語+得了+離合詞要素 ²
D型否定	訳語+不了+離合詞要素

以下、文献[2]に沿って可能補語について説明する。

A型可能補語は、動詞と結果補語あるいは方向補語の間「得/不」を入れたものになる。「得」は肯定形で、「不」は否定形である。

A型可能補語は、主に、ある事柄が実現できるかできないかの意味を表す。

(1) 日: 英語の新聞が読める。 中: 读得懂英语报纸。

(2) 日: 英語の新聞が読めない。 中: 读不懂英语报纸。

説明: 能願動詞では、「不能读」→「読んではいけない」。

中国語動詞がA型可能補語と目的語を同時に持つ際に、目的語の位置は一般的に以下の二種類がある。

第一種目は、可能補語が結果補語あるいは単純方向補語で作られている時には、目的語は可能補語の後ろに置かれる。

第二種目は、可能補語が複合方向補語から作られている場合には、多くの場合、目的語はもとの複合方向補語の間に置かれる。

B型可能補語は、中国語動詞あるいは形容詞の後ろに「得了/不了」を利用したものとなる。「得了」は肯定形で、「不了」は

¹ 結果補語、方向補語が可能補語要素となりうる。また、可能補語要素は動詞ごとに決まる。

² 離合詞とは「動詞+目的語」型の動詞で、二つの形態素に分離することが出来るもの。本稿では、「目的語」の部分に離合詞要素と定義する。例えば、「踊る」の訳語が「跳舞」で、「踊れる」の訳は「跳得了舞」となる。「舞」を示す目的語の部分を離合詞要素という。離合詞の二つの形態素が常に共起していることが特徴で、欠落がある場合は意味が変わってしまう。

否定形である。

(3) 日：一人で動ける。 中：一个人动得了。

(4) 日：一人で動けない。 中：一个人动不了。

説明：能願動詞では、「不能動」→「動いてはいけない」。

B型可能補語は主に以下の二通りの意味を表す。

①.ある事をやり終えることが可能であるかを表す。このような意味を表す時には、前の方には多、少、大、小あるいは具体的な数量などの全体条件がよくあり、後ろの方には人数、時間などの制限を表す言葉がよく入る。

②.動作や事柄それ自身が発生あるいは実現できるかどうかを表す。

B型可能補語と目的語と同時に持つ場合は、一般的に補語の後ろに置くことができる。

C型可能補語は、動詞あるいは形容詞の後ろに「得/不得」がつくものになる。「得」は肯定形で、「不得」は否定形である。

(5) 日：そこへいける。 中：那里去得。

(6) 日：そこへいけない。 中：那里去不得。

説明：能願動詞では、「不能去」→「行ってはいけない」。

C型可能補語は主に以下の二通りの意味を表す。

①.「客観的な条件によってあることが起こることを許すか許さないか」という意味を表す。

②.「ある事が実現できるかどうか」を表す。

C型可能補語と目的語と同時に持つ場合は、目的語を述語の前または補語の後ろに置くことができる。

D型可能補語は、中国語動詞あるいは形容詞が離合詞である場合、その離合詞の間に、「得了/不了」を利用したものとなる。「得了」は肯定形で、「不了」は否定形である。

(7) 日：学生は先生に手伝える。 中：学生给老师帮得了忙。

(8) 日：学生は先生に手伝えない。 中：学生给老师帮不了忙。

説明：能願動詞では、「不能帮忙」→「手伝ってはいけない」。

D型可能補語は、意味的にB型可能補語と同様で、B型可能補語の特別な一族とも言える。機械翻訳上では、離合詞を特別扱いにする必要があるため、D型可能補語を独立させる。D型可能補語の離合詞要素の後に、さらに他の目的語と同時に持つ現象がほとんど見られない。

また、中国語の可能でない言語表現で日本語可能表現文を翻訳しなければならない現象が多数ある。主に、結果補語、方向補語などを用いて翻訳することになる。

(9) 日：資料が用意できた。 中：资料准备好了。

(10) 日：内容を理解できている。 中：内容理解了。

4. 可能表現の分析と翻訳

本節では、まず中国語能願動詞と可能補語の互換性について述べる。次に、日本語可能表現文のアスペクト計算結果、否定概念の有無、文スタイル(疑問文と平叙文)、依頼勧誘と尊敬、仮定の表現を中心に分析する。

4.1. 能願動詞と可能補語との互換性

能願動詞と可能補語との関係は文献[2]に沿って纏める。

A型可能補語：

① ある事柄が実現可能であるかを表す時の肯定形では、意味的に大体「能(可以)+動詞+結果補語/方向補語」の文型を代わりに使うことができる。否定形では、単純に置き換えることはできない。

② ある事柄を許可するかどうかを表す可能形では、「能(可以)」を利用する。

B型とD型可能補語：

③ ある事をやり終える可能であるかを表す時の肯定形では、一般的に「能(可以)+動詞」で表すことができる。否定形では、単純に置き換えることができない。

④ なんらかの事情によりある行動が許されるかどうかを表す時に、能願動詞を用いる。

C型可能補語：

⑤ 客観的な条件によってあることが起こることを許すか許さないかを表す時、一般的に「不能」と置き換えられる。

⑥ ある事が実現できるかどうかを表す際に、「能(可以)」で置き換えられる。

しかし、こうした記述内容のままで機械翻訳に用いられるのが困難であるため、以下、文のアスペクト計算結果や否定概念等の面から分析した結果を纏める。

4.2. アスペクト計算結果に関する分析と翻訳

日本語可能表現文のアスペクト計算結果は、原形、状態形、進行形、完了形の4種類を用いる[4][5]。

原形：彼はよく林檎を食べる。

完了形：彼は林檎を食べた。

状態形：彼を知っている。

進行形：外は雨が降っている。

また、日本語「た/る/ている/テイタ」と対応する中国語アスペクト助字の生成と副詞の補間は文献[6]の方法が適用できる。日本語可能表現の用言に対する中国語での訳し方について、以下に述べる。

旅行会話コーパスに存在している可能表現文のアスペクト計算結果は、90%以上が原形であったため、対象文を変形して、各種のアスペクトを持つ文を作り、可能補語生成への影響について分析した。

(11) ここで観光バスのチケットを買えますか。

中：在这里买得到观光汽车的票吗？

これを以下のように変形した。

(12) ここで観光バスのチケットを買えていますか。

(13) ここで観光バスのチケットを買えましたか。

(14) ここで観光バスのチケットを買えていましたか。

(12)~(14)の三つの文の中国語訳は次のようになる。

中：在这里买到了观光汽车的票吗？

評価対象文が否定概念を含む場合も同様に変形して分析を行う。

(15) 日：新聞を読めません。(原形) 中：读不懂报纸。

これを以下のように変形した。

(16) 新聞を読めていません。 中：没能读懂报纸。

(17) 新聞を読めませんでした。 中：没能读懂报纸。

(18) 新聞を読めていませんでした。 中：没能读懂报纸。

分析した結果を以下のように纏めることができる。

①. 可能表現文のアスペクト計算結果が「原形」の時、表1に示す可能補語パターンで翻訳する。

②. 可能表現文のアスペクト計算結果が「状態形」、「進行形」または「完了形」の時、表2を用いて翻訳する。

表2では、肯定文の場合、能願動詞も可能補語も生成しな

くてよい¹。否定文の場合、「没能+動詞」の共通パターンが存在する。以下、日本語可能表現文のアスペクト計算結果が「原形」でない時、可能補語が生成しなくてもよい理由を分析する。

表2 「状態形・進行形・完了形」の時の翻訳パターン

補語型	出力パターン
A型肯定	訳語+可能補語要素
A型否定	没能+訳語+可能補語要素
B型肯定	訳語
B型否定	没能+訳語
C型肯定	訳語
C型否定	没能+訳語
D型肯定	訳語+離合詞要素
D型否定	没能+訳語+離合詞要素

中国語のA型可能補語は、主に「ある事柄が実現できるかどうか」の意味を表す。日本語可能表現のアスペクト結果が「原形」の時、この意味を表すことができるが、アスペクト結果が「原形」でない時、つまり、日本語可能表現の中に「た/ている/ていた」が存在する場合、用言の意味が主に「ある事実や結果がどうなったか」の意味を表す。この意味に対応する中国語表現は「動詞+結果補語/方向補語」等となるため、可能補語との対応関係が取れなくなることが原因であると考えられる。

BCD型の可能補語も同様に、日本語可能表現の中に「た/ている/ていた」が存在する場合、文の意味は可能補語で表せる意味との対応が取れないのが原因であると思われる。

また、表2に示す翻訳パターンは、下記の例文に示すように、疑問文も平叙文も同様に用いることができる。

- (19) リフト券は買える。中：买得到缆车票。「原形」
 平叙・肯定文：「状態形・進行形・完了形」
 日：リフト券は買え(た/ている/ていた)。
 中：买到了缆车票。
 肯定・疑問文：「状態形・進行形・完了形」
 日：リフト券は買え(た/ている/ていた)か。
 中：买到了缆车票吗？
 平叙・否定文：「状態形・進行形・完了形」
 日：リフト券は買え(なかった/ていない/ていなかった)。
 中：没能买到缆车票。
 否定・疑問文：「状態形・進行形・完了形」
 日：リフト券は買え(なかった/ていない/ていなかった)か。
 中：没能买到缆车票吗？

4.3. 否定概念、文スタイルに関する分析と翻訳

対象文に対して否定概念の有無と疑問文または平叙文との組み合わせにより、能願動詞と可能補語による可能表現を考察した。以下、例を用いて分析する。

- (20) タクシーはどこで拾えますか？ 肯定・疑問
 中：哪里能叫到出租车吗？(能願動詞)
 中：哪里叫得到出租车吗？(A型可能補語)
 分析：同様に自然で正しい。
 (21) タクシーはどこで拾えませんか？ 否定・疑問
 中：哪里不能叫到出租车吗？(能願動詞)
 中：哪里叫不到出租车吗？(A型可能補語)
 分析：「不能叫到」がやや不自然だが、意味が通じる。
 (22) 宿泊客は無料でプールを使えますか？ 肯定・疑問

- 中：住宿客人能免费使用游泳池吗？(能願動詞)
 中：住宿客人免费使用得了游泳池吗？(B型可能補語)
 分析：同様に自然で正しい。

- (23) 宿泊客は無料でプールを使えませんか？ 否定・疑問
 中：住宿客人不能免费使用游泳池吗？(能願動詞)
 中：住宿客人免费使用不了游泳池吗？(B型可能補語)
 分析：「使用不了」がやや不自然だが、意味が通じる。
 例文に示すように、疑問文可能表現は、否定概念の有無と関係せずに、一般的に能願動詞で訳せると考えられる。ただし、一部の動詞は、結果補語や方向補語の補間が必要になる。

- (24) この電話ではカードは使えます。 肯定・平叙
 中：这个电话能用卡。(能願動詞)
 中：这个电话用得了卡。(B型可能補語)
 分析：同様に自然で正しい。

肯定平叙文の場合には、能願動詞と可能補語のどちらでも翻訳することができると考えられるが、用言によって表す意味のずれが多少ある。

- (25) この電話ではカードは使えません。 否定・平叙
 中：这个电话不能用卡。(能願動詞)
 中：这个电话用不了卡。(B型可能補語)
 分析：同様に自然で正しい。

- (26) 夜は眠れません。 否定・平叙
 中：晚上不能睡着觉。(能願動詞)
 中：晚上睡不着觉。(D型可能補語)
 分析：「不能睡着觉」→「眠ってはいけない」の誤訳である。
 否定平叙文の場合、許可や禁止的な意味を表す文は、能願動詞で翻訳できると考えられるが、その他の意味を表す時に、可能補語を利用するのが原則であると思われる。
 また、アスペクト結果が「原形」意外の場合、表2が使える。

4.4. 依頼勧誘と尊敬表現に関する分析と翻訳

日本語可能表現のうち、依頼勧誘と尊敬表現を示すパターンがしばしばある。以下、例文で分析する。

- (27) 説明していただけませんか？
 中：能不能说明一下？(能願動詞)
 中：说明得了吗？(B型可能補語)
 分析：「能不能」は丁寧、自然、正しい。
 (28) プレゼント用に包んでもらえますか？
 中：能不能包装成礼品用？(能願動詞)
 中：礼品用包装得了吗？(B型可能補語)
 分析：「能不能」は丁寧、自然、正しい。
 (29) 教会のバザーに寄付をお願いできませんか？
 中：能不能给教会的义卖捐钱？(能願動詞)
 中：给教会的义卖捐得了钱吗？(B型可能補語)
 分析：「能不能」は丁寧、自然、正しい。
 (30) メッセージをお伝え願えますか？？
 中：能不能转告留言？(能願動詞)
 中：转告得了留言吗？(B型可能補語)
 (31) ウエストを約二センチおつめできます。
 中：腰围能放出大约二公分。(能願動詞)
 中：腰围放得出大约二公分。(A型可能補語)
 分析：丁寧さが同様、「能放出」の方が自然である。
 例文に示すように、「て(で)もらえるか」、「て(で)いただけるか」などの「て接続」形表現文、「お+動詞連用形+願えるか」、「お(ご)+動詞連用形+できる」などの「お(ご)+動詞連用形

¹副詞などの影響で、訳語の前に「能」が付くこともできる。

＋可能助動詞」型の文、「お(ご)名詞＋受給動詞の可能形」等の可能表現文の翻訳規則は、次に示す。

① 疑問文の場合、能願動詞の「能」または「可以」で翻訳すればよいが、丁寧さは、「能不能」、「可不可以」で表現する。

② 平叙文の場合、能願動詞の「能」または「可以」で翻訳できるが、「能不能」、「可不可以」の表現が必要としなくなる。

また、アスペクト結果が「状態形」、「進行形」、「完了形」の際、疑問文も平叙文も表 2 を利用することができる。

4.5. 仮定に関する分析と翻訳

日本語尊敬表現文のうち、仮定を含む表現も多数ある。以下、例文を用いて、分析する。

(32) 診療所に行けなければ往診をお願いします。

中: 如果不能去诊所, 可以到住家问诊。(能願動詞)

中: 如果去不了诊所, 可以到住家问诊。(B 型可能補語)

分析: 同様に自然で正しい。

(33) 五時までにお持ちいただければ、間に合います。

中: 如果到五点能拿来的话, 来得及。(能願動詞)

中: 如果到五点拿得来的话, 来得及。(A 型可能補語)

分析: 同様に自然で正しい。

(34) 十時半でしたらご用意できます。

中: 如果是十点半就能准备好。(能願動詞)

中: 如果是十点半就准备得下。(B 型可能補語)

分析: 「能准备好」が自然だが、結果補語「好」の補間が必須となる。

例文のように、「ば、たら、なら」等の仮定を含む可能表現文は、中国語が一般的に「如果～(就)能～」で訳せると思われる。

また、アスペクト結果が「状態形」、「進行形」、「完了形」の際、疑問文も平叙文も表 2 が使える。

5. 予備実験

日中翻訳システムの翻訳辞書に可能補語分類情報と可能補語要素を付与する必要があることと、予備実験評価の便利性を考慮し、予め、日中翻訳可能補語判定用DBを構築した。5 万 5 千文規模の旅行会話コーパスから 2785 個の動詞とサ変の対訳リストを抽出した。可能補語パターンが動詞依存のため、手作業でDBを作成した。表 3 はDBにある4種類の可能補語の用言数を示す。

表 3 各可能補語の用言数

種別	A型	B型	C型	D型
用言数	1227	1106	105	347

評価文は、旅行会話コーパスからランダム的に 260 文を抽出した。対象文の中に、仮定文が 16 文を含む(ば:8 文、たら:6 文、なら:2 文)。疑問文が 158 文、否定概念が 53 文、中に否定疑問文が 17 文、否定平叙文が 36 文であった。文のアスペクト計算結果は「原形」が 255 文、「完了形」が 5 文、状態形と進行形が 0 文であった。予備実験が次のように行った。

① 評価対象文の 260 文に対して、能願動詞「能、可以、会」で用いて全文を翻訳する。

② 提案手法に基づいて、能願動詞/可能補語/可能でない言語表現の訳し分けを行って全文を翻訳する。

③ 評価対象文の 260 文に対して、疑問文でない文を疑問文に変形して、①と②の手順で評価する。

④ 評価対象文の 260 文に対して、文アスペクト計算結果が

「原形」の文を「完了形」に変形して、①と②の手順で評価する。

⑤ 評価対象文の 260 文に対して、否定平叙文でない文を否定平叙文(46 文が変形できず)に変形して、①と②の手順で評価する。

考案した翻訳手法に基づく翻訳精度を手作業で評価した。表 4 は、前記実験方法①と②、表 5 は前記実験方法③～⑤による評価結果を示す。翻訳精度は、意味が通じる翻訳結果文の数/全評価対象文数*100%と定義する。改善文数は、意味の通じない誤訳から正解になる文の数を示す。改善率は、改善文数/全評価対象文数*100%と定義する。

表 4 実験結果 1

①の精度(%)	②の精度(%)	改善文数	改善率(%)
82.7	98.1	40	15.4

表 5 実験結果 2

方法	①の精度(%)	②の精度(%)	改善文数	改善率(%)
③	93.1	97.3	11	4.2
④	24.2	95.4	185	71.2
⑤	120/214	201/214	81	31.2

表 4 と表 5 での評価結果は、能願動詞「能、可以、会」などの訳しわけ問題と、同一用言が複数の可能補語を持つ場合、その可能補語の訳し分け問題がすでに解決できるとの仮定で得られたものである。可能補語の訳し分けは、用言意味分類情報、文脈情報などの取り入れが必要になると考えられる。

また、表 4 と表 5 での評価結果では、現在の翻訳規則では対処できない例文があった。例えば、副詞、用言の用法により可能補語ではなく、結果補語や方向補語で翻訳しなければいけないケースが存在する。

6. おわりに

本論文では、日中機械翻訳システムにおける日本語可能形から中国語の能願動詞による可能表現と、可能補語による可能表現と、可能でない表現との三つの部分に訳し分け可能な翻訳手法を考案した。また、考案手法に基づいて 260 文の旅行会話可能表現文に対して、手作業で評価を行った。評価の結果、考案手法の有効性を一定の程度で確認した。

今後の予定として、動詞の意味や用法によるモデルの精密化、文脈情報と詳細意味分類の導入による翻訳生成パターンの分析、同一用言が複数の可能補語パターンを持つ場合、その曖昧性の解消などが挙げられる。

参考文献:

- [1] 長田誠也, 徐金安, 山端潔: 旅行会話向け日中機械翻訳システムの開発, 言語処理学会第 14 回年次大会, 2008 年 3 月。
- [2] 劉月華等: 实用現代漢語語法(増訂本), 商務印書館, 2001, pp.582-594。
- [3] 守屋 宏則: やさしくわかる中国語文法の基礎, 株式会社東方書店, 1995 年 3 月 15 日発行, pp.192-208。
- [4] 寺村 秀夫: 日本語の文法, 国立国語研究所, 大蔵省印刷局, 昭和 56 年 1 月 10 日初版発行。
- [5] 金田一春彦: 日本語動詞のアスペクト, むぎ書房, 1976 年 5 月 10 日発行, 1983 年 5 月 31 日 4 刷。
- [6] 謝軍, 卜朝暉, 池田尚志: 日中機械翻訳におけるテンス・アスペクトの処理, 自然言語処理 Vol.10 No.4, pp. 177-200, 2003。