HOKKAIDO UNIVERSITY

Title	対訳コーパスに基づく学習型機械翻訳における省略可能情報を用いた部分対応学習の有効性
Author(s)	寺島, 涼; 越前谷, 博; 荒木, 健治
Citation	電子情報通信学会論文誌. D, 情報・システム, J93-D(3), 377-388
Issue Date	2010-03-01
Doc URL	https://hdl.handle.net/2115/52204
Rights	copyright©2010 電子情報通信学会
Type	journal article
File Information	DTRD93D-3_377-388.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

論　文

対訳コーパスに基づく学習型機械翻訳における省略可能情報を用いた部分対応学習の有効性

寺島　涼[†a)]　越前谷　博[††b)]　荒木　健治[†c)]

Effectiveness of the Learning Method for Partial Correspondence Based on Deletion Possibility in Learning-Type Machine Translation Using Parallel Corpus

Ryo TERASHIMA[†a)], Hiroshi ECHIZEN-YA[††b)], and Kenji ARAKI[†c)]

あらまし　対訳コーパスに基づく機械翻訳手法では，対訳文の原言語文と目的言語文間で句に相当する部分的な対応関係を決定することは有効である．そこで，本論文では省略可能情報を用いた部分対応学習を提案する．本手法では句レベルの部分対応を決定するための抽出ルールを，対訳文中の省略可能な部分に着目することにより自動獲得する．次いで，句に相当する部分の探索範囲を限定するための情報を有している抽出ルールを用いて，効率良く対訳文中の部分対応を決定する．本論文では，対訳コーパスから翻訳ルールを自動獲得することで翻訳を行う，学習型機械翻訳に本手法の適用を行った．性能評価実験の結果，本手法により，効率的かつ汎用的な翻訳ルールの獲得が可能となり，人手による正誤評価及び様々な自動評価において翻訳精度の向上が確認された．

キーワード　機械翻訳，学習，部分対応，省略可能，対訳コーパス

1. まえがき

近年，対訳コーパスに基づく機械翻訳手法の研究が盛んに行われている．そこでは，対訳文の原言語文と目的言語文間における，句に相当する部分的な対応関係の利用が有効である．部分的な対応関係を決定する手法としては，解析的アプローチ [1]〜[3] と統計的アプローチ [4]〜[7] の二つに大きく分類できる．解析的アプローチでは，品詞情報若しくは構文情報を用いることで部分の対応関係を決定する．特に，対応関係の決定が比較的容易な名詞句の部分対応を優先的に利用することは，対訳文からの翻訳知識の自動獲得において有効となる [3]．しかし，解析的アプローチの問題点としては，解析ツールが十分ではない言語には適用が困難となることが挙げられる．また，統計的アプローチでは，大規模な対訳コーパスの使用を前提に，統計情報に基づき部分の対応関係を決定する．そのため，データスパースな対訳コーパスでは，部分対応を効率良く決定することが困難となる．

そこで，本論文では，対訳文中の句に相当する部分の対応関係を効率良く決定する新たな学習手法として，省略可能情報を用いた部分対応学習を提案する．本手法では文中の省略可能な部分に着目することにより抽出ルールを自動獲得する．獲得された抽出ルールは，対訳文中の句に相当する部分の探索範囲を限定するための情報を有している．この抽出ルールを様々な対訳文に適用することにより，対訳文中の句レベルの部分対応を効率良く決定できる．その結果，解析ツールに強く依存することなく多言語への適用が可能となる．

本論文では，省略可能情報を用いた部分対応学習を，学習型機械翻訳 [8], [9] に適用する．本論文で用いる学習型機械翻訳は，翻訳に必要な翻訳ルールを対訳コーパスより自動獲得し，それらを用いて翻訳を行う．このような学習型機械翻訳に本手法を適用することで，学習能力の向上の観点より，本手法の有効性を確認できる．性能評価実験の結果，本手法を適用することで

† 北海道大学大学院情報科学研究科，札幌市
Graduate School of Information Science and Technology, Hokkaido University, Kita 14 Nishi 9, Kita-ku, Sapporo-shi, 060–0814 Japan
†† 北海学園大学工学部，札幌市
Faculty of Engineering, Hokkai-Gakuen University, Minami 26 Nishi 11, Chuo-ku, Sapporo-shi, 064–0926 Japan
a) E-mail: terashima@media.eng.hokudai.ac.jp
b) E-mail: echi@eli.hokkai-s-u.ac.jp
c) E-mail: araki@media.eng.hokudai.ac.jp

翻訳ルールの効率的な獲得が可能となり，人手による正誤評価が 5.12 ポイント向上した．更に，様々な自動評価においてもスコアが向上した．これらの結果より，本手法が対訳コーパスに基づく学習型機械翻訳に有効となることが確認された．

2. 学習型機械翻訳システム

本論文における学習型機械翻訳システムの処理概要を図 1 に示す．まず，対訳コーパスから 1 組の対訳文を取り出し，その対訳文の原言語文を翻訳対象文として翻訳部（Translation process）に入力する．翻訳部では，翻訳ルール辞書中の翻訳ルールを用いて入力文に対する翻訳結果を生成する．フィードバック部（Feedback process）では，翻訳部で使用された翻訳ルールに対して，対訳文の目的言語文より得られる正しい翻訳結果に基づき評価を行う．学習部（Learning process）では，与えられた対訳文より翻訳ルールを自動獲得し，それらを順次，翻訳ルール辞書に登録する．

2.1 翻 訳 部

翻訳部では，入力文に対する翻訳結果を生成する．図 2 に翻訳結果生成の具体例を示す．はじめに，入力文 “Where is the hotel?” に対し適用可能な翻訳ルールを選択する．ここで，翻訳ルールは，原言語文から得られる原言語部（Source language part（以下，原部と記す））と目的言語文から得られる目的言語部（Target language part（以下，目的部と記す））のペアで構成される．適用可能な翻訳ルールとは，翻訳ルールの原部のすべて，若しくは変数を除いた全部分が同じ順で入力文に現れる翻訳ルールである．また，翻訳ルールは，文翻訳ルール（Sentence translation rule）と部分翻訳ルール（Part translation rule）の 2 種類である．文翻訳ルールとは図 2 の（Where is @0?；@0/は/どこ/です/か？）のように文の構造をもつ翻訳ルールである．ここで，日本文中の “/” は語の区切りを示している．なお，本論文における単語分割は日本語形態素解析ツール ChaSen (注1) を用いて行った．また，部分翻訳ルールとは（the hotel；ホテル）のように対訳文の部分を示す翻訳ルールである．翻訳結果は，選択された文翻訳ルールと部分翻訳ルールを組み合わせることで生成される．その結果，図 2 においては，(Where is the hotel?；ホテル/は/どこ/です/か？）が得られ，目的部の “ホテルはどこですか？” が翻訳結果となる．

入力文に対し翻訳結果が複数生成された場合は，各翻訳結果に使用された文翻訳ルールの具象度と正適用度に基づき順位付けを行う．すなわち，具象度の高い文翻訳ルールが使用された翻訳結果，続いて正適用度が高い文翻訳ルールが使用された翻訳結果を上位とする．具象度と正適用度はそれぞれ以下の式 (1)，(2) より求める．

$$\text{具象度} = \frac{\text{翻訳ルールの原部の変数以外の単語数}}{\text{入力文の単語総数}} \tag{1}$$

$$\text{正適用度} = \frac{\text{正適用数}}{\text{正適用数} + \text{誤適用数}} \tag{2}$$

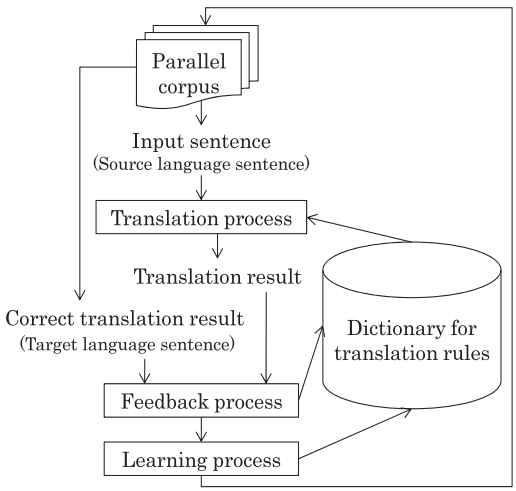


図 1 処 理 概 要
Fig. 1 Outline of processing.

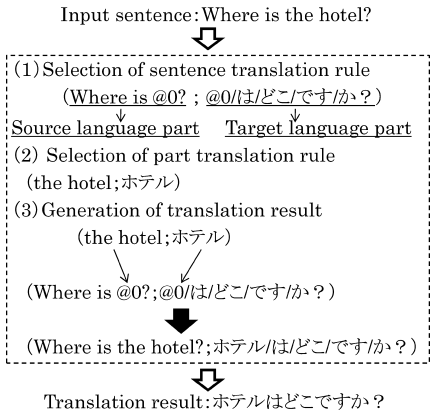


図 2 翻訳結果生成の具体例
Fig. 2 Example of generation of a translation result.

(注1)：奈良先端科学技術大学院大学松本研究室
http://chasen-legacy.sourceforge.jp/

式 (2) の正適用度における，正適用数，誤適用数は **2.2** のフィードバック部で決定される．

2.2 フィードバック部

フィードバック部では，翻訳部で使用された翻訳ルールの評価を行う．正しい翻訳結果と使用された翻訳ルールの目的部が一致する場合，若しくは翻訳ルールの目的部の変数を除いた部分が同じ順序で一致する場合，翻訳ルールの正適用数を 1 増加させる．一致しない場合は，誤適用数を 1 増加させる．

このようなフィードバック処理を行うことで，より良質な翻訳ルールが優先的に使用されるようになる．

2.3 学 習 部

学習部では，従来手法である帰納的学習 [8] と本論文で提案する省略可能情報を用いた部分対応学習を対訳文に適用することで翻訳ルールを自動獲得する．また，対訳文自体も文翻訳ルールとして翻訳ルール辞書に登録される．

2.3.1 帰納的学習による翻訳ルールの獲得

帰納的学習では，二つの対訳文において，共通部分と差異部分の抽出を多段階に行うことで，多様な翻訳ルールを自動獲得する．本論文における共通部分とは，二つの単語列間において表層が一致する共通単語を基本単位とし，それらが連続して出現する場合には連結して得られる部分を指す．更に共通部分以外で 1 語以上の単語が連続している部分を差異部分とする．

帰納的学習による翻訳ルールの獲得処理を図 3 に基づき述べる．対訳文 1 と対訳文 2 においては，“a room” と “部屋” が共通部分となる．したがって，対訳文 1 においては，共通部分の対応関係における組合せ数は 1 となる．差異部分は原言語文中の “I’d like”，“with a balcony” と目的言語文中の “バルコニー/付き/の”，“を/お願い/し/ます” であり，対応関係の組合せ数は 4 となる．共通部分と差異部分の組合せ数が異なる場合，組合せ数の少ない方を抽出する．したがって，この場合，共通部分を抽出することで，部分翻訳ルールとして（a room；部屋）が獲得される．共通部分と差異部分の組合せ数が同じ場合は，共通部分と差異部分のそれぞれを抽出し，翻訳ルールを獲得する．また，共通部分若しくは差異部分において，対応関係の組合せ数が複数となりあいまい性が存在する場合には，すべての組合せを抽出する．そして，すべての組合せから部分間類似度が最も高い組合せを選択し，部分翻訳ルールを獲得する．部分間類似度については **2.3.2** で述べる．更に，抽出した部分を変数 “@0” に置き換えることにより文翻訳ルールを獲得する．図 3 では，対訳文 1 より文翻訳ルールとして（I’d like @0 with a balcony.；バルコニー/付き/の/@0/を/お願い/し/ます.），対訳文 2 より（Can I get @0 for tonight?；今夜/@0/は/あり/ます/か？）がそれぞれ獲得される．このような処理を翻訳ルール対においても適用することで，より一般化された翻訳ルールを獲得する．

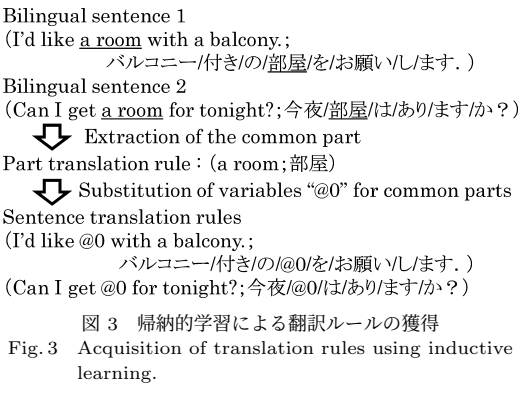


図 3 帰納的学習による翻訳ルールの獲得
Fig. 3 Acquisition of translation rules using inductive learning.

帰納的学習により獲得された部分翻訳ルールは，翻訳処理だけでなく，省略可能情報を用いた部分対応学習による抽出ルールの獲得にも利用される．

2.3.2 部分間類似度

本論文では，部分間の対応関係の確からしさを部分間類似度とし，以下の処理より求める．

（1） 対訳文の原言語文中の部分 P_s の構成単語数を m，構成単語を$W_{si}(i=1,\ldots m)$，また，目的言語文中の部分 P_t の構成単語数を n，構成単語を $W_{tj}(j=1,\ldots n)$と表すとき，全構成単語 W_{si} と W_{tj} の間の類似度 $sim\,(W_{si},W_{tj})$ を Dice 係数により求める．任意の単語 W_s と単語 W_t における Dice 係数は以下の式 (3) により定義される．

$$sim\,(W_s,W_t)=\frac{2f_{st}}{f_s+f_t} \tag{3}$$

f_s は原言語文中における単語 W_s の出現頻度，f_t は目的言語文中における単語 W_t の出現頻度である．また，f_{st} は，単語 W_s，W_t が対訳文に同時に出現する頻度である．出現頻度は，それまでに入力されたすべての対訳文を用いて決定する．

（2） 原言語の各単語 W_{si} に対する最大の類似度の平均$ave_sim_W_s$=$\frac{1}{m}\sum_{i=1}^{m} max[sim\,(W_{si},W_{t1})\ldots sim\,(W_{si},W_{tn})]$ を求める．

（3） 目的言語の各単語 W_{tj} に対する最大の類

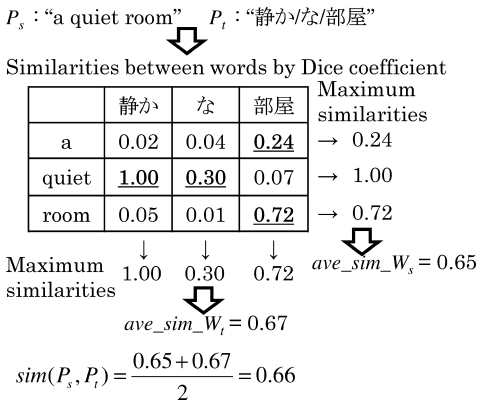


図 4　部分間類似度の計算例
Fig. 4　Example of calculation for similarity between parts.

似度の平均 $ave_sim_W_t = \frac{1}{n}\sum_{j=1}^{n} max[sim\,(W_{s1}, W_{tj}) \ldots sim\,(W_{sm}, W_{tj})]$ を求める．

（4）　$ave_sim_W_s$ と $ave_sim_W_t$ の平均を部分間類似度 $sim\,(P_s, P_t)$ とする．

P_s が “a quiet room”，P_t が “静か/な/部屋” の場合の部分間類似度の計算例を図 4 に示す．はじめに，式 (3) より，Dice 係数を用いて各単語間の類似度を求める．続いて，$ave_sim_W_s$ を求める．図 4 では “a”，“quiet”，“room” に対して，それぞれ 0.24，1.00，0.72 が最も高い類似度であるため，これらの平均値 0.65 が $ave_sim_W_s$ となる．また，$ave_sim_W_t$ は 0.67 ($= \frac{1.00+0.30+0.72}{3}$) となる．したがって，“a quiet room” と “静か/な/部屋” の部分間類似度は $ave_sim_W_s$ と $ave_sim_W_t$ の平均値である 0.66 ($= \frac{0.65+0.67}{2}$) となる．

北村ら [11] は，$\log_2 f_{st}$ を乗算した Dice 係数を用いている．しかし，f_{st} が 1 の単語間の類似度は 0 になるなど，出現頻度の低い単語を含む場合，正しい部分対応であっても部分間類似度が低下することがある．したがって，本論文では重み付けを行わない Dice 係数を用いている．

3.　省略可能情報を用いた部分対応学習

本学習型機械翻訳システムでは，対訳文の部分対応を決定することで翻訳ルールを獲得する．その際，帰納的学習 [8] に基づく従来手法では，句レベルの部分対応が二つの対訳文間の共通部分及び差異部分として出現する必要があった．それに対して，省略可能情報を用いた部分対応学習では抽出ルールを自動獲得し，

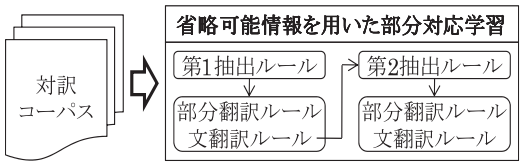


図 5　本手法により自動獲得されるルールの関係
Fig. 5　The relationship of rules acquired by the proposed method.

その抽出ルールに基づき句レベルの部分対応を個々の対訳文から決定する．その結果，効率良くかつ汎用的な翻訳ルールを獲得することが可能となる．本論文では，本手法を図 1 の学習部に導入する．

次いで，本手法により自動獲得されるルールの関係を図 5 に示す．本手法では，帰納的学習で獲得された部分翻訳ルールを用い，対訳文中の省略可能な部分を決定することで第 1 抽出ルールを自動獲得する．次いで，第 1 抽出ルールに基づき対訳文中の部分対応を決定することで部分翻訳ルールと文翻訳ルールを獲得する．また，その文翻訳ルールを用いて，第 2 抽出ルールを自動獲得する．更に，第 2 抽出ルールに基づき対訳文中の部分対応を決定し，より多くの部分翻訳ルールと文翻訳ルールを獲得する．ここで，文翻訳ルールは入力文に対して文の構造をとらえるため，また，部分翻訳ルールは文中の句に相当する部分訳を行うための対訳知識である．第 1 抽出ルール及び第 2 抽出ルールは対訳文中の部分対応を効率良く決定するためのルールである．第 1 抽出ルールは部分対応の先頭及び末尾の部分を含み，その部分も抽出対象となる．第 2 抽出ルールは，対訳文中の部分対応に隣接する部分を含み，その部分は抽出対象とならない．

ここで，第 1 抽出ルールの獲得における省略可能な部分とは，対訳文の原言語文若しくは目的言語文において，名詞句を含む部分を 3 分割し，その中の両側の単語列を連結して得られる部分が対訳コーパスに存在する場合，3 分割された部分の中間の単語列である．名詞句を含む部分とその分割位置は，帰納的学習より得た部分翻訳ルールや単語 bigram 確率を用いて決定する．

また，本手法は様々な言語間の対訳コーパスに対して適用可能であるが，冠詞の存在する言語に最も有効に機能すると考えられる．冠詞は様々な名詞句の先頭に付与され，効率的に部分対応を決定する際に有効な情報となる．したがって，本論文では，冠詞が存在する言語の中で最も広く利用されている英語と身近な言語である日本語で構成される英日の対訳コーパスに適用した．

3.1 第 1 抽出ルールの獲得

以下に第 1 抽出ルール獲得の処理過程を示す.

（1） それまでに獲得された部分翻訳ルールの原部と目的部の両方において，それらの全構成単語が同じ順序で現れる対訳文を選択する．そして，共通単語が 1 語以上連続している部分を共通部分とする.

（2） 対訳文における原言語文若しくは目的言語文の共通部分の数により以下の処理のいずれかを実行する．その際，原言語文，目的言語文中の共通部分の数はともに 2 以下，かつ，どちらか一方の言語文の共通部分の数は 2 でなければならない.

（a） 共通部分の数が 2 の場合，二つの共通部分に挟まれた部分を省略可能な部分と位置づけ，それを共通部分とともに抽出する.

（b） 共通部分の数が 1 の場合，共通部分中の単語と，共通部分以外の単語（対象単語）との bigram 確率を求めることで省略可能な部分を決定する．省略可能な部分の探索は，共通部分の左側に対して行う左側探索と共通部分の右側に対して行う右側探索がある．なお，bigram 確率の計算は，それまで入力された全対訳文を用いて行う．以下にそれぞれの探索の処理過程を示す.

i. 左側探索では，共通部分の左に隣接する単語を除き，共通部分の左側に位置する単語を対象単語として，順次 bigram 確率を求める．その結果，bigram 確率が 0 でなく，更に，対象単語を左側にずらし，bigram 確率が 0 になった場合，探索を終了する．そして，bigram 確率が 0 でない対象単語と共通部分に挟まれた部分を省略可能な部分と位置づける．次いで，省略可能な部分，共通部分をこの並びで抽出する．更に対象単語，省略可能な部分，共通部分をこの並びで抽出する．bigram 確率が 0 のまま文頭に至った場合は，共通部分の左側に位置する全単語を省略可能な部分とする.

ii. 右側探索では i. と同様の処理を共通部分の右側方向に対して行う．bigram 確率が 0 のまま文末に至った場合には，共通部分の右側に位置する全単語を省略可能な部分とする.

（3） 対訳文の原言語文と目的言語文のそれぞれから抽出された部分の組合せに対して部分間類似度を求める．そして，部分間類似度が最大かつしきい値以上の組合せのみを選択する．しきい値には，予備実験に基づき得られた 0.65 を用いる．このしきい値は比較的高い値であるが，良質な第 1 抽出ルールの獲得を優先しているためである.

（4） 選択された原言語部分と目的言語部分の組合せにおいて，省略可能な部分を変数に置き換えることにより第 1 抽出ルールを獲得する.

図 6 に第 1 抽出ルール獲得の具体例を示す．部分翻訳ルールと対訳文間の共通部分の数は原言語文では “a” と “room” の 2 となるため，それらに挟まれた “quiet” を省略可能な部分と位置づけ，“a quiet room” を抽出する．一方，目的言語文では共通部分の数が “部屋” の 1 となるため，bigram 確率により，省略可能な部分を決定する．左側探索においては “部屋” に隣接する “な” を除き，更に左側に位置する単語を対象単語として bigram 確率を求める．対訳コーパス中に “の/部屋” が存在していることで P（部屋｜静か），P（部屋｜の），P（部屋｜付き）はそれぞれ 0，0.06，0 となる．したがって，P（部屋｜の）は 0 でなく，その後の P（部屋｜付き）が 0 となった時点で探索を終了する．その結果，対象単語 “の” と共通部分 “部屋” に挟まれた

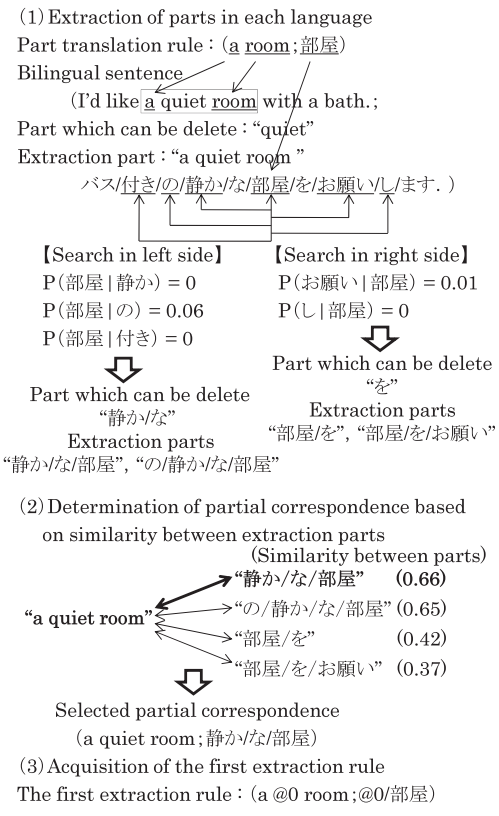


図 6 第 1 抽出ルール獲得の具体例

Fig. 6 Example of acquisition of the first extraction rule.

“静か/な” が省略可能な部分となり，“静か/な/部屋” 及び “の” を含めた “の/静か/な/部屋” が抽出される．右側探索では “部屋/お願い” が対訳コーパスに存在したため，“を” が省略可能な部分となり，“部屋/を” 及び対象単語 “お願い” を含めた “部屋/を/お願い” が抽出される．このように，共通部分の数が 1 の場合には，対象単語を部分対応に含めるべきかどうかを判別するための情報が不足している．したがって，対象単語を含めた場合と含めない場合の両方の部分を抽出し，部分間類似度に基づき決定している．次いで，抽出された部分の全組合せに対して部分間類似度を求める．その結果，部分間類似度が最も高くかつしきい値以上である組合せ（a quiet room；静か/な/部屋）を選択する．更に，省略可能な部分 “quiet” と “静か/な” を変数に置き換えることで，第 1 抽出ルールとして（a @0 room；@/部屋）を獲得する．

また，第 1 抽出ルール間において帰納的学習を適用することで，より汎用的な第 1 抽出ルールを獲得する．例えば，(a @0 room；@0/部屋）と（a @0 hotel；@0/ホテル）の二つの第 1 抽出ルールにおいて，差異部分を変数化することにより，(a @0；@0）を得る．この第 1 抽出ルールは，原言語文中において “a” とその右側に隣接する部分で構成される句レベルの部分が目的言語文中の句レベルの部分と対応することを表している．

3.2 第 1 抽出ルールの利用

本節では第 1 抽出ルールを用いた部分対応の決定及び翻訳ルールの獲得について述べる．以下にその処理過程を述べる．

（1） 第 1 抽出ルールの原部と目的部の変数以外の全部分が同じ順序で現れる対訳文を選択する．

（2） 対訳文の原言語文と目的言語文において，共通部分の数により，以下の処理のいずれかを行う．

（a） 共通部分の数が 2 の場合，共通部分に挟まれた部分を探索範囲とする．

（b） 共通部分の数が 1 の場合において，第 1 抽出ルールの原部若しくは目的部中の変数が共通部分の左側に存在する場合，対訳文の原言語文若しくは目的言語文中の共通部分の左側に隣接する全単語を探索範囲とする．また，共通部分の右側に変数が存在する場合，対訳文の原言語文若しくは目的言語文中の共通部分の右側に隣接する全単語を探索範囲とする．

（c） 共通部分が存在しない場合，対訳文の原言語文若しくは目的言語文の全単語を探索範囲とする．

（3） 対訳文の原言語文と目的言語文に存在する共通部分の数の組合せにより，以下の処理のいずれかを行う．

（a） 対訳文の原言語文と目的言語文に存在する共通部分の数がともに 2 である場合，それぞれの言語文から共通部分に挟まれた探索範囲を抽出する．

（b） 対訳文の原言語文と目的言語文に存在する共通部分の数の片方のみが 2 の場合，以下の処理を行う．

i. 共通部分の数が 2 である言語文（基準言語文）において，共通部分に挟まれた探索範囲を抽出する．

ii. 基準言語文より抽出された部分の探索範囲の先頭から末尾の各単語に対し，もう一方の言語文（対象言語文）の探索範囲の全単語との類似度を式 (3) の Dice 係数より求め，最も高い類似度かつしきい値以上の単語の組合せを選択する．

iii. 対象言語文の探索範囲において，処理（b）の ii. により選択された単語の数が 1 の場合には，その単語を抽出する．単語の数が複数の場合には，二つの単語に挟まれた部分の構成単語数が最大となる部分を抽出する．

（c） 対訳文の原言語文と目的言語文に存在する共通部分がともに 1 以下の場合，以下の処理を行う．

i. 共通部分の数が多い方の言語文（基準言語文）の探索範囲において，共通部分に隣接する単語を選択する．なお，対訳文中に存在する共通部分の数が原言語文，目的言語文ともに 1 の場合には，原言語文を基準言語文として処理（6）までを行う．更に，目的言語文を基準言語文として処理（6）までを行う．

ii. 選択された単語に対し，基準言語文に対するもう一方の言語文（対象言語文）より決定された探索範囲の全単語との類似度を式 (3) の Dice 係数より求め，最も高い類似度かつしきい値以上の単語の組合せを選択する．更に，基準言語文の共通部分に対して次に近くに位置する単語を選択し，同様の処理を行う．この処理を基準言語文の探索範囲において，類似度がしきい値以下の単語，末尾の単語，若しくは先頭の単語に至るまで繰り返す．

iii. 基準言語文の探索範囲において，共通部分に隣接する単語から処理（c）の ii. が終了した際に使用された単語までの部分を抽出する．

iv. 対象言語文の探索範囲において，処理（c）の ii. により選択された単語の数が 1 の場合には，その単語を抽出する．単語の数が複数の場合には，二つの単語に挟まれた部分の構成単語数が最大となる部分を抽出する．

（4） 対訳文の原言語文若しくは目的言語文に共通部分が存在する場合，次の処理を行う．共通部分の数が 2 の場合，処理（3）により抽出された部分とともに共通部分を抽出する．共通部分の数が 1 の場合には，処理（3）により抽出された部分から共通部分までを抽出する．

（5） 抽出された原言語の部分と目的言語の部分間の組合せに対して部分間類似度を求め，その値がしきい値以上の組合せを部分対応とする．

（6） 決定された部分対応を部分翻訳ルールとして獲得する．更に，対訳文中の部分対応の箇所を変数に置き換えることにより，文翻訳ルールを獲得する．

上記の処理（3）の類似度及び処理（5）の部分間類似度に対するしきい値には予備実験に基づき得られた 0.35 を用いる．この値は比較的低い値であるが，多くの翻訳ルールの獲得に有効となる．

第 1 抽出ルールを用いた翻訳ルール獲得の具体例を図 7 に示す．この場合，原言語文においては共通部分の数は “a” の 1 であるため処理（2）の（b）により探索範囲は “return ticket” となる．また，目的言語文においては共通部分は存在しないため処理（2）の（c）により探索範囲は全単語である “帰り/の/航空/券/を/お/持ち/です/か” となる．

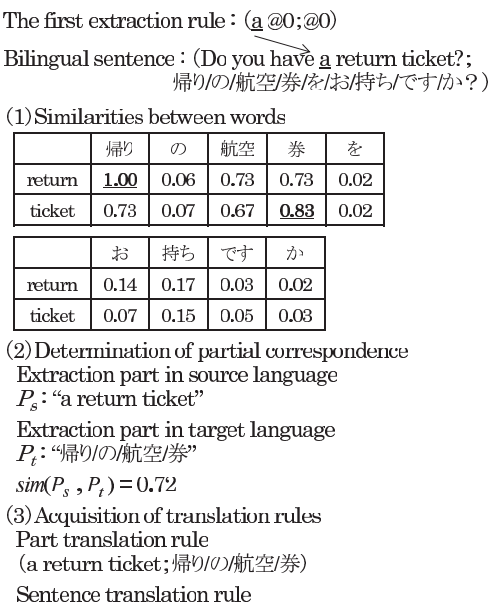


図 7 第 1 抽出ルールを用いた翻訳ルール獲得の具体例
Fig. 7 Example of acquisition of translation rules using the first extraction rule.

次いで，処理（3）以降に基づき，部分対応を決定する．図 7 では，共通部分は原言語文における “a” のみとなるため，処理（3）の（c）により共通部分の数が多い原言語文が基準言語文，目的言語文が対象言語文となる．はじめに “a” に隣接する “return” と対象言語文の探索範囲にある全単語との類似度を求めると，図 7 の (1) より “帰り” との類似度が 1.00 と最も高く，またしきい値 0.35 以上であるため，組合せ(return；帰り)を選択し，処理を継続する．基準言語文の探索範囲中の末尾の単語 “ticket” では，最も高い類似度は “券” との 0.83 であり，しきい値 0.35 を上回っているため，その組合せ(ticket；券)を選択する．したがって，原言語文からは共通部分 “a” も含めた “a return ticket” を抽出する．目的言語文においては，“帰り” から “券” までの “帰り/の/航空/券” を抽出する．抽出した部分間の部分間類似度は 0.72 となり，しきい値 0.35 以上であるため，抽出部分の組（a return ticket；帰り/の/航空/券）は部分対応及び部分翻訳ルールとなる．更に，対訳文中の部分対応の箇所を変数 “@0” に置き換えることにより文翻訳ルールとして（Do you have @0?；@0/を/お/持ち/です/か?）が得られる．また，これらの処理は対訳文だけでなく，変数を含む文翻訳ルールに対しても再帰的に行われるため，より汎用的な文翻訳ルールが獲得される．

3.3 第 2 抽出ルールの獲得とその利用

本手法では，より多くの抽出ルールを獲得するために，第 1 抽出ルールより獲得された文翻訳ルールを用いて，第 2 抽出ルールを獲得する．以下に第 2 抽出ルール獲得の処理過程を示す．

（1） 原部，目的部の両方において変数に隣接する部分が共通部分となる文翻訳ルール対を選択する．その際，原部間と目的部間のそれぞれにおいて，変数に対する共通部分の左右の位置は一致しなければならない．

（2） 文翻訳ルール対における原部と目的部のそれぞれから共通部分と変数で構成される部分を抽出し，それらの組を第 2 抽出ルールとして獲得する．

第 2 抽出ルールの獲得の具体例を図 8 に示す．図 8 では，“from” と “から” が共通部分となり，かつ，原部間，目的部間ともに変数の位置が共通部分に対して一致するため，(from @0；@0/から）を第 2 抽出ルールとして獲得する．第 2 抽出ルールは原言語文中の “from” の右側に隣接する部分と，目的言語文中の “から” の左側に隣接する部分が対応関係にあること

Sentence translation rule 1
（How do I get there from @0?;
@0/から/は/どう/行け/ば/良い/の/です/か？）
Sentence translation rule 2
（Do you have a bus from @0?;
@0/から/の/バス/は/あり/ます/か？）
⇩ Extraction of common part and variable “@0”
The second extraction rule：(from @0;@0/から)

図 8　第 2 抽出ルール獲得の具体例
Fig. 8　Example of acquisition of the second extraction rule.

を表している．

このような第 2 抽出ルールを適用することで，対訳文中の部分対応を決定するとともに翻訳ルールを獲得する．その際の処理過程は **3.2** の第 1 抽出ルールの適用方法と同様である．ただし，**3.2** の処理（4）において，共通部分を部分対応に含めない点のみが異なる．第 1 抽出ルールは，対訳文中に存在する部分対応中の先頭及び末尾の部分が含まれている．そのため，それ自体も部分対応に含まれる．それに対して，第 2 抽出ルールは，部分対応に隣接する部分を含んでいる．そのため，それ自体は部分対応に含まれない．例えば，第 2 抽出ルール（with @0；@0/付き/の）を対訳文（Can I have a room with twin beds?；ツイン/ベッド/付き/の/部屋/は/あり/ます/か?）に適用する場合，共通部分 “with” と “付き/の” を除いた（twin beds；ツイン/ベッド）が部分対応となる．

抽出ルールは，対訳文から部分対応を決定する際に，その探索範囲を限定するための情報をもつ．すなわち，抽出ルールは対訳文中の部分対応の先頭及び末尾の部分若しくは部分対応に隣接する部分を含んでいるため，効率的に部分対応を決定することが可能となる．

4. 性能評価実験

4.1 実験方法

実験データには旅行用英会話 [12]〜[21] における英日の対訳文 1,710 組を使用した．実験データの英文と日本文の平均単語数はそれぞれ 6.6 と 7.6 である．実験に用いたシステムは，**2.** で述べたシステムにおいて，省略可能情報を用いた部分対応学習を適用したシステムと，適用しないシステムの二つである．実験は初期状態として辞書を空の状態にし，1 文入力するごとに翻訳と学習をすべての実験データに対し繰り返し行った．これは与えられたすべての対訳文を用いることで，常にその時点での最適な翻訳結果の生成が可能になると考えられるためである．得られた全翻訳結果のうち，後半の文献 5 冊 [17]〜[21] にある 762 文に対する翻訳結果を，ある程度学習が進んだ段階の結果と位置づけ，評価データとした．このように，本実験では，評価データに対しても学習処理及びフィードバック処理を行っている．フィードバック部で求める式 (2) の正適用度は，翻訳部で利用され，かつ，学習部で獲得される翻訳ルールに付与される．すなわち，翻訳部，フィードバック部，そして，学習部の三つの処理は連動している．したがって，本実験においてはこれら三つの処理を行い，本手法の有効性を確認する．

4.2 評価方法

システムが生成した翻訳結果に対する評価は，人手による正誤評価と自動評価を用いた．システムが複数の翻訳結果を生成した場合は，**2.1** で述べた順位付けの結果，上位 1 位となった翻訳結果のみを評価の対象とする．また，人手による正誤評価では，翻訳結果は有効な翻訳と無効な翻訳に分類される．有効な翻訳結果は，以下の二つである．

（1） 未登録語を含まない正しい翻訳結果
（2） 未登録語を含む正しい翻訳結果

未登録語を含む正しい翻訳結果とは，未登録語が名詞，形容詞，及び名詞句に相当する場合，訳語を与えることにより未登録語を含まない正しい翻訳結果が容易に得られる翻訳結果である．すなわち，有効な翻訳結果とは，入力文に対して文の核となる動詞が正しく訳され，かつ，正しい翻訳結果に対して文の構造が正しく反映された翻訳結果である．これは，本学習型機械翻訳システムが文翻訳ルールに基づいて翻訳を行うことから適切な評価方法と考えられる．また，有効な翻訳結果以外はすべて無効な翻訳結果となる．そして，全評価データにおける有効な翻訳結果の割合を有効な翻訳率として求める．個々の翻訳結果の評価は第 1 著者が行う．

更に，自動評価による評価を行った．自動評価基準には，BLEU, IMPACT [10], NIST, mWER, mPER を用いた．その際，翻訳結果 1 文に対して四つの参照訳を使用した．mWER，mPER はスコアが低いほど良質な翻訳と位置づけられ，その他はスコアが高いほど良質な翻訳と位置づけられる．

4.3 実験結果

表 1 に正誤評価における実験結果を示す．表中の () 内の数値は有効な翻訳結果数である．表 1 より，本手法を適用することで，有効な翻訳率は 5.12 ポイント増加した．また，表 2 に自動評価における実験結果を

示す．表 2 では各自動評価において，本手法を適用したシステムのすべてのスコアが，本手法を適用しないシステムのスコアを上回った．そして，表 3 には，本手法を適用することで無効な翻訳結果から有効な翻訳結果となった翻訳結果の具体例を示す．表中の翻訳不能とは適用可能な文翻訳ルールが存在せず全く翻訳が行えなかった場合である．また，翻訳結果中の英単語は未登録語である．

4.4 考　　察

人手による正誤評価においては，表 3 に示すような無効な翻訳結果から有効な翻訳結果となった翻訳結果は 44 文存在した．逆に有効な翻訳結果から無効な翻訳結果となった翻訳結果は 5 文存在した．したがって，全体として有効な翻訳結果は 39 文増加したことになる．また，図 9 に 100 文ごとの有効な翻訳率の推移を示す．図 9 より，すべての区間において，本手法を適用したシステムが本手法を適用しないシステムを上回り，かつ，学習が進むにつれて有効な翻訳率が増加する傾向にあることが確認できる．有効な翻訳結果の増加は，抽出ルールによる良質な翻訳ルールの増加が原因である．本論文における良質な翻訳ルールとは，原部と目的部の対応関係が正しい翻訳ルールである．

抽出ルールを適用することにより，文翻訳ルールの数は 4,424 から 6,615 に増加した．その中で，良質な文翻訳ルールの数は 3,285 から 4,170 に増加した．したがって，本手法により増加した良質な文翻訳ルールの数は 885 であり，その精度は 40.39%($= \frac{885}{2,191} \times 100$) となる．一方，部分翻訳ルールの数は 1,137 から 2,958 に増加した．その中で，良質な部分翻訳ルールの数は 625 から 1,358 に増加した．したがって，本手法により増加した良質な部分翻訳ルールの数は 733 であり，その精度は 40.25%($= \frac{733}{1,821} \times 100$)となる．このように，精度が低下したにもかかわらず有効な翻訳結果の数が増加した要因について述べる．

本学習型機械翻訳システムでは，文の構造をとらえた，より良質な文翻訳ルールを使用するために，式 (1) より得られる具象度の高い文翻訳ルールから優先的に使用している．そこで，本手法により無効な翻訳結果から有効な翻訳結果となった 44 文を対象に，以下の調査を行った．はじめに，個々の入力文ごとに翻訳処理の際，本手法を適用しないシステムで使用された文翻訳ルールの中で最大の具象度を求めた．そして，同じ入力文ごとに，本手法を適用したシステムの翻訳処理に使用された文翻訳ルールの中で，求めた具象度より大きい具象度をもつ文翻訳ルールを選択し，良質な文翻訳ルールの精度を求めた．その結果，比較的高い 68.25%が得られた．したがって，本手法により翻訳処理を行う際の良質な文翻訳ルールが増加し，その結果，翻訳精度が向上したことを確認した．

更に，今回の実験で獲得された変数をもたない良

表 1　正誤評価における実験結果
Table 1　Experimental results of right-wrong evaluation.

システム	有効な翻訳率	内訳	
		未登録語なし	未登録語あり
本手法あり	55.91% (426)	73.47% (313)	26.53% (113)
本手法なし	50.79% (387)	73.39% (284)	26.61% (103)

表 2　自動評価における実験結果
Table 2　Experimental results of automatic evaluation.

システム	BLEU	IMPACT	NIST	mWER	mPER
本手法あり	0.355	0.549	6.17	0.555	0.460
本手法なし	0.329	0.519	5.79	0.585	0.507

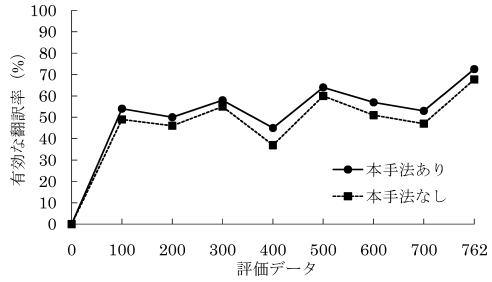


図 9　有効な翻訳率の推移
Fig. 9　Change of the effective translation rates.

表 3　翻訳結果における本手法の有効性
Table 3　Effectiveness of proposed method in translation results.

入力文	本手法なし	本手法あり
I'd like a room with shower.	付きのシャワー部屋をお願いします.	シャワー付きの部屋をお願いします.
Can I have a room with a shower?	a shower ベッド付きの部屋はありますか？	シャワー付きの部屋はありますか？
How much will it take to Sheraton Hotel?	How much will it take to Sheraton Hotel?（翻訳不能）	シェラトンホテルまで料金はどのくらいかかりますか？
I have AMEX.	I have AMEX.（翻訳不能）	AMEX を持ってます.

質な部分翻訳ルールが何度目の出現で獲得されたのかについて調査を行った．その結果，出現頻度が1，2，そして，3以上で獲得された部分翻訳ルールの数はそれぞれ658，194，145となり，出現頻度が1のみで獲得されたものが全体の66.00%を占めていた．また，この出現頻度が1の時点で獲得された部分翻訳ルールの精度は42.59%($= \frac{658}{1,545} \times 100$)であった．本手法を適用しないシステムでは51.00%($= \frac{280}{549} \times 100$)であった．しかし，良質な部分翻訳ルールの数は2.35倍に増加した．したがって，本手法は精度の低下を招くが，良質な部分翻訳ルールの大幅な増加をもたらした．

次いで，良質な翻訳ルールの増加を可能とした抽出ルールについて述べる．表4に獲得された良質な抽出ルールの内訳及びその具体例を示す．本実験により獲得された第1抽出ルールの総数は25，第2抽出ルールの総数は58であった．そのうち，良質な第1抽出ルールの数は13，良質な第2抽出ルールの数は16であった．また，表4より，第1抽出ルールにおいては，冠詞を含むものが数多く獲得された．これは，第1抽出ルールの獲得を可能とした部分翻訳ルールの多くに冠詞が含まれていたためである．冠詞を含まない第1抽出ルールにおいては，“your”のようにその訳語が省略された対訳文が存在したために獲得された．一方，部分対応に隣接する部分を含む第2抽出ルールにおいては，前置詞をもつものがその多くを占めていた．冠詞を含む部分対応の多くは前置詞や動詞に隣接していたが，前置詞はその種類の少なさから獲得が容易であったため，多くの第2抽出ルールに含まれたと考えられる．

誤った抽出ルールの獲得は，第1抽出ルール獲得の際に誤った部分翻訳ルールが使用されることなどが原因であった．例えば，誤った部分翻訳ルール（into dollars；ドル）と対訳文（Could you change 100 Canadian dollars into U.S. dollars, please?；100/カナダ/ドル/を/US/ドル/に/両替/し/て/ください?）においては，英文中の“U.S.”と日本文中の“US”が省略可能な部分となり，誤った第1抽出ルール（into @0 dollars；@0/ドル）が獲得された．このような誤った抽出ルールの獲得に対しては，フィードバック処理を強化するなど，今後解決に向けて取り組む予定である．

表4　良質な抽出ルールの詳細
Table 4　Details of high quality extraction rules.

第1抽出ルール	第2抽出ルール
冠詞あり（10個）	前置詞あり（13個）
(a @0 room；@0/部屋)	(with @0；@0/付き/の)
(the @0 Hotel；@0/ホテル)	(near @0；@0/の/近く)
(a @0；@0)	(at @0；@0/に)
(the @0；@0)	(from @0；@0/から)
冠詞なし（3個）	前置詞なし（3個）
(your @0；@0)	(is @0；@0/は)

また，本学習型機械翻訳システムでは，評価データも学習データに加えるため，入力する文の順序が変わると翻訳結果が変化する．そこで，それがどの程度翻訳精度に影響するのかを調査した．評価データ762文を2分割し，その前後を入れ換えて実験を行った．その結果，有効な翻訳率は56.17%となり，表1の55.91%に比べ大きな変化は見られなかった．これは，評価データも学習することで，学習データが増えていくのに伴い良質な翻訳ルールが増加するという点は，入力文の順序の変化に影響されないためと考えられる．

一方，現在の本学習型機械翻訳システムは学習データに出現していない表現については翻訳ルールとして獲得できず，翻訳が困難となる．例えば，時制や単複による語形変化が生じると，それらは全く異なる単語と認識され，対処することが困難である．また，複文や重文などに対しては，個々の文をとらえた文翻訳ルールが獲得されても，それらを組み合わせることが困難である．したがって，学習データが不十分な場合，それを補うための形態素情報や複数の文翻訳ルールを結合するための規則が必要となる．しかし，それ以前に，学習データが十分であっても，翻訳ルールを効率的に獲得できなければ，適用ドメインは更に小さくなる．このような観点より，本手法を用いて，学習データからのより効率的な翻訳ルールの獲得を図っている．

5. 関連研究

はじめに，対訳コーパスに基づく機械翻訳という点で類似性の高い統計的機械翻訳（SMT）を用いて，本手法を検証する．今回使用したSMTは翻訳モデルにはGIZA++ [4]を用い，言語モデルにはSRILM [5]を用いた．また，翻訳デコーダにはMoses [6]を用いた．一方，本手法を適用したシステムと本手法を適用しないシステムにおいては評価データに対して学習処理とフィードバック処理は行わない．実験は，実験データ1,710文による5分割交差検定を行った．評価方法には，Fluencyと Adequacyの観点での3名のバイリンガルによる5段階評価を用いた．本比較実験の結果を表5に示す．表中の数値は3名の評価値の平

均である．また，図 10，図 11 に実験データ 1,710 文に対する 3 名の Fluency と Adequacy の 5 段階評価値の内訳を示す．図 10，図 11 より，本学習型機械翻訳システムは，評価値が高い 4 と 5 の数においては，SMT と比べて大きな差がないことを確認できる．したがって，表 5 における，本学習型機械翻訳システムと SMT との違いは，評価値が低い 1 と 2 の数の違いと考えられる．これは，本学習型機械翻訳システムでは，良質な部分翻訳ルールが得られても，文翻訳ルールが存在しなければそれを翻訳に反映できないことに起因していると考えられる．したがって，良質な部分翻訳ルールの有効利用が重要となる．これは今後の課題である．

更に，対訳文中の効率的な部分対応の決定の観点での検証を行った．実験データの対訳文の英文中の名詞句において，その名詞句の部分対応の出現頻度が 1 であるものを対象に，それらを部分翻訳ルールとしてどの程度抽出できたかの調査を行った．その際，SMT では文法的な句の抽出のみが目的ではないため，名詞句を含む正しい部分対応がフレーズテーブルに存在していれば，正しい抽出と位置づけた．ただし，文の形を有した節に相当するものはフレーズの抽出の観点より不十分とみなし対象外とした．その結果，本手法を適用したシステムと SMT の抽出率はそれぞれ 30.88%($= \frac{201}{651} \times 100$)，27.65%($= \frac{180}{651} \times 100$)となった．一方，名詞句そのものを正しい抽出とした場合の SMT の抽出率は 18.74%($= \frac{122}{651} \times 100$)となった．これらの結果より，本手法は低頻度の部分対応を効率良くかつ正確に決定できることが明らかとなった．

表 5　比較実験
Table 5　Comparative experiments.

システム	Fluency	Adequacy
本手法あり	2.63	2.82
本手法なし	2.54	2.71
SMT	2.82	3.09

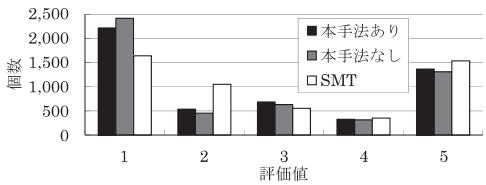


図 10　Fluency における 5 段階評価の内訳
Fig. 10　Details of five-grade evaluation in fluency.

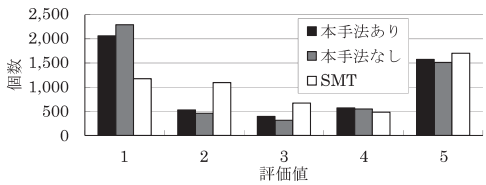


図 11　Adequacy における 5 段階評価の内訳
Fig. 11　Details of five-grade evaluation in adequacy.

Ong ら [22]，Malavazos ら [23]，そして，Güvenir ら [24] は，二つの対訳文間から共通部分及び差異部分を抽出することで，翻訳ルールを獲得している．しかし，句に相当する部分の決定においては，部分対応の構成単語数が多くなると，それが共通部分及び差異部分として出現する可能性は低くなり，翻訳ルールの獲得が困難となる．それに対して，本手法では，自動獲得される抽出ルールとその利用により，個々の対訳文から効率良く翻訳ルールを獲得することが可能である．Chiang [7] は，統計的手法である GIZA++ [4] に基づき部分対応を決定することで翻訳ルールを自動獲得している．しかし，単語の出現頻度が低い場合には，単語間の対応関係において曖昧性が高くなり，正確な部分対応の抽出が困難になると考えられる．それに対して，本手法による抽出ルールは部分対応の先頭や末尾の部分若しくは部分対応に隣接する部分を有しているため，より正確な部分対応の抽出が可能となる．しかし，本手法を適用したシステムは，抽出した正確な部分対応を最大限に生かすには至っていない．したがって，今後は獲得した良質な部分翻訳ルールを最大限に生かした翻訳処理の導入が必要と考えられる．

6. むすび

本論文では，効率的かつ正確に対訳文中の句に相当する部分対応を決定する新たな手法として，省略可能情報を用いた部分対応学習を提案した．そして，本手法の有効性を確認するために，学習型機械翻訳システムに本手法の適用を行った．性能評価実験の結果，翻訳精度が向上し，本手法の有効性が確認された．

今後は，本手法を SMT に適用することで，より良い対訳コーパスに基づく機械翻訳システムの実現を図る．更に，他の言語を用いた性能評価実験や，同一の表現が数多く出現する対訳コーパスを用いた性能評価実験を通して，本手法の有効性を検証する予定である．

謝辞　本研究の一部は，平成 19 年度財団法人北海道科学技術総合振興センター「研究開発助成事業」及び北海学園大学ハイテク・リサーチ・センター研究費「私立大学戦略的研究基盤形成支援事業」の補助金に

よって行われた．

文　献

[1] 池原　悟，阿部さつき，徳久雅人，村上仁一，“非線形な表現構造に着目した重文と複文の日英文型パターン化，”自然言語処理，vol.11, no.3, pp.69–95, July 2004.

[2] 荒牧英治，黒橋禎夫，佐藤理史，渡辺日出雄，“用例ベース翻訳のための対訳文の句アライメント，”自然言語処理，vol.10, no.5, pp.75–92, Oct. 2003.

[3] M. Sasayama, F. Ren, and S. Kuroiwa, “Automatic super-function extraction for translation of spoken dialogue,” Natural Language Processing and Knowledge Engineering 2007, pp.141–148, Beijing, China, Sept. 2007.

[4] F.J. Och and H. Ney, “A Systematic comparison of various statistical alignment models,” Computational Linguistics, vol.29, no.1, pp.19–51, March 2003.

[5] A. Stolcke, “SRILM - An extensible language modeling toolkit,” 7th International Conference on Spoken Language Processing, pp.901–904, Denver, Colorado, USA, Sept. 2002.

[6] P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, C. Dyer, O. Bojar, A. Constantin, and E. Herbst, “Moses: Open source toolkit for statistical machine translation,” Proc. ACL 2007, pp.177–180, Prague, Czech Republic, June 2007.

[7] D. Chiang, “Hierarchical phrase-based translation,” Computational Linguistics, vol.33, no.2, pp.201–228, June 2007.

[8] 荒木健治，自然言語処理ことはじめ—言葉を覚え会話のできるコンピュータ，森北出版，東京，2004.

[9] 越前谷博，荒木健治，桃内佳雄，栃内香次，“翻訳例に基づく再帰チェーンリンク型学習による機械翻訳手法，”信学論（D-II），vol.J85-D-II, no.12, pp.1840–1852, Dec. 2002.

[10] H. Echizen-ya and K. Araki, “Automatic evaluation of machine translation based on recursive acquisition of an intuitive common parts continuum,” Proc. 11th Machine Translation Summit, pp.151–158, Copenhagen, Denmark, Sept. 2007.

[11] 北村美穂子，松本裕治，“対訳コーパスを利用した対訳表現の自動抽出，”情処学論，vol.38, no.4, pp.727–736, 1997.

[12] 石川洋一，トラベル・コミュニケーション研究会，ひとり旅これで十分英会話，実業之日本社，東京，1992.

[13] W. Reed，困った時のトラベル英会話入門，日本文芸社，東京，1994.

[14] 甲斐順子，ひとり歩きの英語自遊自在，日本交通公社出版事業局，東京，1996.

[15] 荒木庸子，J.C. Lee，旅行英会話ポケットブック，日本文芸社，東京，1987.

[16] 旅行英会話研究会，海外旅行英会話，実業之日本社，東京，1980.

[17] K.S. Gilbert，ケントのトラベル英会話，実業之日本社，東京，1995.

[18] 前川　裕，アメリカを自由に歩く旅の英会話，池田書店，東京，1994.

[19] 丸子量二，海外旅行英会話，土屋書店，東京，1997.

[20] 地球の歩き方編集室，旅の会話集 2 米語/英語，ダイヤモンド・ビッグ社，東京，1996.

[21] 斎藤晃雄，六ヶ国語会話 1 ヨーロッパ・アメリカ編，日本交通公社出版事業局，東京，1960.

[22] E. Ong, K. Go, M. Morga, V. Nunez, and F. Veto, “Extracting and using translation templates in an example-based machine translation system,” J. Research in Science, Computing, and Engineering, vol.4, no.3, pp.83–100, 2007.

[23] C. Malavazos and S. Piperidis, “Application of analogical modeling to example based machine translation,” Proc. Coling 2000, pp.516–522, Saarbücken, Germany, 2000.

[24] H.A. Güvenir and I. Cicekli, “Learning translation templates from examples,” Information Systems, vol.23, no.6, pp.353–363, 1998.

（平成 21 年 6 月 2 日受付，9 月 24 日再受付）



寺島　涼　（学生員）

平 19 北海学園大・工・電子情報卒．平 21 北大大学院情報科学研究科修士課程了，現在，同大学院情報科学研究科博士後期課程在学中．自然言語処理の研究に従事．情報処理学会会員．



越前谷　博　（正員）

平 3 北海学園大・工・電子情報卒．平 8 同大大学院工学研究科修士課程了．平 10 北大大学院工学研究科電子情報工学専攻博士後期課程退学．博士（工学）．現在，北海学園大学工学部電子情報工学科助教．自然言語処理，特に機械翻訳の研究に従事．情報処理学会，人工知能学会，言語処理学会，ACL，IEEE 等各会員．



荒木　健治　（正員）

昭 57 北大・工・電子卒．昭 63 同大大学院博士課程了．博士（工学）．同年，北海学園大学工学部電子情報工学科助手．平元同講師．平 3 同助教授，平 10 同教授．平 10 北大・工・電子情報工学専攻助教授．平 14 同教授．現在，北大・情報科学・メディアネットワーク専攻教授．自然言語処理，特に機械翻訳，対話処理などの研究に従事．情報処理学会，人工知能学会，言語処理学会，ACL，IEEE，AAAI 等各会員．