解　説

機械翻訳最新事情：（上）統計的機械翻訳入門

永田 昌明　　渡辺 太郎　　塚田 元
NTT コミュニケーション科学基礎研究所

「統計的機械翻訳」(statistical machine translation) は，互いに翻訳になっている 2 つの言語の文の対から翻訳規則や対訳辞書を自動的に学習し，言語翻訳を実現する技術である．この技術は過去 10 年間に大きく進歩し，アラビア語と英語のような語順が比較的近い言語対では，従来の翻訳手法より精度が高いと言われている．本稿では，上下 2 編に分けて，近年の自然言語処理で最もホットな話題である統計的機械翻訳の技術概要，および，評価型ワークショップを中心とした最先端の研究動向を報告する．

はじめに

インターネットの普及により，ごく普通のユーザが英語・中国語・韓国語などの外国語で書かれた Web ページに接する機会が飛躍的に増えた．世界中に市場を持つ多国籍企業は，マニュアルなどの製品情報を迅速かつ正確に現地語で提供しなければならない．計算機科学のように英語が世界共通語となった分野の専門家は，常に情報の発信や取得を英語で行う必要がある．言語の壁を超えて人とコミュニケーションしたい，または，言語の壁を超えて知識を交換したいという想いは人類の普遍的な要求の 1 つと言ってよい．

コンピュータを利用してある言語を別の言語に翻訳する技術を「機械翻訳」(machine translation) と呼ぶ．機械翻訳の研究はコンピュータの誕生とほぼ同時に 1950 年代から始まり，今日までに多数の機械翻訳ソフトウェアが開発された．しかし，正直なところ現状の翻訳ソフトがユーザのさまざまなニーズを満足させているとは言い難い．

「統計的機械翻訳」(statistical machine translation) は，1990 年前後に従来の機械翻訳手法が抱えている問題点を解決するまったく新しいアプローチとして提案された．このアプローチは 2000 年代に入ると学術コミュニティで主流を占めるようになり，ついに 2007 年には Web ページの自動翻訳サービスや携帯電話の音声通訳サービスなどへの応用が始まった☆1．

本稿「機械翻訳最新事情」では，この統計的機械翻訳に関する最新の研究状況を上下 2 編に分けて解説する．「(上) 統計的機械翻訳入門」では，ほぼ歴史的な経緯に沿って統計的機械翻訳の基本的な考え方と代表的な手法を概観する．「(下) 評価型ワークショップの動向と日本からの貢献」では，世界の研究機関が技術を競う評価型ワークショップとそこでの日本の研究機関の貢献を紹介する．本稿を通じて，少しでも多くの方が統計的機械翻訳の研究に興味を持っていただければ幸いである．

統計的機械翻訳とは

従来の機械翻訳システムでは，新しい言語間の翻訳を実現するために，数人から数十人の言語学の専門家が何年もの歳月をかけてコンピュータ処理向けの文法規則と辞書を作成する必要があった．このような機械翻訳へのアプローチを「知識に基づく機械翻訳」(knowledge-based machine translation) と呼ぶ．

統計的機械翻訳では，互いに翻訳になっている 2 つの言語の文の対から翻訳規則や対訳辞書に相当する数学的なモデルを自動的に学習する．統計的機械翻訳の最大の特徴は，自動学習により，新しい言語間（たとえば英語

☆1 Google Translate BETA や ATR-Trek の「しゃべって翻訳」など．

とアラビア語）や特定の分野（たとえば特許）の翻訳システムを短期間に低コストで作成できることである．

身近にある大規模な多言語データといえばWebである．言語統計が翻訳に有効なことはインターネット検索エンジンを使えば簡単に実感できる．たとえば，英語のメールを書いていて「〜してもらえると有難い」と言いたいが，"I would be grateful if you could .." と "It would be grateful if you could.." のどちらが適切か迷った場合，検索エンジンにこれらの表現をそのままキーワードとして入力し，検索された文書数が多い方を選ぶとよい．すなわち，ある言語（日本語）から別の言語（英語）へ翻訳するときに，翻訳先（英語）の言語データが大量にあれば，正しい翻訳を選べる可能性が高い．

あるいは，たとえば「股関節」に対応する英単語を探したい場合，「股関節」と "of" の2つの単語をキーワードとして入力し，検索結果を眺めるとよい．検索結果に "hip" という単語が何度も出現するので，次に「股関節」と "hip" をキーワードとして入力して検索結果を眺めると，"hip joint" の出現頻度が高いことからこれが股関節の訳語であることが分かる☆2．すなわち，ある言語（日本語）から別の言語（英語）へ翻訳されたテキストが大量にあるとき，互いに翻訳になっている部分に何度も共起する単語のペアを探せば対訳を得られる可能性が高い．

統計的機械翻訳は，このようなヒューリスティックな対訳語の探索や翻訳先言語の文生成を，コンピュータの力を借りて大規模かつ系統的に行う手法に関する研究である．「統計的」の名前の通り，以降の説明は非常に数学的なものになるが，上記の2つの例に共感できれば，技術の詳細は必要に応じて理解すればよい．また大まかな研究の流れさえ分かればよいという場合は，途中を読み飛ばして「おわりに」へ進んでいただいて結構である．

言語翻訳の生成モデル

統計的機械翻訳の研究は，1980年代後半にIBMのワトソン研究所の音声認識グループで始まった[1]．初期の研究がフランス語から英語への翻訳を対象としていたため，統計的機械翻訳の分野では，原言語（source language，翻訳元言語）をフランス語 **f**，目的言語（target language，翻訳先言語）を英語 **e** と表記する習慣があり，本稿もこれに従う．

一般に，あるフランス語の文に対してさまざまな英語の文への翻訳が考えられる．統計的機械翻訳では，あるフランス語の文 **f** に対してすべての英語の文 **e** が翻訳になり得ると考え，すべての文の対 (**e**, **f**) に対して「翻訳者が **f** を **e** に翻訳する可能性」に相当する確率 $P(\mathbf{e}|\mathbf{f})$ を割り当てる．このとき，与えられた **f** に対して確率 $P(\mathbf{e}|\mathbf{f})$ を最大にする $\hat{\mathbf{e}}$ を選べば，フランス語を英語に翻訳する際の誤りを最小にできる．ベイズの法則により結局 $P(\mathbf{e})P(\mathbf{f}|\mathbf{e})$ を最大にする文を探せばよい．

$$\hat{\mathbf{e}} = \arg\max_{\mathbf{e}} P(\mathbf{e}|\mathbf{f}) = \arg\max_{\mathbf{e}} P(\mathbf{e})P(\mathbf{f}|\mathbf{e}) \qquad (1)$$

式 (1) は，「雑音のある通信路モデル」（noisy channel model）を言語翻訳に適用したことを意味する．翻訳すべきフランス語の文は，非常に雑音の多い通信路において英語がフランス語に変形したと見なし，これを元の英語の文へ復元する処理が言語翻訳であると考える．

一般に，英語の文の事前確率 $P(\mathbf{e})$ を計算するためのモデルを言語モデル（language model），英語の文が与えられたときのフランス語の文の条件付き確率 $P(\mathbf{f}|\mathbf{e})$ を計算するためのモデルを翻訳モデル（translation model）と呼ぶ☆3．また言語翻訳は雑音のある通信路による符合化（encode）の逆過程という解釈から，$P(\mathbf{e})P(\mathbf{f}|\mathbf{e})$ を最大化する英語の文を探索する処理をデコード（decode，復号），復号を実行する処理系をデコーダ（decoder，復号器）と呼ぶ．

同じ内容を2つの言語で記述したテキストの集合を対訳コーパス（parallel corpus）または二言語コーパス（bilingual corpus）と呼ぶ．翻訳モデルや言語モデルは対訳コーパスから学習する．対訳コーパスとしては，カナダの国会議事録（Hansards），EUの議会議事録（Europarl），香港の議会議事録，国連の刊行物，多国籍企業の製品マニュアル，通信社のニュース記事などが統計的機械翻訳の研究に利用されている．

単語に基づく翻訳

1990年前後にIBMは順番に少しずつ複雑になるモデル1からモデル5までの5つの翻訳モデルを提案した[1]．この「IBM翻訳モデル」は，対訳文において互いに翻訳になっている単語を結び付ける単語対応（word alignment）という概念を基本としている．**図-1** の上段に単語対応の例を示す．一般に単語対応は多対多対応であり，英語の冠詞や日本語の助詞のように相手言語に対応する単語がない場合もある．

IBM翻訳モデルでは，フランス語の文 **f** と英語の文 **e** が互いに翻訳になっており，その単語対応が **a** である

☆2 "of" は英語の文を含むページを検索するために加えたキーワードなので，英語の高頻度語であれば何でもよい．一部の検索エンジンは高頻度語を無視するので "+of" などと入力する必要がある．

☆3 ベイズの法則を用いた式 (1) のせいで，翻訳モデルの原言語（英語）/目的言語（フランス語）は翻訳システムの原言語（フランス語）/目的言語（英語）と逆になる．この混乱を避けるために原言語と目的言語ではなくフランス語と英語を使う習慣になった（と思われる）．

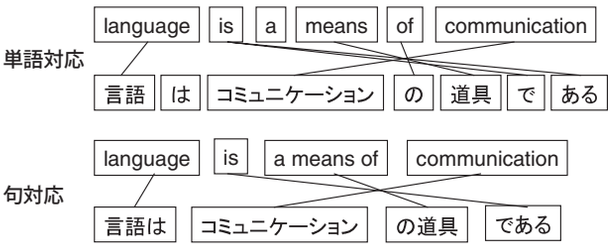

図 -1　単語対応と句対応

ような同時確率分布 $P(\mathbf{f}, \mathbf{a}, \mathbf{e})$ を考え，$P(\mathbf{f}|\mathbf{e})$ をすべての単語対応に関する条件付き確率 $P(\mathbf{f}, \mathbf{a}|\mathbf{e})$ の和として表す．

$$P(\mathbf{f}|\mathbf{e}) = \sum_{\mathbf{a}} P(\mathbf{f}, \mathbf{a}|\mathbf{e}) \tag{2}$$

IBM 翻訳モデルでは，英語からフランス語への単語対応を一対多と制限し，翻訳モデルのパラメタを期待値最大化法（EM アルゴリズム）を用いて推定する．一対多対応に制限することにより，対応関係が正しく表現できない場合が生じるが，可能な単語対応の総数が大幅に減るのでパラメタの推定は容易になる．また局所的な最適値に陥ることを避けるために，同じ訓練データに対して，より簡単なモデルのパラメタの推定値を次のモデルのパラメタの初期値とするという手順によりモデルを推定する．

IBM 翻訳モデルは後述するようにさまざまな限界が指摘され，現在では翻訳モデルとして使用されることは少ない．IBM 翻訳モデルを作成する GIZA++[8] というフリーソフトが公開されて以降は，単語対応を求めるブラックボックス的なツールとして使用されることが多いので，本稿では説明を省略する．詳しく文献 5）等を参照してほしい．

句に基づく翻訳

2000 年前後から翻訳の基本的な単位を単語から句に拡張する研究が盛んになり成功を収めた．ここでいう句（phrase）は，名詞句や動詞句といった言語学的な文の構成要素ではなく，単に連続した単語列を指す．句を翻訳の基本単位とすることにより，局所的な単語の並び替え，複数単語から構成される表現，局所的な文脈に依存する単語の挿入や削除などを句の対応として翻訳モデルの中に表現することができる．

図 -1 に句対応の例を示す．たとえば，"a means of" と「の道具」を句単位で対応させているように，句に基づく翻訳モデルでは，英語における名詞 "means" と助詞 "of" の語順が日本語における助詞「の」と名詞「道具」の語順と逆になるという現象や，英語における冠詞 "a" に対応する日本語の単語が存在しないという現象に対して精緻な確率モデルを用意するのではなく，このような対訳句の用例を大量にテーブルに記憶することで対処する．また単語を翻訳の基本単位とする場合に比べ，句を単位として並び替えを行うことにより並び替え操作の数が少なくなる．

ここでは句に基づく統計的機械翻訳（phrase-based SMT）の代表例として，Koehn ら[4] が提案した翻訳モデルとデコーダを紹介する．この翻訳モデルでは，まず原言語の文 f を I 個の句の列 $\bar{f}_1^I = \bar{f}_1 ... \bar{f}_I$ に分割し，原言語の各句 $\bar{f}_i$ を目的言語の句 $\bar{e}_i$ に翻訳し，句を並び替える．翻訳確率 $P(\mathbf{f}|\mathbf{e})$ は句翻訳確率（phrase translation probability）$\phi(\bar{f}_i \mid \bar{e}_i)$ と相対的な句歪み確率（phrase distortion probability）$d(a_i - b_{i-1})$ の積で近似する．

$$p(\bar{f}_1^I|\bar{e}_1^I) = \prod_{i=1}^{I} \phi(\bar{f}_i|\bar{e}_i)\, d(a_i - b_{i-1}) \tag{3}$$

ここで a_i は，i 番目の目的言語句に翻訳された原言語句の開始位置であり，b_{i-1} は，$(i-1)$ 番目の目的言語句に翻訳された原言語句の終了位置である．

句翻訳確率は抽出された句の相対確率から求める．

$$\phi(\bar{f}|\bar{e}) = \frac{count(\bar{f}, \bar{e})}{\sum_{\bar{f}'} count(\bar{f}', \bar{e})} \tag{4}$$

ここで $count(\bar{f}, \bar{e})$ は，原言語句 $\bar{f}$ と目的言語句 $\bar{e}$ の対応付けの頻度である．句歪み確率は，適当に決めたパラメタ α を用いて，句の移動距離に対して指数的に大きくなるペナルティを与える．

$$d(a_i - b_{i-1}) = \alpha^{|a_i - b_{i-1,-1}|} \tag{5}$$

上記の句歪み確率は，距離や移動方向，原言語句や目的言語句への依存性など，大局的な句の並び替えの傾向を表現していない．Nagata ら[6] は，日本語と英語のような語順が大きく異なる言語向けに，より詳細な句の並び替えモデルを提案している．

互いに翻訳となる句は，単語対応付けされた対訳コーパスからヒューリスティクスを用いて抽出する．**図 -2** に例を示す．まず IBM 翻訳モデルを用いて原言語から目的言語および目的言語から原言語の双方向の単語対応を求め，両者の積集合（intersection）と和集合（union）を求める．積集合の対応点（alignment point）は信頼度が高いと考えられるので，積集合の対応点を起点に新しい対応点を加えて対応点が内部に閉じているような句の対応を取り出す．新しい対応点は和集合の中から既存の対応点に隣接し新しい単語に単語対応を与えるものを選ぶ．

原言語の入力文に対する目的言語の翻訳出力は，ビーム探索により文頭から文末方向に部分的な翻訳を生

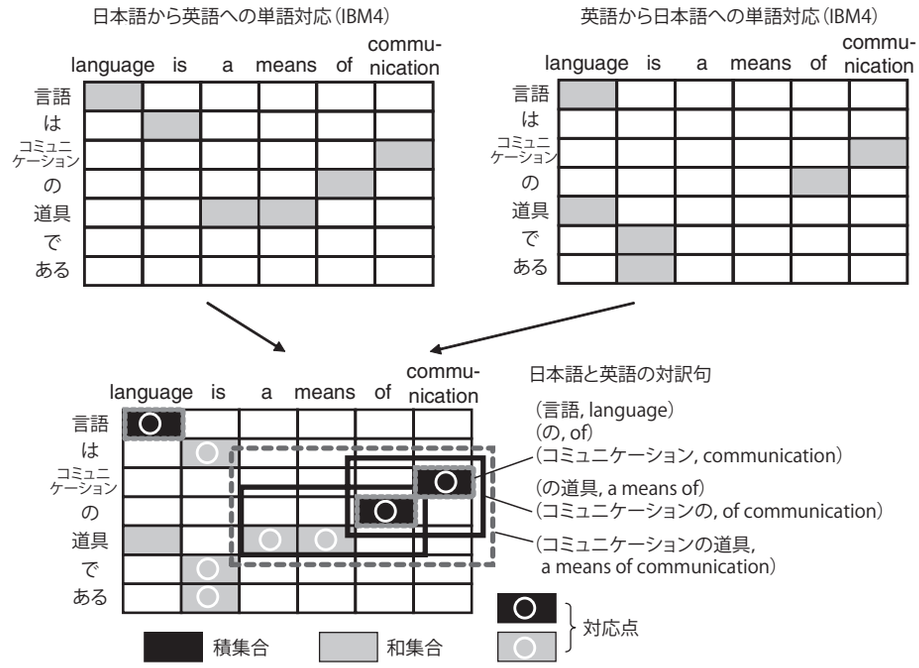

図 -2　単語対応付けからの対訳句の抽出

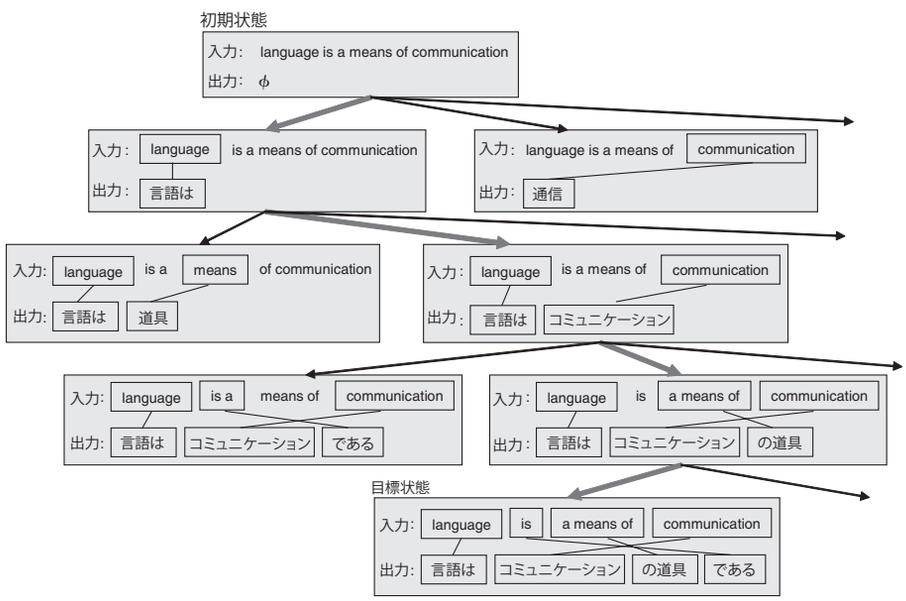

図 -3　ビーム探索によるデコーディング

成する．**図 -3** に例を示す．まず空（empty）の初期仮説（initial hypothesis）から出発し，ある仮説において 1 つの句を翻訳して新しい仮説を作るステップを繰り返す．1 つの句の翻訳では，原言語の文で未翻訳の単語列から原言語の句を 1 つ選び，対応する目的言語の句を目的言語の部分文の文末側に付加する．仮説は優先順位付きキュー（priority queue）で管理し，原言語の文のすべての単語を翻訳したら探索は終了する．

言語翻訳の識別モデル

◆対数線形モデル

統計的自然言語処理における生成モデル（generative model）から識別モデル（discriminative model）への流れに呼応して，2000 年代前半から雑音のある通信路モデルに代わって対数線形モデル（log linear model）を用いて事後確率 $P(\mathbf{e}|\mathbf{f})$ を直接モデル化する方法が主流になった．対数線形モデルでは，M 個の素性関数 $h_m(\mathbf{e}, \mathbf{f})$ と，

各素性（feature）に対する重み λ_m を考え，翻訳の事後確率 $P(\mathbf{e}|\mathbf{f})$ を次式により求める．

$$p_{\lambda_1^M}(\mathbf{e}|\mathbf{f}) = \frac{\exp\sum_{m=1}^{M}\lambda_m h_m(\mathbf{e},\ \mathbf{f})}{\sum_{\mathbf{e}'}\exp\sum_{m=1}^{M}\lambda_m h_m(\mathbf{e}',\ \mathbf{f})} \quad (6)$$

入力文に対する翻訳を求める際には式（6）の分母を計算する必要はなく，素性と重みの線形和を最大とする候補を探索すればよい．

$$\hat{\mathbf{e}} = \arg\max_{\mathbf{e}} P(\mathbf{e}|\mathbf{f}) = \arg\max_{\mathbf{e}} \sum_{m=1}^{M}\lambda_m h_m(\mathbf{e},\ \mathbf{f}) \quad (7)$$

この式は，$h_1(\mathbf{e}|\mathbf{f}) = \log p(\mathbf{e})$, $h_2(\mathbf{e}|\mathbf{f}) = \log p(\mathbf{f}|\mathbf{e})$, $\lambda_1 = \lambda_2 = 1$ とすれば式（1）と同じでなので，このモデルは雑音のある通信路モデルを包含している．パラメタ λ_1 と λ_2 の最適化は，最適なモデルの重み（model scaling factor）を求めることに相当する．通常，素性 $h_m(\mathbf{e}, \mathbf{f})$ としては，翻訳モデル，言語モデル，歪みモデル，単語の長さなどが用いられる．

訓練データとしてS個の文の対からなる対訳コーパス $\{(\mathbf{e}_s, \mathbf{f}_s)|s = 1, \ldots S\}$ が与えられたとき，モデルパラメタ λ_1^M は最尤推定する（すなわちコーパスの事後確率を最大にするように求める）．対数線形モデルの尤度は凸（convex）関数であり，一般化反復スケーリング（Generalized Iterative Scaling）や勾配（gradient）に基づく最適化法により大局的な最適値を求められる．

$$\hat{\lambda}_1^M = \arg\max_{\lambda_1^M} \sum_{s=1}^{S}\log p_{\lambda_1^M}(\boldsymbol{e_s}|\boldsymbol{f_s}) \quad (8)$$

学習の際に式（6）の分母，すなわち，入力文のすべての翻訳候補に関する和を求める必要がある．通常は，確率が大きい順に上位N個の翻訳候補を求め，このN-best候補の確率の和で分母を近似する．

◆翻訳品質の評価尺度

次節で述べる最小誤り率学習への準備として，最も標準的な翻訳精度の評価尺度であるBLEU(Bilingual Evaluation Understudy)[9]を紹介する．翻訳精度の評価は非常に難しい問題であり，BLEUの誕生は機械翻訳の研究に革命的な変化をもたらした．翻訳精度の評価方法は評価型ワークショップと密接に関係するので下編で詳しく解説することとし，上編では必要最小限の説明にとどめる．

BLEUは，機械による翻訳はプロの翻訳者による翻訳（参照訳，reference）に類似しているほどよいと考え，類似度を0から1の間の数値で表す．具体的には，システムが出力した1つの翻訳候補と正解集合（複数の参照訳）の間の異なる長さの単語ngramの適合率（precision）p_n の幾何平均に，短い文へのペナルティであるBPを掛けたものである．

$$BLEU = BP \times \exp\left(\frac{1}{N}\sum_{n=1}^{N}\log p_n\right) \quad (9)$$

ここで単語ngramとは連続するn個の単語列であり，単語ngramの適合率 p_n とは候補に含まれるすべての単語ngramのうち正解集合の含まれる単語ngramと一致したものの割合である．通常はN=4を用いる．

◆最小誤り率学習

式（8）に基づくモデルパラメタの最尤推定の問題点は，尤度最大のパラメタが翻訳精度を最大にする保証がないことである．BLEUのような自動評価尺度が誕生したことから，Och[7]は，翻訳精度の評価尺度を直接最大化するパラメタ推定法である「最小誤り率学習」（Minimum Error Rate Training）を提案した．

参照訳 $\mathbf{r}$ に対する翻訳候補 $\mathbf{e}$ の誤りを評価する関数を $E(\mathbf{r}, \mathbf{e})$ とする．たとえばBLEUを最適化する場合は $E = 1 - BLEU$ とすればよい．訓練データとして対訳コーパス $\{(\mathbf{e}_s, \mathbf{f}_s)|s = 1,\ldots S\}$ が与えられたとき，最小誤り率学習では，訓練データにおける最適候補と参照訳の誤りの総和が最小になるようにモデルパラメタ λ_1^M を求める．

$$\hat{\lambda}_1^M = \arg\min_{\lambda_1^M} \sum_{s=1}^{S} E\left(\boldsymbol{e}_s,\ \arg\max_{\boldsymbol{e}} p_{\lambda_1^M}(\boldsymbol{e}|\boldsymbol{f}_s)\right) \quad (10)$$

ここで $\arg\max_{\mathbf{e}} p_{\lambda_1^M}(\mathbf{e}|\mathbf{f}_s)$ は，パラメタ λ_1^M のモデルで $\mathbf{f}_s$ を翻訳したときの確率最大の候補である．

式（10）の誤り関数は凸関数ではないので，勾配に基づく最適化法は使えない．そこで，まずランダムに選んだ λ_1^M から出発し，他のパラメタを固定して1つのパラメタ λ_m について最小化することを繰り返す．最小誤り率学習は，この1つのパラメタに関する最小化において対数線形モデルの性質をうまく利用することにより計算量を削減している．

構文に基づく翻訳

大局的な語句の並び替えをうまく扱うために，構文理論（syntactic theory），特に自然言語の階層構造を翻訳モデルの中で利用する「構文に基づく統計的機械翻訳」（syntax-based SMT）が2000年代前半から現在まで盛んに研究されている．これまでに言語学的な構文解析に基づくものから形式言語論的な木変換（tree transduction）に基づくものまでさまざまな翻訳モデルが提案されているが，ここでは現在最も精度が良いとされる「階層的句に基づく翻訳」（Hierarchical Phrase-Based Translation）[2]を紹介する．

階層的句に基づく翻訳は「同期文脈自由文法」（Synchronous Context Free Grammar, SCFG）に基づいている．同期文脈自由文法の規則は一般に次式で表される．

$X \rightarrow <\gamma, \alpha, \sim>$

ここで X は非終端記号，γ と α は終端記号と非終端記号の列，$\sim$ は γ と α に含まれる非終端記号の間の 1 対 1 対応を表す．

階層的句に基づく翻訳規則は，ある句が他の句を含むことを許す．これにより非連続な句（変数を含む翻訳規則）や句の並び替え規則を表現できる．以下に例を示す．ここで枠付きの添字は∼でリンクされていることを表す．

$X \rightarrow X_{\boxed{1}}$ である，is $X_{\boxed{1}}$ (11)

$X \rightarrow X_{\boxed{1}}$ の $X_{\boxed{2}}$，$X_{\boxed{2}}$ of $X_{\boxed{1}}$ (12)

$X \rightarrow$ 言語は，language (13)

$X \rightarrow$ コミュニケーション，communication (14)

$X \rightarrow$ 道具，a means (15)

最初の規則は変数を含む翻訳規則の例であり，2 番目の規則は句の並び替えの例である．3 番目以降の規則は前節で説明した句に基づく翻訳の対訳句と同じであり，階層的句に基づく翻訳は句に基づく翻訳の拡張になっていることが分かる．

同期文脈自由文法の導出 (derivation) は，対応付けられた開始記号の対から始まり，各ステップにおいて 1 つの規則の右辺にある 2 つの要素を使って 2 つのリンクされた非終端記号を書き換える．一般には X を開始記号としてもよいが，Chiang[2] は，以下の 2 つの接着規則 (glue rule) を導入し，入力文を並び替えのないチャンク (chunk) の列に分割することを許して，頑健性を高めている．

$S \rightarrow S_{\boxed{1}} X_{\boxed{2}}$，$S_{\boxed{1}} X_{\boxed{2}}$ (16)

$S \rightarrow X_{\boxed{1}}$，$X_{\boxed{1}}$ (17)

図 -4 にこれらの規則を用いた対訳文の導出の例を示す．

階層的な句は単語対応付けされた対訳テキストからの句の抽出を一般化することにより抽出する．まず句に基づく翻訳の場合と同様に，単語対応付けと矛盾しない句を抽出し，これを初期句 (initial phrase) とする．次に句の内部に他の句を含む場合，含まれる句を非終端記号に置換する．たとえば図 -2 において（コミュニケーションの , of communication）という句は（コミュニケーション , communication）という句を含むことから，$X \rightarrow X_{\boxed{1}}$ の，of $X_{\boxed{1}}$ という規則を抽出できる．

上記の方法では非常に多くの規則が生成されるので，実際には，初期句の長さは最大 10 単語，非終端記号は最大 2 個，少なくとも 1 つの対応付けされた単語を含むなどの制約を加える．翻訳モデルの重みは，対数線形モデルの最小誤り率学習により推定する．

デコーディングは，同期文脈自由文法の原言語側の規則を使って原言語の文を解析して原言語の構文木を作成し，これを目的言語の構文木に写像して終端記号を読み出すことにより目的言語の文を得る．各規則に含まれる非終端記号を最大 2 個までに制限しているので，解析には CKY アルゴリズムが使える．

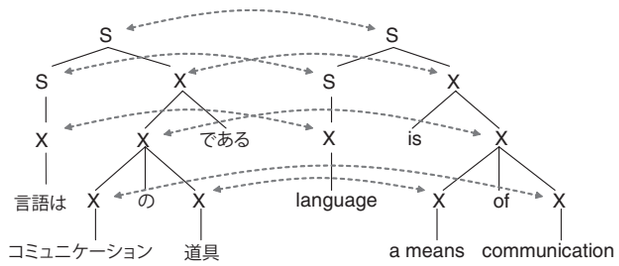

図 -4　階層的句に基づく対訳文の導出

構文に基づく翻訳のデコーダの実装において最も悩ましいのは，言語モデルの重みを統合する方法である．句に基づく翻訳のデコーダでは，目的言語を文頭から文末方向へ連続する単語列として生成するため，単語 ngram による言語モデルを簡単に組み込むことができ，早期の枝刈り（ビーム探索）が可能となる．構文に基づく翻訳のデコーダでは，解析の途中段階において，目的言語を必ずしも連続した単語列として生成できないので言語モデルを組み込めず，効率的な探索が難しい．

Watanabe ら[10] は，翻訳規則を抽出する際に，目的言語側に右辺の先頭要素は必ず終端記号であるという Greibach 標準形と同じ制約を加え，目的言語が文頭から文末方向へ連続する単語列として生成されるように原言語の Earley アルゴリズムによるトップダウン解析を制御することにより，言語モデルを簡単に適用できる方法を提案した．Chiang[2] は，目的言語の言語モデルを組み込んだ CKY アルゴリズムによる原言語の部分解析の k-best 候補を効率よく計算する cube pruning を提案した．Huang と Chiang[3] は，cube pruning に遅延評価 (lazy evaluation) を導入することによりさらに計算量を削減した cube growing を提案した．

おわりに

人類の夢の 1 つである機械翻訳の研究は，少なくともアメリカでは常にその時代の政治・経済状況に影響される．冷戦下の 1950 年代はロシア語を英語に翻訳することが研究目標だった．現在の統計的機械翻訳の研究ブームは，2001 年の同時多発テロ以降，アラビア語および中国語から英語への翻訳に多額の国防予算が投入されたことと無縁ではない．

統計的機械翻訳はこの 10 年間に大きく進歩し，アラビア語と英語のような語順が比較的近い言語対では，従来手法より精度が高いといわれている．今後，構文情報

の利用が進むことにより，日本語と英語のような語順が大きく異なる言語対の翻訳精度も向上するだろう．現在では，IBM 翻訳モデルを作成する GIZA++ や句に基づく翻訳デコーダ moses[☆4] などのオープンソースの統計翻訳ツールが整備されており，対訳コーパスさえあれば，誰でも簡単にほぼ最先端の技術を体験できる．

副次的なことだが，統計的機械翻訳は，機械翻訳の研究スタイルを一変させた．従来の機械翻訳システムは，大勢の研究者が１つのシステムを構築するために，リーダの下に堅固な(陸軍的な)統制体制ができてしまい，研究者個人の貢献が外部に見えないことが多かった．これに対して統計的機械翻訳は，自動学習アルゴリズムの検討など基本的に１人で解決可能ないくつかの部分問題に分割できるので，個人の能力と努力がそのまま研究成果に反映される．それゆえ腕自慢的な評価型ワークショップが生まれ，F1 レースのような競争の下で技術が急速に進歩する．この様子については下編で詳しく紹介する．

統計的機械翻訳は，定跡よりも力任せの探索を重視することで世界チャンピオンと互角に戦うレベルに達したコンピュータチェスに似ているといわれる．膨大な対訳句候補をメモリに記憶し，その組合せを CPU パワーで力任せに探索する「非人間的」なアプローチのせいか，統計的機械翻訳に対しては自然言語処理コミュニティの内部でも拒絶反応を示す研究者が多い．

翻訳がどれぐらい知的な行為であり，本当に力任せの探索で解決するのかは工学的にも哲学的にも興味深い問題である．工学的な立場に立つ筆者らは，コンピュータがプロの翻訳者を超える日を楽しみに，今後も研究を続けたいと思っている．

参考文献

1) Brown, P. F., Pietra, S. A. D., Pietra, V. J. D. and Mercer, R. L. : The Mathematics of Statistical Machine Translation : Parameter Estimation, Computational Linguistics, Vol.19, No.2, pp.263-311 (1993).
2) Chiang, D. : Hierarchical Phrase-Based Translation, Computational Linguistics, Vol.33, No.2, pp.201-228 (2007).
3) Huang, L. and Chiang, D. : Forest Rescoring : Faster Decoding with Integrated Language Models, Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics (ACL-07), pp.144-151 (2007).
4) Koehn, P., Och, J. F. and Marcu, D. : Statistical Phrase-Based Translation, Proceedings of the Joint Conference on Human Language Technologies and the Annual Meeting of the North American Chapter of the Association of Computational Linguistics (HLT-NAACL-03), pp.127-133 (2003).
5) 永田昌明：確率モデルによる自然言語処理，言語と心理の統計，pp.59-128, 岩波書店 (2003).
6) Nagata, M., Saito, K., Yamamoto, K. and Ohashi, K. : A Clustered Global Phrase Reordering Model for Statistical Machine Translation, Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics (COLING-ACL-06), pp.713-720 (2006).
7) Och, F. J. : Minimum Error Rate Training in Statistical Machine Translation, Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics (ACL-03), pp.160-167 (2003).
8) Och, F. J. and Ney, H. : A Systematic Comparison of Various Statistical Alignment Models, Computational Linguistics, Vol.29, No.1, pp.19-51 (2003).
9) Papineni, K., Roukos, S., Ward, T. and Zhu, W.-J. : BLEU : a Method for Automatic Evaluation of Machine Translation, Proceedings of the 40th Annual Meeting of the Association for Computational Lnguistics (ACL-02), pp.311-318 (2002).
10) Watanabe, T., Tsukada, H. and Isozaki, H. : Left-to-Right Target Generation for Hierarchical Phrase-Based Translation, Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics (COLING-ACL-06), pp.777-784 (2006).

(平成 19 年 12 月 6 日受付)

永田　昌明（正会員）
nagata.masaaki@lab.ntt.co.jp

1987 年京都大学大学院工学研究科修士課程修了．現在，コミュニケーション科学基礎研究所主幹研究員．工学博士．統計的自然言語処理の研究に従事．

渡辺　太郎
taro@cslab.kecl.ntt.co.jp

2003 年京都大学大学院情報学研究科博士後期課程研究指導認定退学．現在，コミュニケーション科学基礎研究所リサーチスペシャリスト．京都大学博士（情報学）．統計的機械翻訳の研究に従事．

塚田　　元（正会員）
tsukada@cslab.kecl.ntt.co.jp

1989 年東京工業大学大学院理工学研究科修士課程修了．現在，NTT コミュニケーション科学基礎研究所主任研究員．統計的機械翻訳および音声言語処理の研究に従事．

☆4 http://www.statmt.org/moses/