■2 群（画像・音・言語）- 10 編（自然言語処理）

4 章 機械翻訳

【本章の構成】

本章では，ルールベイス翻訳（4–1 節），用例翻訳（4–2 節），統計翻訳（4–3 節），評価尺度（4–4 節），翻訳支援（4–5 節），機械翻訳の活用（4–6 節），対訳用語抽出（4–7 節），翻字（4–8 節）について述べる．

■2 群 - 10 編 - 4 章

4-1 ルールベイス翻訳

■2 群 - 10 編 - 4 章

4-2 用例翻訳

■2 群 - 10 編 - 4 章

4-3 統計翻訳

■2 群 - 10 編 - 4 章

4-4 評価尺度

■2 群 - 10 編 - 4 章

4-5 翻訳支援

（執筆者：阿辺川武，影浦　峡）［2009 年 8 月 受領］

機械翻訳が究極の理念的目標として人間の翻訳と同様の翻訳を自動的に行わせることを想定しているのに対して，翻訳支援は人間が行う翻訳をできる限り効率化することを目指すもので，自然言語処理分野ではとりわけ ケイ（Kay）による影響力の強い論文が発表されて以来[1]，研究開発が盛んになった．現在では，商用システムも含め，複数の翻訳支援システムが実用に供されている．

4-5-1 概要と技術的展開

翻訳支援システムを広義に捉えるならば，電子辞書や，オンライン上の辞書を横断的に検索するシステム，下訳段階で利用される機械翻訳システムなども翻訳支援システムの一環を構成すると考えることができるが，通常，翻訳支援システムといった場合に指す範囲はもう少し狭く，もっぱら翻訳を支援するために特化して開発されたシステムを指し，翻訳メモリ，用語管理機能，文書管理機能のすべてあるいは何れかを備えていることが特徴である．実用システムは，特許翻訳を含む産業翻訳を中心に導入されている．

翻訳メモリ（Translation Memory）は，多くの翻訳支援システムにおいて中核的機能を構成するもので，原文と翻訳文を文単位，場合によっては句などのより細かい単位でデータベース化し，それらの再利用を可能にするメカニズムである[2]．曖昧マッチング機能を備えたものが多く，全く同じ文だけでなく類似の文の先行訳例を活用することができる．これにより同じ文章を繰り返し翻訳する手間が省けるほか，翻訳表現の統一も図ることができる．複数の翻訳者が分担で作業するときにも一定の基準を提供するため，グループで進める翻訳プロジェクトのクォリティコントロールにも有用である．

産業翻訳の分野では，文学翻訳とは異なり，用語の統一が極めて重要になる．そのため，用語管理（Terminology Management）も翻訳メモリと並び，多くの翻訳支援システムに組み込まれている．事前に作られた用語集（Glossary）の登録活用，翻訳文書からの用語抽出と登録，用語翻訳の統制と共有といった機能が提供されていることが多い．

文書管理機能は，翻訳そのものを支援するというよりは翻訳マネージメントを支援するという色彩が強く，文書のバージョン管理，複数の翻訳者が共同作業している場合の制御，文書の分割とマージ，校正支援など，単言語の文書管理システムと同等の機能が提供されていることが多い．

研究の前線としては，いくつかのテーマがある．第一に，既往の対訳文書から翻訳メモリ化する対訳単位抽出の精度の向上や対訳用語抽出の精度向上などであり，これらは特定の翻訳支援システム改善の一部としてだけでなく自然言語処理の基礎技術として研究が進められている．第二に，翻訳メモリを有効に活用するための曖昧マッチング技術の改善であり，入力のコンプリーションとして翻訳メモリから訳文を提示する方法の最適化などを含んだ研究が，とりわけ構造的に近い言語間で進められている[3]．関連する研究課題として，辞書引きにおける柔軟なイディオム検索の実現などがある[4]．第三に，アプリケーションシステムとしての使い勝手をめぐる改善と研究がある[5, 6]．実際，様々な機能をどのようなインタフェースに実現するかは，要素技術の高度化が進んだ現在，翻訳支援システムの使い勝手を左右す

る極めて重要な要因となっている．例えば，前述の入力のコンプリーションとして候補を提示するシステムでは，翻訳に熟達したユーザに対してキーボード入力速度にコンプリーションが間に合わない，コンプリーションにより入力のリズムが崩されるため機能を ON にしないといった問題が生じている．

翻訳支援を広義に捉えた場合には，翻訳者の熟練度に応じて翻訳支援に要求される機能が異なるため，技術的な方向性も多岐にわたることになる．原言語に完全に精通していないユーザであれば，辞書引きによる読解支援機能や，機械翻訳を一つの参照訳として手軽に参照・修正できる機能が必要とされる．一方で，熟達した翻訳者であれば，複数の辞書からの訳語候補が一覧でき最適な表現が想起できる機能や，書籍や Web サイトで用いられている既訳表現を確認する機能といった，翻訳のクォリティを向上させることを目的とした機能が必要とされる．これらを実現するためには，辞書構築や Web 処理など幅広い基礎技術が必要となる．

4-5-2 主なシステムと応用の展開

翻訳支援は実用システム指向で開発が進められてきたこともあるため，ここでは代表的なシステムを簡単にあげることにする．商用の翻訳支援システムとして，SDL Trados[7]，Transit[8]，TraTool[9]，Deja Vu[10] などがある．強調点は異なるものの，いずれも産業翻訳業界に向けた標準的な翻訳支援システムとして上述の基本的な機能を備えたもので広い範囲で利用されている．富士通「ATLAS」[11]や東芝「The 翻訳」[12]のように，機械翻訳システムの展開として翻訳メモリ機能を組み込み，翻訳支援システム的な側面を打ち出しているものもある．

応用の観点から，最近ではソフトウェアのローカライゼーションやオンライン文書の翻訳を中心に，プロの翻訳者ではなくボランティア翻訳者が担う翻訳の領域が増えてきた．こうしたボランティアに向けて，Omega-T[13]のようなオープンソースのフリーな翻訳支援システムも開発されている．Google Translators' Toolkit[14] はオープンソースではないが，専門の翻訳者よりも広い範囲の翻訳者を対象としたもので，無料で使うことができる．

産業翻訳に翻訳支援システムが浸透すると同時に，オンライン翻訳やソフトウェアのローカライゼーションといった多様な翻訳の現場が目に見えるようになってくるなかで，基本的な翻訳支援システムの機能は備えつつ，新たな翻訳作業に特化したシステムも提案され始めている．オンライン文書の翻訳に特化した翻訳支援システム/翻訳文書ホスティングシステムとして公開されている「みんなの翻訳」[15]はその一つで，オンライン翻訳者が辞書引きとウェブ検索に翻訳時間の多くを割いていることに注目し，高品質辞書の検索から Wikipedia 検索，ウェブ検索までをシームレスに行うことができる翻訳環境を提供している．

これら複数の翻訳支援システムが実利用されているなかで，翻訳データを複数の支援システム間で相互運用するために，TMX（Translation Memory eXchange）や XLIFF（XML Localisation Interchange File Format）といった，ベンダーに依存しないオープンな XML 標準が策定されている．こうした標準化を通して機能面ではデータなどの互換性が進む一方，応用システムとしては，今後，「翻訳」と一括りにするのではなく，産業翻訳支援，オンライン翻訳支援などなど，多様に存在する翻訳の類型に特化した翻訳支援システムの開発が進むことが予測される．

なお，翻訳支援システムの普及とボランティア翻訳者の登場は，翻訳という産業に対して様々な影響を及ぼす可能性があり，現に及ぼしつつある．翻訳支援システムの導入により時

間当たりの翻訳量が増えた場合，翻訳者の時間単価が上がるわけではない．例えば，翻訳メモリに収録された訳文と合致している訳文は合致率に従い翻訳料金を割り引く契約なども存在し，クライアントと翻訳会社側には時間短縮のほか経済的メリットがあるが[16]，熟練翻訳者の活動と養成が中長期的にどのような影響を受けるかは現在のところ明らかになっていない．また，ボランティア翻訳者の登場により，産業翻訳者とボランティア翻訳者ではほとんどの場合，翻訳対象が全く異なるにも関わらず，産業翻訳まで無料化に近いのが当然であるという風潮が広まる恐れもなくはない．翻訳支援システムの活用はあくまで翻訳者あってのことであり，技術開発に関わる研究者も，応用システムの社会的な影響や効果を考慮する必要がますます高まっていると言えよう．

■参考文献

1) M. Kay：“The proper place of men and machine in language translation,” Machine Translation, 12, pp.3-23, 1997.（元論文は 1980 年に Xerox PARC Working Paper として公開配布された）
2) F. Gow：“Metrics for Evaluating Translation Memory Software,” PhD Thesis, University of Ottawa, 1993.
3) E. Macklovitch：“TransType2: The last word,” LREC2006, pp.167-172, 2006
4) K. Takeuchi, et. al.：“Flexible automatic look-up of English idiom entries in dictionaries,” MT Summit XI, pp.451-458, 2007.
5) 熊野　正, 西脇正通, 田中英輝：“「翻訳パレット」を用いた翻訳支援の提案,” 言語処理学会第 12 回年次大会, P6-5, 2006.
6) 大倉清司, 富士　秀,　徐国偉, 長瀬友樹, 潮田　明：“Cliche: さらなる翻訳効率化のための翻訳支援インターフェース,” P6-4, 2006.
7) http://www.translationzone.com/jp/
8) http://www.star-group.net/star-www/description/transit/star-group/eng/star.html
9) http://www.tratool.com/
10) http://www.atril.com/
11) http://software.fujitsu.com/jp/atlas/
12) http://pf.toshiba-sol.co.jp/prod/hon_yaku/index_j.htm
13) http://www.omegat.org/
14) http://translate.google.com/toolkit/
15) http://trans-aid.jp/
16) http://kyutoku.cocolog-nifty.com/translation/2006/06/post_1cad.html

■2 群 - 10 編 - 4 章

4-6 機械翻訳の活用

（執筆者：潮田　明）[2009 年 8 月 受領]

機械翻訳の研究開発の最終目標は完全自動翻訳の実現であるが，翻訳とは本来人間の持つ様々な言語知識，世界知識や，推論などの知識処理機能を総動員して行われる高度な知的作業であり，それを機械に行わせる試みはどこまで行っても「完成」といえるレベルに達することはないと言えるほどゴールへの道のりは遠い．翻訳精度向上への探求は今も日本を含む世界各国で地道に行われているが，一方で，精度的には不完全ながら役に立つ機械翻訳の使い道が模索されている．ここでは，これまでに行われてきた機械翻訳の様々な活用法のなかから，有効な活用法の代表例として，翻訳支援，多言語コミュニケーション支援及び音声翻訳について概説する．

4-6-3 機械翻訳と翻訳支援

現在コンピュータを利用した翻訳技術のうち，翻訳の現場で実際に活用されている技術には，大きく分けて 2 つの流れがあると言える．一つは，機械翻訳結果を活用するアプローチ，もう一つは，実際の訳例を多数集めてデータベースを構築し，翻訳の現場でそれらを検索しながら再利用しようというアプローチで，「翻訳メモリ」と呼ばれている．

前者の機械翻訳は，英語があまり得意ではないコンシューマレベルのユーザが英文を読むためのツールとしては，比較的有効な情報収集の手段となってきてはいるが，まだプロの翻訳者が下訳として利用できるレベルの訳文形成には至っていない．一方，後者の翻訳メモリは，完全な翻訳を求めるのではなく，「翻訳支援」という考え方のもとに，文レベルあるいはフレーズレベルの辞書引きができればよいというものであり，技術的には機械翻訳よりも格段にシンプルではあるが，少なくとも翻訳業界には機械翻訳よりも遥かに浸透してきている．もっとも，翻訳メモリもその用途は，コンピュータソフトのローカライズや製品マニュアルの改版など，過去の翻訳文の繰り返しの多い翻訳業務に限られており，産業翻訳全般に役立つ技術とはなっていない．

機械翻訳が長い研究開発の歴史を持ち，翻訳メモリに比べて遥かに高度な技術を駆使しているにも関わらずまだまだ翻訳の現場では主流のツールになり得ていないのは，翻訳品質が十分でないという理由もあるが，これまで機械翻訳が「完全自動翻訳」を大前提として開発が行われ，ユーザ中心の視点に欠けていたという点も大いに関係していると言える．

そこで，あくまで機械翻訳もユーザ（例えば翻訳者）の手助けをするツールの一つという観点から，機械翻訳単独ではなく，辞書引きツールや翻訳メモリなど他のツールと組み合わせた「翻訳支援システム」として開発を進める新しい動きが企業や研究者の間で盛んになってきている．翻訳支援システムの具体的構成は，翻訳メモリ中心のものや，機械翻訳中心のものなど様々であるが，共通しているのはユーザ主体の視点であり，システムは補助手段に過ぎないという点である．そのため，従来の機械翻訳の研究開発ではあまり重視されなかった UI やツールとしての使い勝手という要素が研究開発の重点項目として登場することになる．

システムの評価においても，従来の機械翻訳の場合，出力文の品質を 1 文ずつ評価して翻訳精度何%と数値化していたが，翻訳支援システムの評価の場合は，翻訳の作業効率という点から総合的に評価するなど，新しい指標が必要になってきている．作業効率の評価は，例

えば，一人の翻訳者，あるいは翻訳チームが一定量の翻訳をシステムを用いて行ったときの，全翻訳文の質と全翻訳工程にかかった時間などをもとに行われる．前者と後者のシステム評価の根本的違いは，前者が純粋にシステム自体の評価であるのに対して，後者はシステムを使ったユーザのパフォーマンスの評価であり，ユーザによって評価結果が異なる点にある．

4-6-4 多言語コミュニケーション支援

上述のようにユーザの視点に立つと，機械翻訳に対して自動翻訳装置から翻訳環境の主要機能へと視点を変えることが有用だが，また同時に，単一装置からネットワーク化されたコラボレーション環境の担い手へと見方を広げることも可能である．計算機環境とネットワークの発達により従来単体ソフトとして使われてきた翻訳システムも，最近では以下の例のようにインターネットなどのネットワークで繋がれた様々な言語リソースと連携して使われようとしている．これらの言語リソースには，辞書や訳例などのコンテンツ，及び辞書引きツール，訳例検索装置などの言語処理機能などが含まれる．

機械翻訳を含む各種言語処理機能と辞書や訳例などのコンテンツが，ネットワークを通じて共有可能な形で提供された初期の事例として，総務省「アジア・ブロードバンド計画」のコアプロジェクトの一つである「国際情報通信ハブ形成のための高度 IT 共同実験（2003～2005 年度)」が挙げられる．本共同実験では，自動翻訳/翻訳支援機能を活用した多言語翻訳プラットフォームの基盤構築や，機械翻訳機能を活用したコミュニケーション環境，いわゆる「異文化コミュニケーション環境」の構築が行われた．

多言語翻訳プラットフォームとは，既存の機械翻訳エンジン及びその要素機能を部品として利用し，更に曖昧訳例検索など他の言語処理機能と連携することにより，実用的な翻訳環境を提供するための枠組であり，アジア諸国言語を対象とした機械翻訳システムの共通基盤として設計された．日本と中国を結ぶ専用光ファイバ通信網を用いた実証実験（2003 年 12 月）においては，日中間で企業情報交換を行うための企業情報フォーラムが多言語翻訳プラットフォーム上に構築され，プラットフォームの有効性が確認された．また，本共同実験では同じ専用回線を用いて，異文化コミュニケーション環境の実証実験が，京都大学が中心となって行われた．本実証実験においては，チャットやビデオ会議などコラボレーションの相手と直接対応しながら同時に作業する「同期型多言語コラボレーション」環境と，同一情報を共有しながらも，相手と別々に作業する「非同期型多言語コラボレーション」環境が構築され，機械翻訳機能を用いたコミュニケーションの成立過程の評価などが行われた．

京都大学では，情報通信研究機構などと共同で上記異文化コミュニケーション環境を更に発展させた「言語グリッド」の開発を進めている．「言語グリッド」は，インターネット上の多言語サービス基盤の構築を目指すもので，「インターネット上の言語資源（対訳辞書など）や言語処理機能（機械翻訳など）を自由に組み合わせて使うことができ」そして「自分たちのコミュニティが作った言語資源を追加し，自分たちの活動のための言語サービスを容易に作り出すことができる」ことを目標として掲げている．このような機械翻訳を軸とする翻訳環境の発展は，将来的には集合知という形で機械翻訳にフィードバックされ，更なる機械翻訳精度の向上へと繋がっていくものと期待される．

4-6-5 音声翻訳

機械翻訳技術は，また他の情報処理技術と組み合わせることで用途を大きく広げることができる．その代表例が音

声技術との連携であり，音声認識→機械翻訳→音声合成と繋げることで自動通訳システムが実現できる．我が国では，1986 年 4 月に，音声翻訳実現を目指した初の産学官連携の組織として自動翻訳電話研究所（ATR）が設立され，音声翻訳の基礎研究及び実用化研究が進められてきた．最近では，携帯電話向け翻訳サービスなど実用化に向けた取組みが行われている．ATR 方式を始め開発初期の音声翻訳は，機械翻訳の方式として辞書や文法規則に基づくルールベース翻訳が用いられており，統計処理をベースとする音声認識技術との連結が大きな課題であった．ルールベース翻訳方式は今日でも日英・英日翻訳においては殆どの実用システムで採用されている方式であり，かつテキスト翻訳の方式としては今なお最も翻訳精度の高い方式であるといえるが，音声翻訳用に用いるには多くの課題を抱えていた．

まず，文法規則はもともと文法的に正しいとされる表現を対象に体系化されたものなので，省略表現や言い間違え，言いよどみなどを多く含む話し言葉の翻訳に対しては適用が難しく，また新たに会話文用に文法規則を補充しようとしても，多様な会話表現を網羅的にカバーすることは非常に難しい．文解析の際に適用可能なルールが 1 つも見つからない場合は翻訳自体が失敗することもあり得る．更に，音声認識と機械翻訳を連結させるうえでの本質的な課題として，エラーの伝播・増幅の問題がある．単純化した例を挙げると，仮に音声認識と機械翻訳単体の精度がそれぞれ 80 % だとしても，音声認識と機械翻訳を連結させたシステムの精度は通常 64 % に届かない．なぜなら機械翻訳の精度 80 %というのは，100 % 正しいテキストに対して 100 文中 64 文正しく翻訳するという意味であって，もともと 20 % のエラーを含んだ音声認識結果に対しては，エラーの伝播・増幅が生じ，通常 80 % の精度は出せないからである．

一方，2000 年代に入り統計翻訳（SMT）の研究開発が進展してくると，音声翻訳の大きな課題の幾つかが解決あるいは緩和されるようになってきた．まず，統計翻訳ではあらゆる翻訳候補にゼロより大きい評価値が割り当てられるため，どのような入力に対しても必ず何がしかの翻訳結果を返すことが可能である．また，音声認識と機械翻訳の双方が同じ統計的枠組みのなかで定量的に評価できるため，音声認識の複数候補とそれぞれに対する機械翻訳の複数候補を同時に評価して最適な組合せを探索することが可能になってきており，このことにより，エラーの伝播・増幅の抑制が可能となる．

音声翻訳において大変効力を発揮している統計翻訳であるが，学習用データである対訳文が大量に必要なこと，学習用データとは全く異なる分野の翻訳は苦手であること，日英翻訳のように全く言語構造や語順の異なる言語間の翻訳においては長文翻訳が困難であること，など固有の問題も抱えており，将来的には従来のルールベース翻訳と統計翻訳の長所をミックスしたハイブリッド方式が有望視されている．

携帯電話上でのアプリケーションとしては，音声翻訳のほかに，OCR 連携翻訳もメーカなどで開発が進められている．これは，携帯カメラで写した文字画像を OCR ソフトでテキスト化したうえで翻訳を行うもので，外国でのレストランメニューの翻訳などの応用が考えられている．

■参考文献

1) 石田 亨：“言語グリッドと異文化コラボレーション,” 電子情報通信学会誌, vol.91, no.6, pp.515-517, 2008.

2) A. Fujii et. al.：“Overview of the Patent Translation Task at the NTCIR-7 Workshop,” Proceedings of NTCIR-7 Workshop Meeting, December 16-19, 2008, Tokyo, Japan.
3) P. Koehn：“Pharaoh: a Beam Search Decoder for Phrase-Based Statistical Machine Translation Models,” in 6th Conference of the Association for Machine Translation in the Americas, AMTA, 2004.

■2 群 - 10 編 - 4 章

4-7 対訳用語抽出

（執筆者：宇津呂武仁）[2009 年 8 月 受領]

機械翻訳の技術において，対訳辞書は最も重要な言語知識の一つである．通常，英語以外の主要な言語においては，主に英語への翻訳及び英語からの翻訳のための対訳辞書が人間の手によって整備されている．それらの対訳辞書においては，一般語の収録語数は十分であると言えるが，多様な専門分野における専門用語については，十分な収録語数であるとは言えない．特に，学術分野の発展や文化・風俗の変容に伴って，新しい専門用語が造られるため，これに追随して対訳辞書を人手によって整備していくことは容易ではない．また，言語によっては，単言語辞書や英語との間の対訳辞書などの言語資源の整備が十分でない言語も多数存在し，それらの言語においては，専門用語に限らず一般語のための辞書資源の整備が大きな課題となっている．

これに対して，機械翻訳分野においては，電子化された多言語テキストを情報源として，翻訳のための対訳辞書資源を獲得する手法の研究が行われている．それらの研究は，多言語の言語資源が整備されはじめた 1990 年代初頭から開始された．更に，近年では，ウェブ上において，世界中の多様な言語，及び，多様な分野・ジャンルの文書が収集可能となったことを背景として，手法の適用範囲を広げている．

それらの研究において，初期の頃から最もよく研究されたのが，図 4･1 に模式的に示す対訳テキスト（二言語間で内容が対応したテキスト．図 4･1 のように，文単位での対応がとれている場合が，最も良質の対訳テキストである) を情報源として，

> 二言語間での出現位置の相関が強いほど，それらの表現の組は二言語間で対訳関係にある可能性が高い．

という考え方に基づいて，対訳表現を獲得する手法である [?]．

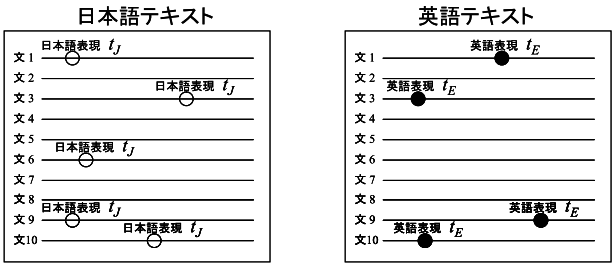

図 4･1　対訳テキスト中の対訳表現

図 4･1 における日本語表現 t_J，英語表現 t_E を，それぞれ x 及び y とみなして，分割表

	y (y が出現する)	$\neg y$ (y が出現しない)
x (x が出現する)	$f(x,y)$	$f(x,\neg y)$
$\neg x$ (x が出現しない)	$f(\neg x,y)$	$f(\neg x,\neg y)$

に従って，二言語間で対応する文対において，x と y がともに出現した頻度 $f(x,y)$，片方のみ出現した頻度 $f(x,\neg y)$，$f(\neg x,y)$，ともに出現しなかった頻度 $f(\neg x,\neg y)$，をそれぞれ測定する．そして，これらの頻度を用いて，相互情報量，Dice 係数，ϕ^2 統計，対数尤度比など，多様な統計的共起尺度を用いて，x と y が対訳表現である度合いを推定する．

この方法は，良質な対訳テキストが利用可能な場合には極めて有望であるが，現実には，多様な言語対・専門分野にわたって，そのような良質な対訳テキストが利用できることは稀である．そこで，情報源となるテキストに対する条件を緩めて，「二言語間で内容が対応したテキスト」ではなく，「二言語間で分野や話題がよく似たテキスト」を情報源とする手法の研究も行われてる [?]．この条件を満たすテキストコーパスをコンパラブルコーパスと呼ぶが，コンパラブルコーパスからの対訳辞書獲得研究において用いられた代表的手法が，以下に示す文脈ベクトルを用いる方法である．この方法では，既存の対訳辞書に含まれていない専門的な用語の周囲に出現する一般語の頻度ベクトル表現を作成し，既存の対訳辞書に含まれる一般語の対訳関係を利用して，これらの文脈ベクトルの二言語間における類似性を測定し，対訳辞書の獲得を行う（以下の例では，「黄色ブドウ球菌」の訳語として「Africa」が排除され「Staphylococcus aureus」が選ばれる）．

「黄色ブドウ球菌」の文脈ベクトル
= { 食中毒/147, 感染/90, 皮膚/84, タンパク質/69, 症候群/53, . . . }
「Staphylococcus aureus」の文脈ベクトル
= { infection/170, illness/26, spread/17, skin/17, virus/13, . . . }
「Africa」の文脈ベクトル
= { South/109, African/32, China/20, ties/15, diplomatic/14, . . . }

これらの基本的技術のほかにも，図 4･2 に示すように，複合語の構成要素の訳語を連結して訳語候補を生成するモジュール（要素合成法）と，生成された訳語候補を，目的言語のコーパスを用いて検証するモジュールを併用する手法 [?]，図 4･3 に示すように，一つの言語（例えば日本語）のウェブページにおいて，重要な専門用語の英訳語が併記されている部分を収集して対訳辞書の獲得を行う手法 [?]など，対訳テキストやコンパラブルコーパス以外の情報源を積極的に活用して対訳辞書の獲得を行う技術が研究されている． 更に，近年，Wikipedia における多言語辞書資源・言語資源を多様な言語処理・知識獲得に活用する研究の流れが活発になっているが，そのなかでも，Wikipedia 中の多言語ページを利用した対訳辞書獲得の研究が幾つか行われている [?]．それらの研究では，Wikipedia 中の言語間リンクとして記述されている対訳辞書知識の範囲を越えて，二言語のエントリーの間の翻訳関係を推定することにより新規の対訳関係を獲得する．

また，そのほかにも，統計的機械翻訳モデル（本章 4–3 節参照）の適用が可能であるような大規模な対訳テキストが利用できる言語対・分野において，統計的機械翻訳モデルの学習によって得られる翻訳モデルを精選することにより，対訳辞書の獲得を行う技術 [?]や，日英

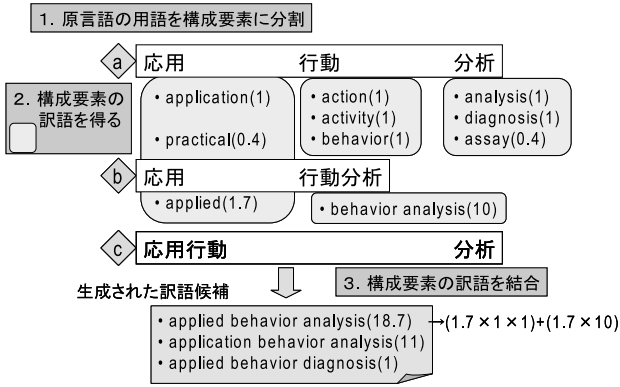

図 4･2　要素合成法による訳語推定

図 4･3　日本語ウェブページにおける日本語用語・英訳語併記の例（黄色ブドウ球菌/Staphylococcus aureus）

対訳辞書及び中英対訳辞書のエントリーに対して統計的機械翻訳モデルを適用することにより，日中対訳辞書を獲得する技術 [7)] の研究なども進められている．

■2 群 - 10 編 - 4 章

4-8 翻字