5 対訳テキストから辞書を自動生成

春野雅彦
ATR人間情報通信研究所

対訳テキストの利用と必要な処理

社会の国際化に伴い雑誌，新聞等で翻訳された記事を読んだり，翻訳されたマニュアル，論文等に目を通す機会が増えている．近年このような対訳文書の多くはコンピュータ上で作成されるため電子的に利用できることが多い．これら電子化された対訳テキストをうまく処理して利用することができれば，分野に依存した特殊な表現や，細かい意味の違いを実際のデータに基づいて理解することが可能となり，我々が英文を読んだり作成したりする際に大きな助けになる．また，対訳テキストから抽出される知識は機械翻訳システムの精度向上にも役立つことが期待される．

このような対訳テキストの利用例として図-1に筆者らが作成した適応的辞書システムAIDA[1]の画面を示す．このシステムでは対訳テキストから自動的に対訳表現対を取り出し（右側の画面），その表現を含む例文を検索することができる（左側の画面）．図-1の例は‘アルツハイマー病患者-Alzheimer patients’を含む例文を検索した結果である．AIDAでは対訳テキストから得られる項目と既存の電子辞書とが同一ウィンドウ上でリンクされるためユーザは自分の好みの対訳テキストを用いて既存の辞書をカスタマイズすることができる．

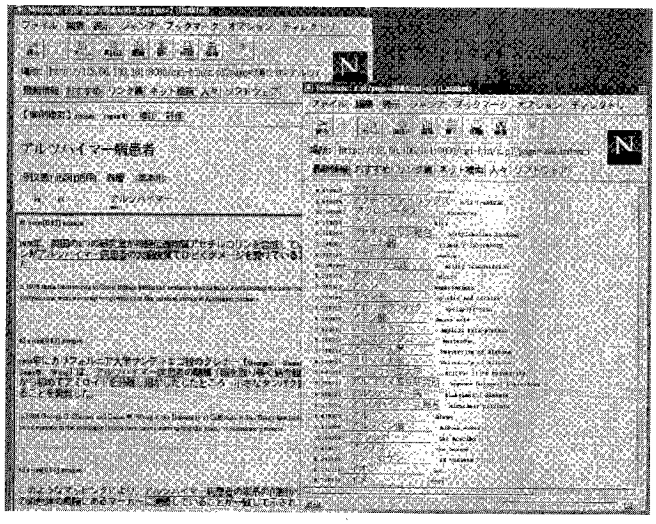

図-1 適応的辞書環境AIDA

では実際に対訳テキストを処理して利用するにはどのような技術が必要になるのであろうか．ここでは主な対象を日本語と英語の対訳テキストに絞り図-2に基づいて話を進めよう．入力される日本語，英語のテキストに対してまず形態素解析を行う．形態素解析は入力文を単語に分割し，それぞれの単語に品詞を割り当てる．日本語では単語間に明示的な境界がないため，形態素解析が特に重要となる．形態素解析による情報は以後の各段階の処理で利用される．

与えられたテキストはお互いに対訳となっていることは分かっているが，どの文がどの文に訳されているかの文対応は与えられていない．対訳テキストを有効に利用するには図-1の左側画面のように，文レベルでの対応を知る必要がある．現実の対訳テキストにおいては，一方の言語の1文が他方の言語では2文や3文に分割して翻訳されていたり，逆に完全に省略されてしまう場合もあるため文体応付けが重要な研究課題となってくる．

次に対訳テキストを辞書的に使用したり機械翻訳システムで利用するには，より詳細な文内の表現レベルの対応が必要となる．図-1の右側画面の例は日経サイエンスの対訳テキストを文体応付けした後，自動的に抽出した対訳表現対である．以下では自動文対応付けと対訳表現抽出の技術について主に筆者らがNTTで行った研究を中心に紹介することにしよう．

対訳テキストの自動文体応付け

自動文対応付けの手法は大きく2種類に分類で

図-2 対訳テキストに対する処理の概要

きる．1つは文内に含まれる単語数[2]や文字数[3]に基づく方法，他方は2言語間の単語の対応関係を利用する語彙に基づく方法である．

前者の手法は**一方の言語で長い文は翻訳された言語でも長くなる**という仮定の下に単語数や文字数を用いた文対応の条件付き確率を記述し，動的計画法を用いて全体の対数尤度を最大にする文対応関係を決定する．この手法で対象となるのは英語-フランス語などの構造的に類似した言語間の硬い翻訳に限られる．これは日本語-英語などの構造が異なる言語間では文字体系，文法体系の違い，さらには，形態素解析システムの違いによって上記の仮定が高い精度では成り立たなくなるためである．

後者の語彙に基づく方法は，（'本' と 'book' のように）**対応する単語対を多く含んでいる文どうしは対応する可能が高い**という仮定に基づいている．ここではまずKay等による反復法[4]を説明し，続いて筆者らが日英対訳テキストの文対応付けを行うために既存の対訳辞書を用いて反復法を拡張した文体応付けシステムBACCS[5]について述べることにする．

図-3に反復法の概略を示す．図のグラフ中で縦軸，横軸はそれぞれ英語テキスト，日本語テキストの文番号を示す．システムの初期状態では両言語の最初と最後の文の対応のみが仮定されている．このように文対応がすでに仮定された点をアンカーと呼ぶ（図中では黒丸で示す）ことにすると，自動文対応付けはアンカーを順に見つける操作に相当する．反復法では以下の3ステップを繰り返すことで確信度の順にアンカーを見つけていく．

第1ステップ 現在のアンカーから文対応可能なペア（図-3でアンカーと曲線で囲まれた領域の格子点に相当）を構成する．

第2ステップ 第1ステップの文対応可能ペアから統計的に訳語対を抽出する．

第3ステップ 文対応可能ペアのそれぞれが含む訳語対の数を累計する．その値が現在の閾値を超えたペアを新たなアンカーに設定し，閾値を下げる．

第1ステップで現在のアンカーから文対応可能なペアを構成する．直観的には可能な文対応の範囲はアンカーの近くで小さく，アンカーから離れるほど大きくなる☆1．

第2ステップでは第1ステップで構成した現在の文対応可能ペアから統計的に訳語対を求める．この訳語対は第3ステップで新たなアンカーを設定するのに利用される．ここでの処理は**出現の様子が近い日英の単語どうしは訳語対である**という仮定に基づいている．対訳コーパス中に現れる任意の日本語単語w_{jpn}と英単語w_{eng}の出現の類似度を評価するために表-1に示すような分割表を考えよう．分割表の構成要素は（a）w_{jpn}とw_{eng}の両方が現れた文対応可能ペアの日本語文数，（b）w_{eng}のみが現れた文対応可能ペアの日本語文数，（c）w_{jpn}のみが現れた文対応可能ペアの日本語文数，（d）w_{jpn}とw_{eng}のどちらも現れない文対応可能ペアの日本語文数，の4種類である．

もしw_{jpn}とw_{eng}が正しい訳語対であれば出現の仕方が近いのでaが大きく，bとcが小さくなるはずである．逆にw_{jpn}とw_{eng}が訳語対でなければaが小さく，bとcが大きくなるはずである．これらを定量的に評価するために（1）式の（自己）相互情報量を利用する．相互情報量の値は各単語が独立に現れると仮定した場合の両単語の共起確率と実際の共起確率の大小関係を示している．この相互情報量の値が閾値☆2を超えた単語対は訳語対であると認定する．

$$\log \frac{prob\,(w_{jpn}, w_{eng})}{prob\,(w_{jpn})\,prob\,(w_{eng})} \qquad (1)$$

$$prob\,(w_{jpn}) = \frac{a+b}{a+b+c+d}$$

$$prob\,(w_{eng}) = \frac{a+b}{a+b+c+d}$$

$$prob\,(w_{jpn}, w_{eng}) = \frac{a}{a+b+c+d}$$

第3ステップでは決められた個数（閾値）以上の訳語対を含む文のペアを新たなアンカーに設定する．新しく見つかったアンカーを利用することで，次回の文対応可能ペアの範囲がより絞り込まれる．

以上の3ステップを閾値の緩和を行いながら繰り返すことで，確信度（含まれる訳語対の数）の順にアンカーが決定される．緩和法を利用することでアルゴリズムの終盤で起こる対応付け誤りが，全体の対応付け精度に及ぼす影響を少なくすることができる．語彙に基づく方法では利用できる訳語対の数がテキスト中の場所によって大きく異なるため，単語数や文字数に基づく手法で用いた動的計画法よりも緩和法が適しているのである．

☆1 2つのアンカーの中央辺りでは1文に対して$O(\sqrt{L})$（Lは2つのアンカーの間に存在する文数）の文を対応付ける．これは統計学的に分散が$O(\sqrt{L})$でモデル化できるからである[4]．

☆2 相互情報量は出現頻度の低い事象に対して大きくなる性質があるので頻度の小さい単語はあらかじめ除外している．

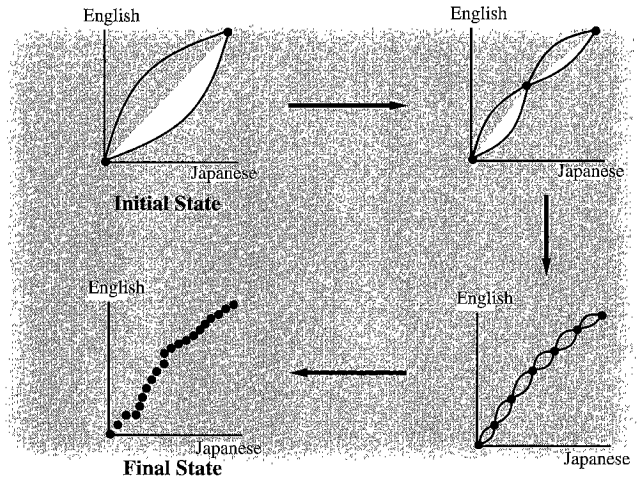

図-3　反復法による文対応付けの過程

	W_{jpn}	
W_{eng}	a	b
	c	d

表-1　単語対の分布に関する分割表

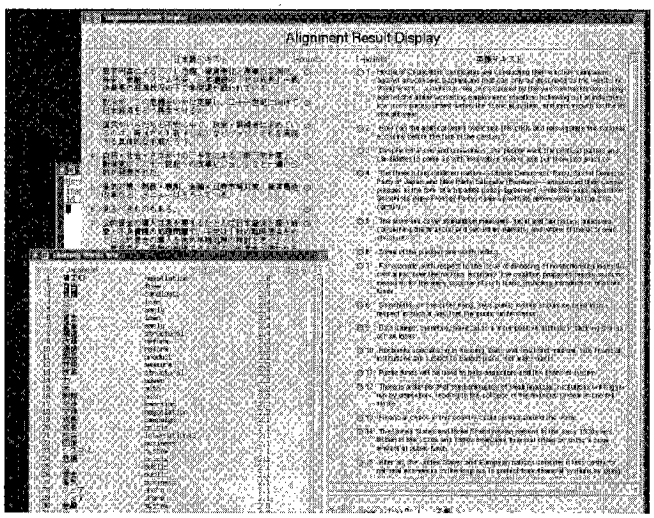

図-4　統合的文対応付け環境BACCS

反復法を日英の対訳コーパスに適用すると第2ステップで統計的に得られる訳語対の数が十分でないため文対応付けが部分的にしか得られないことが多い．筆者らが作成した文対応付けシステムBACCS（図-4）[5]では，統計的に得られる訳語対（図-4左側の画面）と既存の対訳辞書に含まれる訳語対の両者を利用した反復法で文対応付けを行っている．具体的には上記の第2ステップを以下のように変更している．既存の対訳辞書の利用により，統計的手法により得られる分野依存の訳語対（図-4左側）と同時に，コーパス中に1回しか現れない訳語対も利用可能となる．その結果図-4右側の画面に示すように，1対2や2対1等の対応も含めて高い精度で文対応付けを行えるようになった．

第2ステップ　文体応可能ペアから統計的な訳語対と既存の対訳辞書に登録されている訳語対を抽出する．

次に，文対応付けの副産物として統計的に得られる訳語対について検討してみよう．**表-2**に日経サイエンスの脳科学に関する記事で相互情報量の高かった統計的訳語対を示す．表中の‘NMR’，‘MEG’，‘PET’，‘CT’，‘MRI’，‘Functional MRI’はどれも脳を外部から測定するための装置名であり，このテキストのキーワードである．これらの専門用語は通常の対訳辞書には含まれていない．また‘MEG’の正しい日本語訳は‘脳磁図’であるが，形態素解析システムがこれらを脳，磁，図に分解してしまっている．システムは統計的な類似性を考慮することにより‘磁’や‘図’と‘MEG’の対応を正しく捉えている．このように統計的訳語対には既存の辞書に含まれない専門用語や形態素解析を誤る未知語も文対応付けに利用できる利点がある．

対訳表現抽出

この章では対訳テキストから対訳表現を抽出することを考えよう．主に文対応がすでに与えられている場合を考えることにする．**図-5は日英対訳テキストの日本語表現○と英語表現●のテキスト中での出現パターンが似ていればこれらは対訳表現である**という対訳表現抽出の基本的な考え方を示している．このことから対訳表現抽出の問題は，日本語表現○と英語表現●としてどの範囲を考慮するかという問題と，出現の類似性をどう評価するかの問題に分けて考察できる．たとえば，前章の文対応付けの過程で副産物として得られた表-2の訳語対は表現として単語レベル，類似度として相互情報量を用いたものである．単語レベルを選択したことで‘ファンクショナル–functional’，‘MRI-MRI’という対応は抽出されるが‘ファンクショナルMRI-functional MRI’という単語列の対応を得ることは原理的にできない．このように，対訳表現抽出においてはその用途に応じて適切な表現と類似度を選択することが重要となる．

対訳表現抽出の対象となる表現とそれぞれに必要な前処理の種類を次にまとめる．もちろん前のものほど構造が単純で抽出が容易である．

(1) 単語レベルの対応[6]．形態素解析だけが必要となる．

(2) 名詞句レベルの対応[7), 8)]．形態素解析に加え

日本語	英語	相互情報量	日本語	英語	相互情報量
記録	recording	3.58	ＰＥＴ	PET	3.10
リアルタイム	real	3.51	解像度	spatial	3.10
ニューロン	neuron	3.51	そのようだ	such	3.10
フィルム	film	3.51	代謝	metabolism	3.06
グルコース	glucose	3.51	言う	verb	3.04
増加	increase	3.51	科学者	scientist	2.95
磁	MEG	3.51	地図	mapping	2.92
解像度	resolution	3.43	大学	university	2.92
電気	electrical	3.43	思考	thought	2.90
グループ	group	3.39	化合物	compound	2.82
電気	recording	3.39	標識	label	2.82
記録	electrical	3.39	オートラジオ	radioactivity	2.77
言う	generate	3.33	視覚	visual	2.77
提供	provide	3.33	信号	signal	2.77
図	MEG	3.33	リアルタイム	time	2.69
ＮＭＲ	NMR	3.17	オートラジオ	autoradiography	2.67
ファンクショナル	functional	3.17	能力	ability	2.63
機器	equipment	3.17	ＣＴ	CT	2.63
臓器	organ	3.15	聴覚	auditory	2.15
注射	compound	3.10	心	mental	2.05
水	water	3.10	ＭＲＩ	MRI	1.8
標識	radioactive	3.10			

表-2　文対応付けの過程で統計的に得られた訳語対

名詞句識別プログラムが必要となる．

(3) 一般の連続した単語列[9), 10)]．候補となる単語列を選ぶための識別プログラムが必要となる．

(4) 離れた単語列が構成する表現[9), 11), 12)]．実用には高精度の構文解析システムが必要となる．

また，出現パターンの類似度の計算に利用されるものとして以下のものがある．これらのうち複数を組み合わせることもよく行われる．

(1) 対応確率[6), 7), 19)]．EMアルゴリズム等を用いて計算される．

(2) 統計的ヒューリスティクス．共起頻度，相互情報量[9)]，Dice係数[10), 12)] など．

(3) 既存の日英対訳辞書を用いたヒューリスティクス[8), 11)]．

次に，対訳表現抽出法の一例として筆者らの行ったソーティングを用いた方法[9)] を説明する．主に対象とする表現は一般の連続した単語列（上記(3)）で類似度として統計的ヒューリスティクス（上記 (2)）の一種である相互情報量（(1) 式）を利用する．連続した候補単語列を効率よく抽出するためにソーティングを用いる．

あらかじめ形態素解析，および文対応付けされた日英各テキストに対して図-6左側のように単語ごとのポインタを設定する．個々のポインタは，文中のその場所以降の部分文字列（suffix）を表現している☆3．テキスト中での文字列の出現頻度を調べるため，このポインタをソートし図-6右側に示す文字列の出現頻度テーブルを作成する．図中のstring，coincidenceはそれぞれソートされた文字列および隣り合う文字列が前から何文字一致しているかを示す．

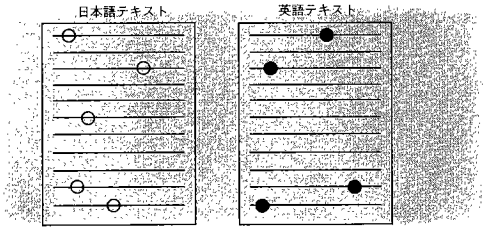

図-5　対訳テキストを用いた対訳表現の抽出

このcoincidenceを利用して各単語列がテキスト中に何回出現したかをカウントする．基本的にあらかじめ決められた回数以上出現した単語列が候補単語列となる．ただし，この段階で形態素情報により相応しくないものを取り除く☆4．次にもう一度出現文番号（sent no.）でポインタをソートし，文内最長一致の原則で最終的な（adopt=1）日英の候補単語列w_{jpn}，w_{eng}を抽出する．

w_{jpn}，w_{eng}に関して文対応付けの場合の表-1と同様な分割表を考えよう．この場合，aはw_{jpn}，w_{eng}の双方が現れた対応文ペア数，bはw_{jpn}だけが現れた対応文ペア数，cはw_{eng}だけが現れた対応文ペア数，dはどちらもが現れなかった対応文ペア数ということになる．w_{jpn}とw_{eng}の統計的類似度として (1) 式の相互情報量を用い，徐々にその閾値を下げることで連続した単語列から構成される対訳表現を抽出できる．

表-3に日経サイエンスの脳科学に関する1766ペアの対訳テキストから抽出した対訳表現例を示す．表から分野に依存した専門的単語列が抽出されて

☆3 実際には，表現の先頭として相応しくない品詞を指すポインタはこの段階で切除され必要なメモリが削減される．

☆4 候補単語列として相応しくない文字列は各言語約50の正規表現規則で記述されている．

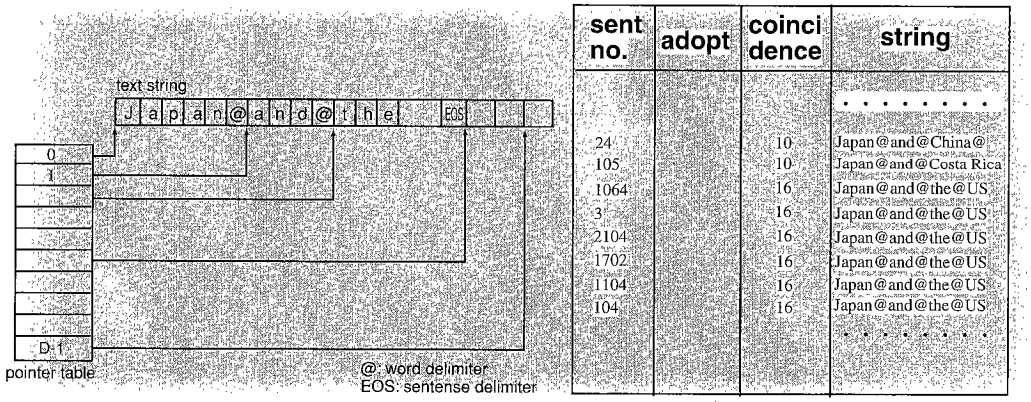

図-6　単語レベルのソートによる候補単語列の抽出

いることが分かる．科学記事という性格上，名詞句が多く抽出されている．これらを表-2の単語対応と比べると面白い．対訳表現抽出の精度に関しては最終的に929個の対訳表現が得られ，その88.5%が正しい対訳表現となっていた[13]．

次に上記の方法で得られた連続した単語列の対応表現を組み合わせることでより獲得が困難な離れた単語列の対応表現を得る試みも行った[9), 13)]．株式分野の対訳テキスト1114文で実験を行ったところ87個の対訳表現が得られその36.8%だけが正しいものであった．この結果から連続した単語列と離れた単語列では抽出の難しさが大きく異なり，離れた単語列で構成される表現に関しては形態素解析と正規表現によるパターンだけでは十分な抽出が行えないことが明らかになった．これらの対訳表現の抽出には日本語，英語に対する高精度の構文解析システムと新たな類似度の基準が必要となろう．

今後の展望

1980年代末から始まった対訳テキストを用いた自然言語処理研究では，対訳テキストの増加に伴い，文対応付け，対訳表現抽出を中心に多くの手法が提案され，その対象言語も広がってきた．代表的な対訳テキストとしてはLiguistic Data Consortium（LDC（http://morph.ldc.upenn.edu/））から入手可能なHansard French/English，UN Parallel Textや，英中対訳テキストであるHong Kong Hansard[14)]がある．日本でも電子協を中心に環境，経済，科学技術各白書の電子化ならびに文対応付与の作業が進められている[15)]．また，各新聞社のホームページに掲載されている社説なども興味深い対訳テキストの一例である．今後インターネット，CD-ROMなどを介してより多くのテキストが利用可能となるであろう．

このような対訳テキストの増加に伴い，テキストにノイズが含まれていたり，一方のテキストが要約になっていたりすることも多くなっている．これまでの研究で対象とされたテキストは比較的きっちりした対応を持つ逐次訳的なものが多かったが，今後はよりロバストな手法の研究[16), 17)]とともに，どの範囲の対訳テキストなら有効に利用できて収集価値があるかを見極めることも重要な課題となろう．また，紙数の関係で本稿では触れることができなかったが，対訳テキストを用いて自動翻訳システムを構築する研究も盛んに行われている[18)～20)]．この方向の研究は近年の統計的構文解析手法[21), 22)]の進歩により，前章で触れた離れた単語列から構成される表現列の抽出とも関連して今後ますます発展するものと思われる．

本稿では対訳テキストから辞書を自動生成する際に必要となる文対応付けと対訳表現抽出の技術について我々の仕事を例に紹介してきた．社会の国際化，対訳テキストの増加に伴って対訳テキストを利用した自然言語処理技術の発展はますます重要となっており，現在も多くの研究が精力的に行われている．このような研究に興味を持たれた読者はAssociation for Computational Linguistics（ACL）のAnnual MeetingやWorkshop on Very Large Corporaの予稿集などを参照されるとよいであろう．

参考文献
1）春野雅彦: 適応的辞書環境AIDA, 自然言語処理研究会NL-120-18, pp.117-124 (1997).
2）Brown, P. et al.: Aligning Sentences in Parallel Corpora, the 29th Annual Meeting of the Association for Computational Linguistics, pp.169-176 (1991).
3）Gale, W. and Church, K.: A Program for Aligning Sentences in Bilingual Corpora, Computational Kinguistics, Vol.19, No.1, pp.75-102 (1993).
4）Kay, M. and Roscheisen, M.: Text-Translation Alignment, Computational Linguistics, Vol.19, No.1, pp.121-142 (1993).
5）春野雅彦: 辞書と統計を用いた対訳アライメント, 情報処理学会論文誌, Vol.38, No.4, pp.719-726 (Apr. 1997).
6）Dagan, I. and Church, K.: Termight: Identifying and Translating Technical Terminology, Proc. Fourth Conference on

No.	Japanese	English	No.	Japanese	English
1	アデニル酸シクラーゼ	adenylyl cyclase	38	アドレナリンレセプター	adrenoceptors
2	シナプス前促通	presynaptic facilitation	39	オートラジオグラフィー	autoradiography
3	1次視覚野	primary visual cortex	40	カリフォルニア大学ロサンゼルス校	University of California at Los Angeles
4	古典的条件付け	classical conditioning	41	ヘッブ型シナプス	Hebb synapses
5	長期記憶	long-term memory	42	受容野	receptive fields
6	調節ニューロン	modulatory neuron	43	一卵性双生児	identical twins
7	脳梁	corpus callosum	44	異性愛男性	heterosexual men
8	テトロドトキシン	tetrodotoxin	45	機能詞	functors
9	プロテインキナーゼ	protein kinase	46	顔細胞	face cells
10	フリーラジカル	free radicals	47	失語症患者	aphasic patients
11	神経原線維変化	neurotibrillary tangles	48	潜在的学習	implicit learning
12	接続詞	conjunctions	49	潜在的再現像	implicit representations
13	辺縁系	limbic system	50	脊椎動物	vertebrates
14	カテコールアミン	catecholamines	51	遅延反応課題	delayed-response tasks
15	注視する	gaze	52	カリフォルニア大学サンディエゴ校	University of California at San Diego
16	コロンビア大学	Columbia University	53	脳活動	brain activity
17	パーキンソン病	Parikinson's disease	54	連想記憶	associative memory
18	モノアミン	monoamines	55	老人斑	senile plaques
19	外套膜	mantle shelf	56	ポスドク	postdoctoral fellow
20	皮質下	subcortical	57	樹状突起	dendrites
21	アミロイドタンパク質	amyloidprotein	58	側枝	side branches
22	ノルアドレナリン	norepinephrine	59	米国立精神衛生研究所	National Institute of Mental Health
23	シナプス前終末	presynaptic terminals	60	盲点	blind spot
24	眼優位性カラム	ocular dominance columns	61	幻覚	hallucinations
25	解剖学的	anatomic	62	運動ニューロン	motor neuron
26	抗うつ薬	antidepressants	63	NMDA受容体	NMDA receptors
27	神経伝達物質	neurotransmitters	64	媒介系	mediation systems
28	主溝	principal sulcus	65	高次精神機能	higher-mental functions
29	条件付け	conditioning	66	両側	bilateral
30	新皮質	neocortex	67	グルタミン酸受容体	glutamate receptor
31	臨界期	critical period	68	シンガー	Singer
32	染色体	chromosome	69	ジヒドロテストステロン	dihydrotestosterone
33	痴呆	dementia	70	ホルモン環境	hormonal environment
34	短期記憶	short-term memory	71	一夫多妻	polygynous
35	長期増強	long-term potentiation	72	一酸化炭素中毒	carbon monoxide poisoning
36	無条件刺激	unconditioned stimulus	73	音刺激	auditory stimuli
37	クロサピン	clozapine	74	遅発性ジスキネシア	tardive dyskinesia

表-3　日経サイエンスから得られた対訳表現

Applied Natural Language Processing, pp.34-40 (1994).
7) Kupiec, J.: An Algorithm for Finding Noun Phrase Correspondences in Bilingual Corpora, the 31st Annual Meeting of the Association for Computational Linguistics, pp.17-22 (1993).
8) Kumano, A. and Hirakawa, H.: Building an MT Dictionary from Parallel Texts Based on Linguistic and Statistical Information, Proc. 15th International Conference on Computational Linguistics, pp.76-81 (1994).
9) Haruno, M., Ikehara, S. and Yamazaki, T.: Learning Bilingual Collocations by Word-level Sorting, Proc. 16th International Conference on Computational Linguistics, pp.525-530 (1996).
10) Kitamura, M. and Matsumoto, Y.: Automatic Extraction of Word Sequence Correspondences in Parallel Corpora, Proc. Fourth Workshop on Very Large Corpora, pp.79-87 (1996).
11) Matsumoto, Y., Ishimoto, H. and Utsuro, T.: Structural Matching of Parallel Texts, the 31st Annual Meeting of the Association for Computational Linguistics, pp.23-30 (1993).
12) Smadja, F. and McKeown, K.: Translating Collocations for Use in Bilingual Lexicons, ARPA Human Language Technology Workshop 94, pp.152-156 (1994).
13) Haruno, M. and Ikehara, S.: Two-Step Extraction of Bilingual Collocations by Using Word-Level Sorting, IEICE Transaction on Information and Systems, Vol.E81-D, No.10, pp.1103-1110 (1998).
14) Wu, D.: Aligning a Parallel English-Chinese Corpus Statistically with Lexical Criteria, the 32nd Annual Meeting of the Association for Computational Linguistics, pp.80-87 (1994).
15) Isahara, H.: JEIDA's English-Japanese Bilingual Corpus Project, First International Conference on Language Resources and Evaluation, pp.471-474 (1998).
16) Fung, P.: A Pattern Matching Method for Finding Noun and Proper Nouns Translations from Noisy Parallel Corpora, Proc. 33rd Annual Meeting of the Association for Computational Linguistics, pp.236-243 (1995).
17) Melamed, I.: A Portable Algorithm for Mapping Bitext Correspondence, Proc. 35th Annual Meeting of the Association for Computational Linguistics, pp.305-312 (1997).
18) Sato, S. and Nagao, M: Toward Memory-Based Translation, Proc. 13th International Conference on Computational Linguistics, pp.247-252 (1990).
19) Brown, P., Pietra, S., Pietra, V. and Mercer, R.: The Mathematics of Statistical Machine Translation: Parameter Estimation, Computational Linguistics, Vol.19, pp.263-312 (1993).
20) Alshawi, H., Buchsbaum, A. and Xia, F.: A Comparison of Head Transducers and Transfer for a Limited Domain Translation Application, Proc. 35th Annual Meeting of the Association for Computational Linguistics, pp.360-365 (1997).
21) Collins, M.: A New Statistical Parse Based on Bigram Lexical Dependencies, Proc. 34th Annual Meeting of Association for Computational Linguistics, pp.184-191 (1996).
22) Haruno, M., Shirai, S. and Ooyama, Y.: Using Decision Trees to Construct a Practical Parser, Machine Learning, Vol.34, No.1, pp.131-150 (1999).

(平成11年3月11日受付)