

テキストマイニング分析による日本人英語学習者の

第二言語習得への学習動機評価モデルの検討

照井 雅子 (大阪大学大学院言語文化研究科)

宮永 正治 (フリーランス翻訳者)

三崎 敦子 (サイマル・アカデミー)

野口 ジュディー (武庫川女子大学薬学部)

1. はじめに

日英翻訳は複雑なプロセスだが、応用言語学の ESP (English for Specific Purposes) アプローチを取り入れ、ジャンル概念とコーパスツールを利用した翻訳者ワークショップでは、翻訳プロセスの効率が上がった (宮永ら 2009)。本ワークショップの 11 名の参加者間で交換された 445 通のメールを、形態素解析プログラム「茶筌」「JUMAN」を利用したテキストマイニングのためのフリーソフトウェア KHCorder (樋口 2005) を用いて計量的に分析することを試みたので報告する。

翻訳者ワークショップの参加者たちは第二言語である英語の学習者でもあるが、高度な英語運用能力を有しながらなお自立的に英語学習に取り組んでいる。こうした日本人英語学習者の第二言語習得への学習動機を、テキストマイニングによって評価するモデルを検討した。

2. 日英翻訳ワークショップ

2.1 概要

本ワークショップは、民間の大手英語学校日英翻訳クラスの受講生が、既習クラスの枠

を超え、自分たちの学習意欲をさらに満たす状況で学びたいとの思いから、英語学校の事務局、英語母語話者講師、日本人講師、受講希望者間で話し合いを重ね、実現に至った。

2006 年 7-8 月、2007 年 1-2 月及び 3-4 月の各月 1 回 120 分、夏・中間・春の計 3 回のワークショップが実施された。参加者は 10 名で、翻訳者、通訳者、人材派遣会社勤務、特許事務所勤務、銀行員、証券トレーダー、公務員、高校教員等、職業は様々である。

英語母語話者講師から開催の条件として示された「全員が一人一人のできる部分で自発的に貢献する、参加型のワークショップにする」という提案を受け、参加者が自分たちで発案、企画、運営した点が本ワークショップの大きな特徴である。

2.2 内容

第 1 回のワークショップでは、参加者 10 名の話し合いで、翻訳需要の高い「ビジネス」分野で、製造業企業と非製造業企業 (商社) から各 1 社を選び、翻訳に取り組む文書として、「社長挨拶」と「環境方針」というジャンルを選んだ。各ジャンルにおいて、英語母

語話者が書いたと思われる文書データを各自で集めて一つにまとめることで、全員が利用できるコーパスを作成した。参加者は計5組のペアとなり、各ペアでお互いの翻訳をチェックし、翻訳の苦労や工夫と、パートナーから得た助言をどう自分の翻訳に活かしたかという経緯をメーリングリストに投稿し合った。ワークショップでは、メーリングリストを通じて既に共有している情報に基づいて活発な議論がなされた。ワークショップ後に、各自で自分の翻訳を修正・改善し、講師に課題を再提出し、校閲を受けた。

第2回として実施された中間ワークショップは、第1回から次回開催まで半年以上あり、また、課題の量が多かった第1回は時間の制約で十分議論が尽くせなかった面もあったので、全員で議論をさらに続け、ある程度の意見の一致をみた試訳を「完成させたい」との声が受講生から強く上がり、日本人講師がまとめ役となって実施された。第1回の課題の1つである「環境方針」をより深く検討する目的で行われた。全員で議論を深めて試訳を完成する中で、各自の問題点を一つの表にまとめることもできた。

第3回は、全員で一つの課題を丁寧に最後まで完成させることを目的として、課題を製造業企業1社の「プレスリリース」というジャンルに絞り込んだ。全員でコーパスを作成し、英語母語話者によって書かれたプレスリリースのレトリック構造を明らかにする Move (Swales 1990) 分析も行った。ワークショップ前の翻訳第一稿を、ワークショップ初日の議論に基づき翻訳第二稿として仕上げ、講師の校閲を受けて、ワークショップ2日目に再度議論した。

3. メーリングリスト

3.1 概要

ワークショップ参加者は英語運用能力が高く、翻訳作業には慣れているが、上述のように、本ワークショップでは ESP アプローチやジャンル概念の導入、コーパスの作成や利用といった新しい翻訳手法を取り入れたため (野口 2008)、様々な情報の共有と意見交換の必要があった。対面での連絡が困難だったため、Yahoo のグループ機能のメーリングリストを活用した。ワークショップ当日だけでなく、参加者それぞれが準備やまとめの作業に多くの時間を割き、連絡を取り合った結果、2006年6月より2007年9月までに計445通のメールがやりとりされた。

3.2 KHCorder

上述のメーリングリストのテキストマイニングを行うために KH Coder を利用した。KH Coder は、様々な日本語のテキスト型データの内容を計量的に分析するためのフリーソフトウェアで、自然言語処理に広く用いられている形態素解析プログラム「茶釜」 「JUMAN」 を利用している。

3.3 テキストマイニング用データの作成

Yahoo メーリングリストにある計445通のメールを1通ずつ開封し、コピーし、MS Word ファイルに貼り付けた。その際、以下のようなアバターや広告は分析の妨げとなるため削除した。

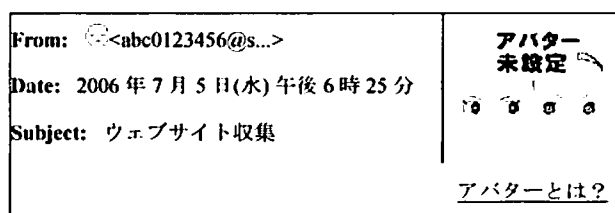


図1 メーリングリストにより配信されるメール中の広告

Yahoo! Groups Sponsor

■□ 自動車保険は安いほうがいい! □■

複数の保険会社から、『無料』で一気に見積もりが可能
 自動車保険見積もりで、自分にぴったりの保険をチェック
 今ならもれなくYahoo!ポイントプレゼント
 ↓詳しくはこちら↓
<http://.....>

図2 メールングリストにより配信されるメール中の広告の例

ヘルプページ: <http://help.yahoo.co.jp/help/jp/groups/>
 グループページ: <http://groups.yahoo.co.jp/group/>
 グループ管理者: <mailto:X-owner@yahoo.co.jp>

・モバイル: <http://rd.yahoo.co.jp/egroups/050616info/1.html>
 ・移行手段: <http://rd.yahoo.co.jp/egroups/050616info/2.html>
 ・利用規約: <http://rd.yahoo.co.jp/egroups/050616info/3.html>

図3 メールングリストにより配信される連絡事項

他に、メール内のリンク内容がテキストデータとして自動的にメールに貼り付けられていた場合は削除した。また、投稿者が単純に前のメールをそのまま含めて返信している場合は、前のメールを削除した。誰かのメールを引用した上で投稿者が新たなコメントを加えている場合は削除していない。添付忘れ等で同じ内容のメールが2通投稿された場合は、同じ内容の部分は削除した。添付ファイル情報も削除した。

すべてのメールを含んだファイルと、投稿の時期と内容から「ワークショップ夏」「中間ワークショップ」「ワークショップ春」「ワークショップ以降」の4つのファイルを作成し、それぞれをテキストデータに変換した。

一方で、各メールの内容を改行毎にエクセルファイルに記録した。項目は、「メール番号(通し)」「差出人」「宛先」「(投稿)日付」「時間」「件名」「段落番号」「文章番号」「メッセージ」とした。このファイルは全体の流れを把握するため利用した。

表1 メール例

No.	差出人	宛先	日付	時間	件名	段落No.	文章No.	メッセージ
1	D	Nさん Mさん	60614	9:36	NさんML作成 ありがとうございました。	1	1	おはようございます。
1	D	Nさん Mさん	60614	9:36	NさんML作成 ありがとうございました。	1	2	やはり連絡遅れの心配もなく 便利ですわ。
1	D	Nさん Mさん	60614	9:36	NさんML作成 ありがとうございました。	1	3	早速ありがとうございます。
1	D	Mさん	60614	9:36	NさんML作成 ありがとうございました。	2	1	メールアドレスについて、情報 ありがとうございました。
1	D	Mさん	60614	9:36	NさんML作成 ありがとうございました。	2	2	翌に、コーパスの作成という 大変な作業・・・
1	D	Nさん Mさん Yさん Tさん	60614	9:36	NさんML作成 ありがとうございました。	3	1	次回のTIMEを聴く会は必ず 出席させていただきたく ので、17日の会社の選定 などは、宜しく願います。
1	D	引用	60614	9:36	NさんML作成 ありがとうございました。	4	1	> Maybe we can start a "Translators' Workshop" on our own and run mini- seminars! Judy Noguchi
1	D	講師	60614	9:36	NさんML作成 ありがとうございました。	5	1	J-sensei, it was really encouraging to read !!
1	D	講師	60614	9:36	NさんML作成 ありがとうございました。	5	2	Whether we can do it at all depends on our efforts and commitments.
1	D	講師	60614	9:36	NさんML作成 ありがとうございました。	5	3	I'll make effort to keep up with other members.

4. データ分析

表2 投稿回数と投稿文字数

投稿者	投稿回数	投稿文字数	備考
MS	117	53,993	企業選定・企業情報の提供 コーパスやデータの管理 会場手配・予約
TY	52	18,410	企業選定・企業情報の提供
TM	45	24,110	運営事務
MA	36	23,982	日本人講師
NJ	34	17,278	WS発案者
AY	31	2,840	
DY	28	8,314	
JN	26	2,900 words 827文字(引用)	ネイティブ講師 主に英語で投稿
IT	21	5,229	
YT	19	9,176	企業選定・企業情報の提供
SF	18	3,855	中間WSより参加 1/11初投稿
Auto	10		ブリーフケースにファイル がアップロードされた際の Yahooからの自動応答
Office	6	1,005	英語学校からの事務連絡
NM	4	1,405	New comer 7/16初投稿

表3 投稿メール数と投稿文字数

	夏	中間	春	WS以降
投稿メール数	168	142	125	10
投稿文字数	93,026	91,160	52,939	5,652

表4 総抽出後数と異なり語数

	夏	中間	春	WS以降
総抽出語数	43,507	49,010	26,439	2,789
異なり語数 (使用)	4,279 (3,795)	3,573 (3,019)	2,837 (2,408)	733 (592)
文	3,397	3,558	2,317	213

表5 頻出語

夏		中間		春		WS以降	
名前の呼びかけ	436	名前の呼びかけ	540	名前の呼びかけ	350	名前の呼びかけ	44
ファイル	176	課題	278	先生	127	発表	24
挨拶の言葉	156	ワークショップ	186	ファイル	104	皆さん	17
皆さん	93	皆さん	158	課題	103	先生	15
翻訳	93	先生	152	アップ/ダウンロード	77	ありがとう	13
環境	92	速刷	107	挨拶の言葉	74	翻訳	12
お願い	83	分析	107	ワークショップ	73	お願い	11
コーパス	78	意見	104	グループ	66	関係	11
メーカー	73	お願い	101	ありがとう	62	タイトル	9
会社	73	お加	99	お加	62	関係	10
先生	69	ありがとう	98	日程	62	使用	9
ありがとう	67	コーパス	96	皆さん	52	言語	8
グループ	63	候補	75	変更	51	ワークショップ	7
提出	61	内容	73	お願い	49	シンポジウム	6
メッセージ	62	挨拶の言葉	69	提出	48	回答	6
ワークショップ	53	ジャンル	69	プレスリリース	37	協力	5
お加	33	ファイル	65	連絡	37	挨拶の言葉	5
送付	34	企業	61	メール	36	字会	5
コメント	28	会社	58	提案	36	許可	4
添付	27	メッセージ	53	投票	31	アンケート	3

5. 考察とまとめ

全体を通して、メーリングリストは主にワークショップの準備により多く利用された。これは、第1回のワークショップの課題の再提出を行った後、第2回のワークショップの準備開始までに約3ヶ月のブランクがあったこと、第3回のワークショップ終了後に交わされたメールがわずか10通であったことから明らかで、参加者がワークショップ当日の議論をより深く活発なものにし、より良い翻訳文を完成させたいために、大きな努力を払ったことがわかる。

参加者は「全員が一人一人のできる部分で自発的に貢献する、参加型のワークショップにする」というコンセプトを十分に理解し、実現するための努力を惜しまなかった。その努力の原動力を探るためにテキストマイニングを試みたわけだが、計445通のメールに含まれる272,777文字をKHCorderを用いて分析するためには、事前のデータ作成に多くの手間と時間を要した。しかし、KHCorder

上で「こんにちは」と「こんばんは」を同じ「挨拶の言葉」と見なすコーディングを行った上で抽出した頻出語を見ると、データを単に時系列で観察していた際に得ていた直感が裏付けられた。

つまり、参加者はメール投稿時に挨拶を欠かさず、頻繁に互いの名前を呼び合い、「ありがとう」「お願い」と感謝とねぎらいの言葉を掛け合っていた。この背景には、課題の選定やコーパス作成に、参加者の一部が、細やかで的確な指示を何度も出し、最終合意に至らせた多大な貢献があった。

結局のところ、本ワークショップの参加者は、第二言語として英語を単に学習するだけでなく、さらに高度で複雑な作業である日英翻訳の質を高めたいという強い意欲を持ち続けていることが、学習動機となっていることが読み取れた。翻訳にはBest answerがなく、Better answerが複数存在するため、質を高めるといふハードルは常に高く設定することができ、学習動機を維持し続ける一つの要因となっていると言える。

References

樋口耕一 (2005). 計量テキスト分析の方法と実践. 大阪大学大学院人間科学研究科 平成16年度 博士論文

Noguchi, J. (2008). ESP and its Role in Education in Japan Today. In Fukui, K. (Ed.), ESP e-Learning for Global Competency (pp. 3-15). Tokyo: University Academy Press, Inc.

宮永正治・照井雅子・三崎敦子・野口ジュディー (2009). 翻訳プロセスへのコーパス分析とジャンル概念導入の有効性. 言語処理学会第15回年次大会発表論文集 pp. 422-425.

Swales, J. M. (1990). Genre Analysis: English in academic and research settings. Cambridge, UK: Cambridge University Press.