

「シンポジウム」

話し言葉コーパスの構築と利用 JECPRESE: JSL と EFL ユーザーのために

野口ジュディー

1. はじめに

JECPRESE, the Japanese-English Corpus of Presentations in Science and Engineering (<http://www.jecprese.sci.waseda.ac.jp/>) は留学生のための専門日本語教育 (JSL, Japanese as a second language) を支援する研究発表コーパスとしてスタートした。日本の大学院生の日本語プレゼンテーションに加えて、米国の大学生や国際学会の英語プレゼンテーションも収められており、EFL (English as a Foreign Language) の学生にも利用可能なコーパスとなった。理工系のプレゼンテーション (発表) の特徴をわかりやすくするために、各発表について、ESP (English for Specific Purposes) の手法であるジャンル分析に依拠したセクションやステップで検索することができるように設計してある。単語や表現の検索もできるようにになっている。

2. プレゼンテーションコーパスの必要性

科学技術分野では、研究の成果を専門分野の研究者と共有することが必須である (Gross 1990, Myers 1990, Canagarajah 2002)。研究成果を共有する方法としては、専門誌への論文投稿や学会での口頭発表、ポスター発表などがある。論文の書き方についての教科書は数多く出版されており、また、論文テキストによるコーパスを構築すると、執筆時に参考にすることができる。しかし、口頭発表やポスター発表の資料は入手するのが難しい上に、コーパス化するためには音声を書き起こすことが必要になる。この問題を解決するために、研究発表の日本語と英語を集めたコーパスを構築した。

最初に、日本語による研究発表をコーパス化することから着手した。日本の大学には、科学技術を学ぶために海外から訪れる留学生が多いため、日本語での研究発表コーパスがあれば、各自の研究発表のみならず、他の学生による日

本語での発表を聞く際の助けにもなるであろうと考えられた。これは、後述する ESP (English for specific purposes) のジャンルの概念に基づくものである。

3. ESP (English for specific purposes) のジャンル分析

グローバル規模で科学技術研究者が高度なレベルでの情報交換を行うには、科学の内容を正確に伝えるための英語力が必要となる。しかし、科学技術を専攻する学生や研究者の全員が高い言語運用能力を有するとは限らない。良い研究を行っても、研究内容を上手に発表することができないと研究に対する理解を得ることは難しいであろう。そのような状況で、学生や研究者の助けになるアプローチが専門英語研究 (ESP) から生まれたジャンル分析の方法である (Swales 1990)。

ESP では、英語運用能力全般を引き上げることを目標としない。そのことより、特定の文書や話し言葉の特徴を明らかにして、その使い方の向上を目指す。例えば、パーティーで流暢に会話を行うことより、良い研究論文を書くことを優先する。それは、文書の専門性が高ければ高いほど、より定型的なパターンが認められ、学びやすいためである。

このパターン化を簡単な例で検証してみよう。学会での口頭発表を行いたい場合、まず、発表内容を把握できるような要旨を学会に提出する。この内容が学会の趣旨と合致し、良い研究であると判断されると、口頭発表を行うことが許可される。最近、ハワイで開催された国際学会での口頭発表の要旨について、ジャンル分析 (パターン分析) をしてみよう。

冒頭の括弧内には、それぞれのセクションの目的が書いてある。文中に下線を施した部分は、読み手にヒントを与える表現である。これらの表現を手がかりにすると、書き手がそのセクションで何を伝えようとしているのかが理解しやすくなる。このように、文章の目的をわかりやすくすることで文書が読みやすくなる (Tojo et al. 2014)。

【Purpose of presentation】 This paper presents corpus linguistic findings that show how differences between Japanese and English influence the way in which science concepts are transmitted in science education today. **【Background】** We first review how Western science concepts entered Japan via translation from Dutch during the Edo Period (1603 to 1868) and then the impact on science education after World War II of American science textbooks being translated for

classroom use. **Today**, with increasing globalization, the Japanese Ministry of Education, Culture, Sports, Science and Technology has designated Super Science High Schools and encouraged the delivery of lectures in English at universities. **[Research aim & methods]** **To find ways to help** instructors prepare their lectures in English and to support students who need to listen to them, since 2010, **we have been working with** corpora of university engineering lectures presented in English. **However**, due to the language distance between English and Japanese, communication differences between Western and Japanese cultures, and prevalent teaching traditions in Japan, **we became aware of the need to** compare the English lectures with those delivered in Japanese. From 2016, being interested in how scientific knowledge was being imparted to the students, we began adding lectures delivered in Japanese to our corpus. **[Results]** **Comparison of** these Japanese lectures with the English ones **has revealed differences** in the lecturers' perceptions of how science should be taught. **[Discussion]** **We discuss how** language and cultural differences can influence the way concepts are transmitted in science education. (Noguchi, J., Kunioishi, N., & Tojo, K. Language and cultural influences on “doing science in Japanese”, The IAFOR International Conference on Education – Hawaii, January 4 to 6, 2018, Honolulu)

このように、明確な目的がある文書に用いられる表現には、繰り返し使用されるうちにより効率よく内容が伝わるようにパターン化されたものが多い (Dressen-Hammouda 2008)。書いたり話したりして発信するときも、読んだり聞いたりして受信するときも、専門分野の書き言葉や話し言葉が有するパターンと高頻度で使用されるヒント表現を知っていると、受発信することが容易になる。パターンを知るための方法のひとつに、コーパスを利用して、頻回に使用される表現を検索することがある。ESP の概念に基づく文書の捉え方を理解することができる、必要な専門文書をより平易に学ぶことができるようになる。

4. アカデミックプレゼンテーションコーパスの構築

4.1 JECPRESE の日本語プレゼンテーションコーパス

日本の工学系大学院で研究を行う留学生を支援する目的で、日本人学生による日本語の口頭発表を 2006 から 2008 年にかけて録音し、書き起こしを行った。

データ収集は、各大学院の各専攻に許可を得た上で、発表者個人からも許可書を得て行った。収録を許可されたコーパスにおけるデータの内訳は、物質科学が34、分子化学が31、知能機能創成工学が34、電気工学が13、機械工学が61であり、日本語の話し言葉（理系プレゼンテーション）としては約120万語であった (Kunioshi et al. 2012)。

4.2 JECPRESE の英語プレゼンテーションコーパス

英語を第二言語として使用する学生のためには、英語でのプレゼンテーションのコーパスが役立つであろうと考えられたため、2種類（2つのジャンル）の英語の口頭発表をコーパスに加えた。1つ目は、化学分野の国際学会における、さまざまな国のベテランの研究者による口頭発表であり、データ数は45、英語の話し言葉（化学系プレゼンテーション）としては約29万語であった。2つ目は、米国の大学の、工学部の学生によるグループ発表であり、データ収集に際しては、約1年かけて大学の倫理委員会に収録の許可を得た上で、学生個人の承諾書も取得した。収録の対象となった学生数は、グループごとに3～4名、合計54名であり、データの内訳は、電気工学が2、土木・環境工学が5、機械工学が6、コンピュータ科学が2であり、英語の話し言葉（理系プレゼンテーション）としては、約44万語であった。

5. JECPRESE の検索システム（インタフェース）

収集した話し言葉日英（理系プレゼンテーション）コーパスは、利便性を考慮して Fig. 1 のような検索システム（インタフェース）からアクセスできるように整備した。検索窓に言葉を入力すると、コンコーダンスラインを得ることができる。しかし、英語や日本語の発表に不慣れな学生にとっては、入力する言葉を思い付くこと自体が難しいと考えられたため、セクション (Sections) とステップ (Steps) という選択肢を設けた。セクションとステップの選択肢についての詳細は、Table 1 を参照されたい。例えば、セクションの *S* (Start) のチェックボックスにチェックを入れると、プレゼンテーションの開始時に発話された言葉が表示される。日本語での口頭発表の開始時には、Table 2 に示したような典型的な表現が繰り返し使用されていることが明らかになった。また、プレゼンテーションの終了時には、「以上です。」という表現が18件、「発表を終わらせていただきます。」という表現が14件認められた。

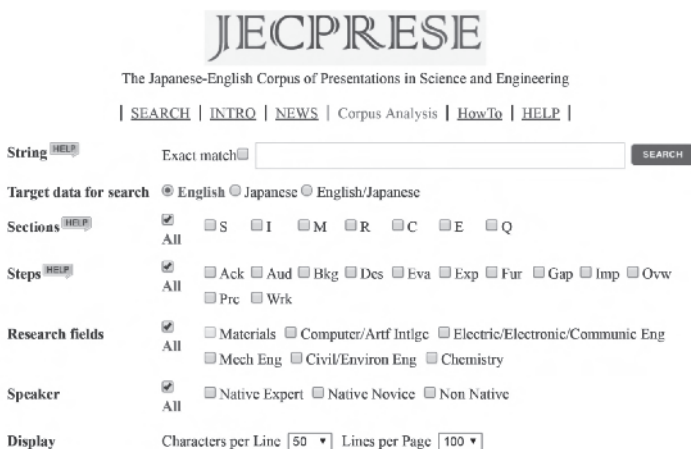


Fig. 1 JECPRESE の検索ページ

Table 1 List moves found in English and Japanese presentations

Tag	Section	Tag	Description
S	Start	Ack	Acknowledgments
I	Introduction	Aud	Audience orientation
M	Materials and methods	Bkg	Background
R	Results and discussion	Des	Description
C	Conclusion	Eva	Evaluation
E	Ending	Exp	Explanation
Q	Question and answer	Fur	Further research
		Gap	Gap
		Imp	Implication
		Ovw	Overview
		Prc	Procedures
		Wrk	Present work

Table 2 日本語での口頭発表の開始時における表現

高頻度使用表現	回数
TITLE と題しまして SURNAME 研究室の SURNAME が発表させていただきます.	16
それでは、発表させていただきます.	14
SURNAME 研の SURNAME です.	5
SURNAME 研究室の SURNAME です.	4

JECPRESE の検索システムでは、口頭発表の各セクション (Sections) におけるステップ (Steps) での詳しい説明の表現を検索することもできる。例えば、英語での口頭発表の導入部で先行研究の隙間 (Gap) をどのように説明しているかを知りたい場合、検索対象 (Target data for search) の項目で英語のプレゼンテーション (English) をクリックし、セクション (Sections) の *I* (Introduction) とステップ (Steps) の *Gap* (研究の意義を説明するステップ) にチェックを入れると、以下のような例が表示される。学生は、表示された表現を参考に各自のプレゼンテーションを準備することができる。

And **we were also asked to design** a sustainable alternative for the reuse of their grey water.

For our project we were to design an onsite waste water facility to accommodate a new residential

So for our problem statement, we are to construct a machine that can replace the current system for

So they are looking for more durable system to make it easier for the river guides.

The problem with this is that you have two materials that are difficult to separate, as you know,

There are two types of challenges.

What we need, however, is a deeper understanding of complex structures.

なお、日本語の口頭発表では、Gap を述べる際に「しかし、」と「そのため、」が頻用されることが判明した。

6. 姉妹サイト OnCAL の紹介

JECPRESE の姉妹サイトとして、JECPRESE と類似した検索システム (インタフェース) を有する OnCAL (The Online Corpus of Academic Lectures) (<http://www.oncal.sci.waseda.ac.jp/>) がある。OnCAL コーパスのデータ数は、英語での工学系講義 [MIT OCW (Massachusetts Institute of Technology, Opencourseware) と SEE (Stanford Engineering Everywhere)] が 430 で、英語の話し言葉 (工学系講義) としては約 350 万語 (395 講義時間) である (Kunioshi et al., 2015)。収録

データは、OnCAL の検索システムを用いて検索することができる。Fig. 2 に OnCAL 検索システムのホームページのイメージを示した。この検索システムでは、JECPRESE の検索システムとは異なるアプローチで検索を行うことができる。

Fig. 2 OnCAL の検索ページ

OnCAL は、英語での工学系講義の流れと表現を明らかにして、日本の大学で英語による理工系の講義を支援することを目的として構築した。OnCAL の検索システムで検索できる項目の中に、Pedagogical Function（教育的機能）がある。例えば、Pedagogical Function の選択肢から、*Thought Experiment* を選択して検索すると、*let's suppose that* や *let's imagine that* などが頻出することがわかる。また、同じく選択肢から、*Cause / effect* を選択して検索すると、*as a consequence of that* や *as a result of this, can lead to, causes it to* などが提示される。このような検索を行うと、英語で講義を行うのに必要な表現を知ることができるだけでなく、英語での講義を聞くのにも役立つであろうと考えられる。

7. 終わりに

話し言葉コーパスとして、日英（理系プレゼンテーション）コーパス JECPRESE とその検索システムを紹介し、姉妹サイトとして英語（工学系講義）

コーパス OnCAL についても簡単に紹介した。是非 JECPRESE と OnCAL の検索サイトにアクセスして、試していただきたい。

謝 辞

JECPRESE は、科学研究費助成事業（科学研究費補助金）基盤研究（C）21520601「理工系口頭発表コーパスに基づいた専門日本語・英語の教育法の開発」の支援を受けた。OnCAL は、科学研究費助成事業（科学研究費補助金）基盤研究（B）24300273「英語を介した理工系高等教育の向上を支援するシステムの開発」の支援を受けた。構築に携わった研究員は、国吉ニルソン、東條加寿子、林洋子、野口ジュディーであった。

参考文献

- Canagarajah, S. (2002) Multilingual writers and the academic community: towards a critical relationship. *Journal of English for Academic Purposes*, 1 (2002) 29–44.
- Dressen-Hammouda, D. (2008) From novice to disciplinary expert: Disciplinary identity and genre mastery. *English for Specific Purposes*, 27, 233–252.
- Gross, A. (1990) *The Rhetoric of Science*. Cambridge, MA: Harvard University Press.
- Kunioshi, N., Noguchi, J., Hayashi, H. & Tojo, K. (2012) An online support site for preparation of oral presentations in science and engineering, *European Journal of Engineering Education*, 2012, 1–9.
- Kunioshi, N., Noguchi, J., Tojo, K. & Hayashi, H. (2015) Supporting English-medium pedagogy through an online corpus of science and engineering lectures, *European Journal of Engineering Education*. <http://dx.doi.org/10.1080/03043797.2015.1056104>
- Myers, G. (1990) *Writing biology: Texts in the social construction of scientific knowledge*. Madison, Wisconsin: The University of Wisconsin Press.
- Swales, J. M. (1990) *Genre analysis: English in academic and research settings*. Cambridge, UK: Cambridge University Press.
- Tojo, K., Hayashi, H., & Noguchi, J. (2014) Linguistic dimensions of hint expressions in science and engineering research presentations. *JACET Selected Papers* Vol. 1(2014), 131–163.

(神戸学院大学名誉教授 Email: jnoguchi@gc.kobegakuin.ac.jp)