# 話し言葉コーパスを用いた理工学系留学生のための日本語学習支援システム
## ―『理工学系語彙・用例学習支援システム レインボー』の開発―

伊藤夏実・遠藤直子・菅谷有子・成永淑・古市由美子・森幸穂

【キーワード】 理工学系話し言葉コーパス、語彙・用例検索ツール、
オンライン学習支援システム、専門分野の複合名詞、
共起表現

## １．はじめに

『理工学系語彙・用例学習支援システム　レインボー[1]』（以下レインボー）は、東京大学大学院工学系研究科コーパス研究チームが、理工学系留学生[2]（以下留学生）のために開発したオンライン日本語学習支援システムである。本研究チームは2007年より工学系研究室のゼミ内の日本語による発表、質疑応答を含む自然発話を音声データとして収録し、文字化した理工学系話し言葉コーパス（以下SESJコーパス【仮称】）[3]を構築している。レインボーはこのSESJコーパスを資源とした語彙・用例を検索することができるオンライン日本語学習支援システムである。本稿では、まずSESJコーパスの構築からレインボー開発までの経緯について述べ、次にレインボーの具体的な作成過程を紹介する。

## ２．レインボー開発の背景

本節では、レインボー開発の背景にあるSESJコーパスの構築およびSESJコーパスを基にした教材の有用性について述べる。

### ２－１ SESJコーパス構築

東京大学大学院工学系研究科では、基本的には英語で研究活動を行う体制が整えられているが[4]、研究の状況によっては、日本語でのコミュニケー

ションが必要とされる場合がある。このような状況の中、日本語の必要性を感じながらも研究中心の生活を送る留学生には日本語学習に割く時間が十分に取れない、といったジレンマが生じている。日本語教育の現場の視点からすれば、彼らにとって必要なのは各分野の研究現場のニーズに合った日本語を効果的に学習できるような日本語教材である。しかし、様々な状況下にある留学生を包括的に支援するアカデミックな日本語教材、特に理工学系日本語教材の整備は十分になされているとは言えない。本研究チームは、留学生たちの研究生活および研究現場の実情やニーズにあった日本語教育および教材を提供するために、2005年からCDS(Can-do Statement)の作成（古市他2008）を行い、次いでプログラム評価（菅谷他2008）を実施した。その結果、研究場面で使用される日本語の自然発話の実態解明の必要性が浮かび上がってきた。そこで2007年より工学系研究室のゼミ内の日本語による発表、質疑応答を含む自然発話を音声データとして収録し、SESJコーパスとして構築している。表１に示したように、これまで、工学系の４分野（電気系工学、都市環境工学、都市計画、建築学）において約80時間収録し、形態素解析にはKH Coder[5]を使用し、延べ語数1,177,834、異なり語数51,277が得られている。現在も新たに3分野（社会基盤学、化学システム工学、電子情報学）でのコーパスを構築中である。

**表1　分野別音声データ収録時間と期間**

| 専攻分野 | 収録期間 | 収録時間 |
|---|---|---|
| 電気系工学 | 2007年11月～2008年2月 | 約20時間 |
| 都市環境工学 | 2008年12月～2009年2月 | 約23時間 |
| 都市計画 | 2008年12月～2009年2月 | 約12時間 |
| 建築学 | 2009年6月～2009年12月 | 約26時間 |

### 2－2 SESJコーパスを基にした教材の必要性

コーパスの分析に基づいた教材作成は有効な手段であり（大曾2006）、理工学系学習者を対象とした専門日本語教育、とりわけ語彙教育においては、市販の日本語教材の語彙だけでは十分ではなく、SESJコーパスを活用した日本語教材の開発が望まれる（菅谷他2009、単他2011）。

書き言葉コーパスを基にした教材に続き、話し言葉コーパスを基にした教材も徐々に整備されつつあるが、専門用語など独特な日本語表現を取り上げている理工学系話し言葉コーパス自体がまだ少なく、それを基にした日本語教材は管見の限り殆どない。名古屋大学が国際化拠点整備事業の一環として作成した教材『留学生のための専門講義の日本語』（2010）は、話し言葉コーパスを基に作成されているが、これは教員や大学院生による学部留学生向けの講義形式の模擬授業を題材にした教材である。また、林他（2012）が開発したオンライン日英語口頭発表表現検索サイト『JECPRESE（The Japanese-English Corpus of Presentations in Science and Engineering）』の日本語データは、理工系の修士論文口頭発表を収集したコーパスを資源としており、本研究チームのゼミ内の複数話者によるやりとりを収集したSESJコーパスとはやや質が異なる。語彙教育のみならず留学生の日本語によるコミュニケーション能力を培うためにも、本コーパスを資源とした日本語教材の開発は重要と言えるだろう。

## 3．レインボーの基本的枠組み

SESJコーパスをいかに日本語教材として活用するかを課題に、本研究チームが本コーパスの研究と分析を重ねた結果、生まれたのがレインボーである。レインボーでは出現頻度が高い名詞を見出し語とし、その名詞が含まれる用例を提示することとした。以下、それぞれの理由について述べる。

### 3－1 見出し語

まず、KH CoderでSESJコーパスデータを解析し、出現頻度20以上の語彙全てに対訳[6]を付与した。次に、レインボーの検索に必要な見出し語として、これらの語彙のうち、出現頻度20以上の名詞を選び出した。名詞を見出し

語として取り上げたのは、SESJコーパス調査の結果から、名詞に各専門分野の特徴が反映されていることがわかったためである。専門分野で使われている名詞についての詳細は3－1－1で詳しく述べる。

### 3－1－1 専門分野の名詞

鎌田他（2004）、林（2004）、小宮（2005）、三國・小森（2008）などは、専門分野の語彙の特徴として、述語部分に頻出する動詞や形容詞は旧日本語能力検定試験の3、4級レベルのような基本的なものが多く、それらと統語的に共起しうる名詞は専門分野の特徴が関与している可能性を指摘している。また、本研究チームがSESJコーパスの出現頻度が高い語彙の調査をしたところ、それらには各分野の専門性が反映されていることがわかった（猪狩他2009、菅谷他2009、単他2009、山口他2010）。各専門分野にどのような名詞が出現しているかについては、菅谷他（2009）の調査結果から表2、表3、表4、表5に、それぞれ、電気系工学データ、都市環境工学データ、都市計画データ、建築学データの出現頻度20以上の名詞を示す。表2を見ると「制御」「出力」「発電」「系統」「周波数」「蓄電池」など、電気系分野の専門性の特徴が表れている語彙がみられ、表3、表4、表5の語彙についても同様のことが言える。

表2 電気系工学データの名詞の出現頻度

| 出現頻度 | 名詞 | 旧JLPT | 出現頻度 | 名詞 | 旧JLPT |
|---|---|---|---|---|---|
| 1位 | 制御 | 級外 | 11位 | エリア | 級外 |
| 2位 | 計算 | 2級 | 12位 | 容量 | 級外 |
| 3位 | 図 | 2級 | 13位 | 事故 | 3級 |
| 4位 | 出力 | 級外 | 14位 | 蓄電池 | 級外 |
| 5位 | 発電 | 2級 | 15位 | ピッチ | 級外 |
| 6位 | 風 | 4級 | 16位 | 利益 | 2級 |
| 7位 | 変動 | 1級 | 17位 | モデル | 2級 |
| 8位 | 系統 | 2級 | 18位 | 風力 | 級外 |
| 9位 | 周波数 | 級外 | 19位 | 電力 | 2級 |
| 10位 | 値 | 2級 | 20位 | 電圧 | 級外 |

表3　都市環境工学データの名詞の出現頻度

| 出現頻度 | 名詞 | 旧JLPT | 出現頻度 | 名詞 | 旧JLPT |
|---|---|---|---|---|---|
| 1位 | 水 | 4級 | 11位 | 実験 | 2級 |
| 2位 | 分解 | 2級 | 12位 | 添加 | 級外 |
| 3位 | 塩素 | 級外 | 13位 | ろ過 | 級外 |
| 4位 | 濃度 | 2級 | 14位 | データ | 1級 |
| 5位 | 細菌 | 1級 | 15位 | 培養 | 級外 |
| 6位 | ベンゼン | 級外 | 16位 | 感じ | 2級 |
| 7位 | 浄水 | 級外 | 17位 | 形 | 3級 |
| 8位 | 菌 | 1級 | 18位 | 生成 | 級外 |
| 9位 | 水素 | 2級 | 19位 | 微生物 | 級外 |
| 10位 | メタン | 級外 | 20位 | 硫酸 | 級外 |

表4　都市計画データの名詞の出現頻度

| 出現頻度 | 名詞 | 旧JLPT | 出現頻度 | 名詞 | 旧JLPT |
|---|---|---|---|---|---|
| 1位 | 人 | 4級 | 11位 | 研究 | 3級 |
| 2位 | 地域 | 2級 | 12位 | 社会 | 3級 |
| 3位 | 都市 | 2級 | 13位 | 関係 | 3級 |
| 4位 | コミュニティー | 級外 | 14位 | あと | 4級 |
| 5位 | 計画 | 3級 | 15位 | 意味 | 4級 |
| 6位 | スラム | 級外 | 16位 | カフェ | 級外 |
| 7位 | 一つ | 4級 | 17位 | 住宅 | 2級 |
| 8位 | 話 | 4級 | 18位 | 組織 | 2級 |
| 9位 | 自分 | 4級 | 19位 | 利用 | 3級 |
| 10位 | 年 | 4級 | 20位 | 公園 | 4級 |

表5 建築学データの名詞の出現頻度

| 出現頻度 | 名詞 | 旧JLPT | 出現頻度 | 名詞 | 旧JLPT |
|---|---|---|---|---|---|
| 1位 | F1 | 級外 | 11位 | 都市 | 2級 |
| 2位 | 建築 | 2級 | 12位 | 水 | 4級 |
| 3位 | 解体 | 級外 | 13位 | 技術 | 3級 |
| 4位 | 住宅 | 2級 | 14位 | 先生 | 4級 |
| 5位 | 研究 | 3級 | 15位 | 改修 | 1級 |
| 6位 | 人 | 4級 | 16位 | 設計 | 2級 |
| 7位 | 話 | 4級 | 17位 | 家 | 4級 |
| 8位 | 空間 | 1級 | 18位 | 関係 | 3級 |
| 9位 | 建物 | 4級 | 19位 | 形 | 3級 |
| 10位 | 論文 | 2級 | 20位 | 感じ | 2級 |

### 3－1－2 専門分野の複合名詞

名詞に専門性が反映される傾向があるという結果から、名詞を見出し語として採用したが、さらに各分野の専門家によって選定された複合名詞も見出し語とした。専門分野の語彙の中で複合名詞が重要な位置を占めていることは先行研究でも指摘されている。内山他（1999）は「専門分野で使用される専門用語には多くの複合語が含まれている。その複合語のほとんどは複合名詞である」と述べ、専門用語の複合語は分野毎に意味概念体系が異なるとしている。また小山（2010）は研究成果を記述するために用いられる言語記号を用語と呼び、用語の特徴として「基本的には文章内で名詞的機能を持つ」、「日本語では多くの用語は語幹レベルでの複合語として出現する」、「用語として認められるかどうかは、多分に主観的な判定基準が入ってくる側面も存在する」などを挙げている。

本研究チームもかねてより複合名詞の重要性を認識していた。しかしながら、日本語教師には専門分野の知識がないことから、ある語のどこまでがひとつの意味概念体系を示すのか判定基準を持ちえず、その選定は懸案事項であった。今回初めて専門家の協力を得て専門分野の複合名詞を見出し語として追加することが可能となった。専門家に選定された複合名詞の例

は4－2に示す。

## 3－3 用例の選定基準

研究現場に即した語彙、および表現を効果的に学習するために、レインボーでは見出し語がどのように使われているかを示す用例を選定した。ここでは用例の選定基準について述べる。

習慣的に共起する語と語のつながりは、一般的にコロケーション（collocation）[7]と呼ばれている。外国語教育では、コロケーションを重視した語彙教育のアプローチがある。Lewis(2000)はコロケーション能力の習得は、自然で母語話者に近い表現を産出するのに不可欠であるとしている。日本語教育では、曹・仁科（2006）、三好（2007）、三國・小森（2008）も母語話者によって産出された言葉の共起関係を提示することは、学習者が文を産出するために有効な指導法であると述べている。またSESJコーパスの研究では、専門性の高い名詞と旧日本語能力検定試験3、4級レベルの動詞や形容詞が組み合わされて使用される傾向が明らかとなり、コロケーションに注目した教材開発の必要性を主張している（猪狩他2009、菅谷他2009、単他2009；2010；2011、山口他2010）。

これまで､本研究チームではSESJコーパスの語彙全体の特徴や傾向を把握するために量的な調査を行っており、Stubbs(2002)の「中心語の左右数語の範囲内の高頻度共起」という定義を採用してきたが、SESJコーパスの教材化に際し、さらに広い範囲にわたって言葉のつながりに注目する必要性を感じた。砂川（2011）は広範囲の言葉のつながりについて、以下のように指摘している。

> 語と語の結びつきだけでなく、副詞とモダリティ要素との結びつきや、句と文末の否定辞やヴォイスとの関わりなど、語を超えた節レベルや、語より小さい形態素レベルの共起関係についても考える必要がある。
>
> （砂川2011：11）

砂川(2011)の指摘のように､本研究でも広範囲の様々な言葉の結び付き、

すなわち、見出し語の名詞と共起する様々な表現、動詞のヴォイスやアスペクト、副詞や連体修飾節、内容節などを含めて「共起表現」と定義し、共起表現を含む用例を優先的に選定することとした。なお、選定した用例の詳細については4－4と4－5で述べる。

## ４．レインボーの内容

本節ではレインボーの具体的な内容について説明する。レインボーは、研究室での発表やディスカッションで使用されている語彙や表現に対する留学生の理解を促し、また彼らが当該語彙や表現を運用する際の一助となることを目指している。レインボーは開発の途上にあり、現時点ではSESJコーパスデータから抽出した出現頻度20以上の名詞と専門家の判断によって取り上げられた複合名詞を見出し語としている。検索キーに調べたい語を入力すると、「見出し単語」「品詞」「よみ」「ローマ字」「対訳語」「カテゴリ」「学習項目」「例文」の順に表示される。「例文」とは本稿では用例と呼んでいるものである。用例は、見出し語の名詞が含まれる発話をSESJコーパスデータから適宜抽出したものである。見出し語検索では、漢字、ひらがな、ローマ字、英語のいずれかによる入力が可能である。専門分野は、単分野、複数分野の指定が可能である。

図1に複合名詞「脱塩素[8]」を検索した際のインターネット上の画面[9]を示す。画面上部に検索キーがあり、中央には言語や分野などが条件設定できる検索項目が位置する。画面下部には指定した検索結果が表示される。

現時点では「脱塩素」を検索するには、「だつ」という語の前方一致検索が可能である。ところが「えんそ」という後方一致検索はできず、この場合「塩素」の用例が表示されてしまう。そのため複合名詞を検索しやすくする部分一致検索システムの開発を考案中である。

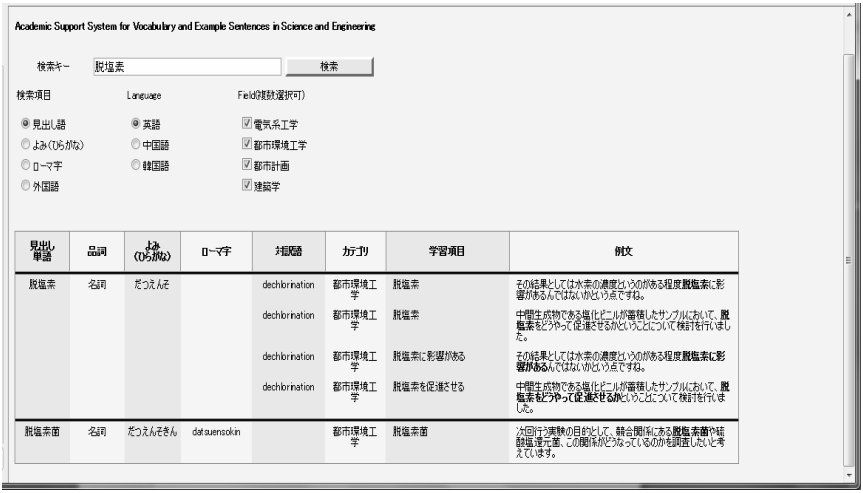


**図1 レインボー 検索結果画面**

次に、レインボーを構成する要素である対訳語や見出し語、用例などの内容について具体的に述べる。

## 4－1 対訳の付与

まず形態素解析により得た出現頻度数20以上の語彙全てに、英語、中国語、韓国語の訳語を付与した。表6は都市環境工学データの名詞語彙リストの一部である。

表6 頻度20以上の対訳付き名詞語彙（都市環境工学データより）

| 品詞 | 語彙 | アルファベット(ヘボン式) | ふりがな | 英訳 | 韓国語訳 | 中国語訳 |
|---|---|---|---|---|---|---|
| 名詞 | オゾン | ozon | おぞん | ozone | 오존 | 臭氧 |
| 名詞 | 量 | ryoo | りょう | quantity, amount, volume | 양, 분량, 도량 | 量 |
| 名詞 | 硫酸 | ryuusan | りゅうさん | sulfuric acid | 황산 | 硫酸 |
| 名詞 | 差 | sa | さ | difference | 차이 | 差 |
| 名詞 | 最後 | saigo | さいご | the last | 최후 | 最后 |
| 名詞 | 細菌 | saikin | さいきん | bacterium | 세균 | 细菌 |
| 名詞 | 最初 | saisho | さいしょ | the beginning | 최초 | 初始 |
| 名詞 | 酢酸 | sakusan | さくさん | acetic acid | 초산 | 醋酸 |
| 名詞 | サンプル | sanpuru | さんぷる | sample | 샘플 | 样本 |

## 4－2 見出し語の選定と複合名詞の見出し語への追加

レインボーの見出し語には、出現頻度20以上の名詞と専門家によって選定された複合名詞を立てた。専門家に選定された複合名詞には、以下のようなものがある。

専門家に選定された複合名詞の例：
活性汚泥、従属栄養細菌、残留塩素、界面活性剤、浄水場、高度浄水、脱塩素

例えば「従属栄養細菌」という複合名詞であるが、専門家によるとこの塊でheterotrophic plate countという一つの意味概念体系を示しており、一

語として扱われるべきであるという。しかし、当該分野の知識を持たない者から見れば「従属」と「栄養細菌」または、「従属」「栄養」「細菌」という形態素に分けることが可能であるようにも見えてしまい、専門外の者にとって、複合名詞の判定は極めて困難であることがわかる。またKH Coderではこのような複合名詞の抽出には限界がある。

## 4－3 専門家による学習優先度の判定作業

学習者個々人のニーズに沿った目的別学習を可能にするため、それぞれの見出し語について、分野の専門家に学習優先度の判定を依頼した。学習優先度[10]とは各分野の専門家から見て、その分野の研究において必要な語彙の学習優先度を指す。学習優先度は「G」General（大学入学以前に知っておくべき語彙）と「B」Basic（大学院での研究に必要な基礎専門語彙）と「S」Specialized（個々の専門性の高い語彙）の3つのカテゴリーに分類した。表7には都市環境工学のデータから例を挙げた。

**表7 学習優先度の3つの分類（都市環境工学のデータより）**

| |
|---|
| 「G」General＝大学入学以前に知っておくべき語彙 |
| 水、水素、硫酸、二酸化炭素、濃度、遺伝子、速度、微生物、塩素、活性、サンプル、データ、ベンゼン、アンモニア、イオン、メタン、実験、変化、培養、抽出、処理、利用、反応、分析、分解、平衡、影響、検出、検討、残留、汚染 |
| 「B」Basic＝大学院での研究に必要な基礎専門語彙 |
| イオン交換樹脂、硝化、浄水、排水、活性炭、バイアル、オゾン、**活性汚泥、従属栄養細菌、残留塩素、界面活性剤、浄水場、高度浄水** |
| 「S」Specialized ＝個々の専門性の高い語彙 |
| BS培地、**脱塩素** |

「G」判定は大学入学以前に知っておくべき語彙で、「水素、硫酸、メタン、ベンゼン、実験、変化」などが挙げられる。漢語、カタカナ語、サ変動詞の名詞部分が見られる。「B」判定は大学院での研究に必要な基礎専門語彙を指し、「イオン交換樹脂、活性炭、オゾン」などが挙げられる。「S」判定は各専門分野に特化した専門性が高い語彙で、「BS培地、脱塩素」が

挙げられる。

表7の太字で示した語彙は、専門家によって選定された複合名詞である。このように複合名詞は、カテゴリー「B」の基礎専門語彙と「S」の専門性の高い語彙に分類されることが多い。

現在、レインボーの見出し語には、学習優先度の判定が付与されていないが、将来的に学習優先度のレベルによって、語彙・用例を検索可能にすることなども検討している。いずれにせよ、効果的な日本語学習を進めるためにも、学習優先度の判定は重要な作業であると考えている。

## 4－4 用例の選定と整形作業

用例を選定する際に留意した点は次の3点である。

1）ゼミ発表やディスカッションの内容の理解を学習者に促すもの。
2）学習者が見出し語を含む表現を産出する場合に参考になるもの。
3）できるだけ格関係が明確であるもの。

以上の点を考慮しながら、日本語教師の視点で学習効果が高いと判断した用例をSESJコーパスのデータから目視で取り出した。次いで、取り出した文について研究チームのメンバーでその適否を検討したのち、最終的に用例として選定した。

用例の整形作業で考慮したのは、以下の4点である。話し言葉の特徴を生かしつつ、レインボーをより学習効果の高いシステムとして機能させるという観点から作業を行った。

1）長い発話は見出し語の名詞を中心にして、学習者が格関係をできるだけ明確に捉えられるように整形する。
2）フィラー、言いよどみ、繰り返しは基本的に削除する。
3）個人名は伏せ、固有名詞も差しさわりのあるものは、適宜判断して伏せる。
4）オノマトペ、倒置、無助詞、縮約形などの表現はそのまま残す。

上記2）のフィラー、言いよどみ、繰り返しは、話し言葉の特徴を表すものと考えられるが、レインボーの利用者として初中級レベルの日本語学習者も想定しており、文意を理解しやすくするために、これらは基本的に削除した。表8は見出し語が「検出」の整形前と整形後の例で、太字で記した

部分が削除および修正の対象となった箇所である。

**表8 用例整形**

| 見出し語 | 整形前 | 整形後 |
|---|---|---|
| 検出 | それなりの数の、**その**、検出事例っていうのが出てこないと、それを両者の相関っていうのは、なかなか、**あのー**、結論が出しづらいですよね。 | それなりの数の、検出事例っていうのが出てこないと、それを両者の相関っていうのは、なかなか、結論が出しづらいですよね。 |

## 4－5「学習項目」の選定

用例を整形したのち、「学習項目」という新たな項目を設け、一部の用例に付与した。この項目には、主に見出し語と共起関係にあると考えられる表現、および複合名詞を含む表現を選んだ。これらの表現は日本語教師の観点で重要であると考えられるもの、あるいは汎用性があるため学習効果が高いと判断したものである。レインボーではこれらの表現を効果的に学習させたいという狙いから「学習項目」と呼ぶ。

「学習項目」は、基本的に格関係を明らかにし、述語は汎用性をもたせるため原則的に辞書形とした。連体修飾節や内容節、ヴォイス、アスペクトの情報が見出し語と強く結びついていると判断した場合は、それらの形のまま取り上げた。具体的な「学習項目」の例は表9に示した。

表9 「学習項目」の例

| 見出し語 | 英訳 | 学習項目 | 用例 |
|---|---|---|---|
| 濃度 | concentration | **濃度が薄まる** | 地質由来のもので、**濃度が薄まってしまう**というか、比率として偏ってしまうっていうことはあり得ます。 |
| | | **濃度を規格化する** | まず、標準試料中の希土類元素濃度を用いて、**濃度を規格化します。** |
| | | **メタンが水の中に溶けている濃度** | 今現在、**メタンが水の中に溶けている濃度**を示しているわけじゃないんでしょ、これ。 |
| 従属栄養細菌 | heterotrophic plate count | **従属栄養細菌が検出される** | **従属栄養細菌が検出されました。** |

さらに4－2で述べた専門家によって選定された複合名詞だけでなく、日本語教師が選定した複合名詞も「学習項目」として積極的に取り上げることにした。ここでいう複合名詞は、本研究では「見出し語の語彙を含む2語以上の名詞が、助詞や助動詞の介入なしに結合した名詞」と定義する。表10に「希土類元素濃度」と「残留塩素濃度」を「学習項目」として取り上げた例を示す。

表10 複合名詞の「学習項目」

| 見出し語 | 英訳 | 学習項目 | 用例 |
|---|---|---|---|
| 濃度 | concentration | **希土類元素濃度** | まず、標準試料中の**希土類元素濃度**を用いて、濃度を規格化します。 |
| | residual chlorine | **残留塩素濃度** | 【地名】浄水場の**残留塩素濃度**が、低かったことですね。 |

表11には、「大腸菌の濃度」「従属栄養細菌の指標性」のように「名詞の名詞」といった表現を含む用例を示した。これらも言葉と言葉の結びつきが強いと考え、「学習項目」に取り上げた。

表11 「名詞の名詞」の「学習項目」

| 見出し語 | 英訳 | 学習項目 | 用例 |
|---|---|---|---|
| 濃度 | concentration | **大腸菌の濃度** | **大腸菌の濃度ですか。** |
| 従属栄養細菌 | heterotrophic plate count | **従属栄養細菌数の指標性** | **従属栄養細菌数の指標性**について検討を行いたいと思います。 |

「状態」のように名詞によっては内容節をとるものがある。このような内容節を含む表現も「学習項目」として取り上げた。表12にその例を示す。

表12 内容節をとる名詞の「学習項目」

| 見出し語 | 英訳 | 学習項目 | 用例 |
|---|---|---|---|
| 状態 | condition/ state | **菌が少ない状態** | じゃあ単離のためで、非常に**菌が少ない状態**で、付いてくれるのを待っている状態だよね。 |
| | | **ビニルクロライドが蓄積している状態** | ただ、こちらに示すのが、最初の、系2と呼ばれる試験結果なんですけども、**ビニルクロライドが蓄積している状態**で水素の添加を行っても脱塩素は進行しない。 |

表13にはサ変動詞を含んだ用例を示した。サ変動詞は「調査する」のように「する」の前項に名詞が位置するので、これらの語彙の場合、名詞の機能に加え、動詞として機能するものも「学習項目」に取り上げた。

表13 サ変動詞を含む用例の「学習項目」

| 見出し語 | 英訳 | 学習項目 | 用例 |
|---|---|---|---|
| 添加<br>(する) | addition | **水素を添加する** | この$Na_2SO_4$を添加した系がほかの系列と違ったのは、**水素を添加していない**場合に、まずジクロロエチレンがほとんど分解されなかったという点が挙げられます。 |
| | | **添加を行う** | 2番目が100ppmになるように、毎回**添加を行った**ものと、一番下が1000ppmのケースです。 |

表14には、助数詞を含んだ用例を示した。助数詞を含んだ用例の「学習項目」は、汎用性を持たせるため数字部分は【数字】と表記し、用例中には、差し支えのない限り実際に出現した数字を記載した。

表14 助数詞を含む用例の「学習項目」

| 見出し語 | 英訳 | 学習項目 | 用例 |
|---|---|---|---|
| 濃度 | concentration | **【数字】mL中の濃度** | この値というのは、その**500mL中の濃度** |

同一用例から複数の「学習項目」が採用可能な場合には、複数の「学習項目」欄を設けた。表15にその例を示す。

表15 同一用例中にある複数の「学習項目」

| 見出し語 | 英訳 | 学習項目 | 用例 |
|---|---|---|---|
| 濃度 | concentration | **メタンが水の中に溶けている濃度** | 今現在、**メタンが水の中に溶けている濃度**を示しているわけじゃないんでしょ、これ。 |
| | | **メタンが水の中に溶けている濃度を示す** | 今現在、**メタンが水の中に溶けている濃度を示している**わけじゃないんでしょ、これ。 |

## ５．課題

以上、レインボー開発までの経緯、そしてレインボーの開発に関わる具体的な作成過程について述べた。

レインボーは、SESJコーパス研究の成果物のひとつであり、理工学系分野の研究に従事する留学生の研究支援を目指したオンライン日本語学習支援システムである。現段階のレインボーは未だ試作版の域を出ず、さまざまな課題を抱えている。レインボーは工学系専門家および日本語教師の観点に基づいて開発されたものであり、今後、留学生の意見を反映させていく必要があるといえる。現在、本研究チームは留学生を対象としてモニター調査を行い、さまざまなフィードバックを得ている。例えば、「用例中の漢字の振り仮名や複合名詞の部分一致検索機能などが必要」、「用例の中には文脈が不明確なものがあるので、文脈がわかるようにしてほしい」、「頻度が低くても、重要な語彙があるので用例とともに載せてほしい」という意見もあった。

レインボーのデータは現在、電気系工学、都市環境工学、都市計画、建築学の4分野の中でも限られた研究室で録音されたものであり、かつ、出現頻度が20以上の名詞を中心に取り上げているため、理工学系のあらゆる分野のあらゆる語彙を網羅しているわけではない。また、ある語が、ある一定の頻度で使用されているからと言って、それが理工学系分野の話し言葉総体の特徴を表しているとは限らないことを、利用者にあらかじめ伝えておく必要がある。

SESJコーパスデータ拡充のために、これまでの工学系4分野からのデータに加え、電子情報学などの3分野でのデータも収集中である。今後は出現頻度が高い語彙のみならず、各分野において重要と考えられる語彙についても用例と「学習項目」を選定することを検討している。

またSESJコーパスを日本語教育にいかに活用していくことができるか、レインボーの他にもSESJコーパスの分析結果を用いた口頭表現の教材や漢字教材など、さらなる教材開発も進めたい。

最後に、レインボーの開発は、SESJコーパスの用例使用および公開の許諾をはじめとし、語彙の学習優先度の判定、複合名詞の選定など、各分野の専門家の協力と支援なしには成就できなかった。この一連の取り組みは、

各専攻の研究室との連携が不可欠であり、日本語教育を軸とした、フィールドを超えた連携の試みともいえる。今後、このSESJコーパス研究をいかに日本語教育の場の実践につなげ、理工学系分野に所属する全ての留学生、そして留学を目指している人々にいかに役立てることができるか、さらなる検討を重ねていきたい。

## 注

1.『理工学系語彙・用例学習支援システム　レインボー Rainbow』は工学系4分野（電気系工学、都市環境工学、都市計画、建築学）からのコーパスデータをもとに開発中の学習支援システムで、現在専門家からの許可を得た都市環境工学のデータを試作版として公開している。この4分野以外にも、現在電子情報学を含む3分野でのデータを収集中であり、将来的には工学系のみならず理学系の語彙や用例もデータに追加した包括的な学習支援システムの構築を目指している。そのためシステム名に「理工学系」と付けた。

2. 本稿では東京大学大学院工学系研究科に所属する日本語学習者だけでなく、非母語話者で理工学系分野の研究に従事する全ての人々を「留学生」と総称する。

3. 本研究チームが構築中の理工学系話し言葉コーパスをScience and Engineering Spoken Japanese Corpus at University of Tokyoの一部をとって、本稿ではSESJコーパス（仮称）とする。SESJコーパスの詳細については山﨑他（2010）の研究成果報告書を参照されたい。

4. 工学系分野の特性として、英語での研究活動が一般的である。中でも東京大学大学院工学系研究科では、修士、あるいは博士課程で入学してくる留学生に対して、英語特別コースが設置されており、このコースでは、授業、発表、論文作成などが全て英語で行われている。しかしながら、このような状況でも研究分野によっては日本語によるコミュニケーションが必要とされる場合がある。

5. KH Corderは、日本語のテキスト型データを計量的に分析するために開発されたツールである。その詳細については

http://khc.sourceforge.net/index.htmlを参照されたい。また本研究では品詞の認定はKH Coderの分類に準じている。例えば「研究する」のようなサ変動詞は、KH Coderでは「研究」はサ変名詞、「する」は動詞として扱われている。

6. 出現頻度20以上の語彙全てに、英語、中国語、韓国語の対訳語を付与した。レインボーはまだ開発中であり、中国語訳と韓国語訳については作業途中である。今後他の品詞や対訳語の表示についても検討中である。

7. コロケーションの定義は研究者によって異なるが、概ね狭義のコロケーションと広義のコロケーションに大別できる。村木（2007）はコロケーションを「自立的な単語のくみあわせで、命名（名づけ、さししめし）の側面のみをになった文法単位」（p.7）と狭義の定義をしている。一方、近年の日本語教育や日本語学の研究では、大曾・滝沢（2003）の「慣習によってまとまって使われる語の連鎖」（p.237）や野田（2007）の「語（または成分）と語（または成分）のつながり」（p.18）のように広義の定義が用いられることが多い。本研究チームでは語彙全体の量的な特徴や傾向を把握するためにStubbs(2002)の「中心語の左右数語の範囲内の高頻度共起」という比較的広義の定義を採用してきた。

8. レインボーはまだ試作版であり、本稿では既に専門家の公開許可を得た都市環境工学分野のデータを提示している。

9. レインボーの検索システムの開発にあたっては、合同会社シンタックスの協力を得た。

10. 語彙の学習優先度は各分野の専門家の判定によるものである。

## 参考文献

猪狩美保・岩﨑夕子・菅谷有子・単娜・古市由美子・村田晶子・山口真紀・山﨑佳子(2009)「工学系話し言葉コーパスにおける日本語の使用実態－使用頻度の高いサ変動詞の共起名詞を中心とした分析－」『言語文化と日本語教育』38，66-69．お茶の水女子大学日本言語文化学研究会

内山清子・竹内孔一・吉岡真治・影浦峡・小山照夫（1999）「専門分野における複合名詞の語構成要素の品詞相当カテゴリーに関する一考察」『学術情報センター紀要』11，49-57.　国立情報学研究所
大曾美恵子・滝沢直宏(2003)「コーパスによる日本語教育の研究－コロケーションおよびその誤用を中心に－」『日本語学』22(5)，234-244. 明治書院
大曾美恵子(2006)「日本語コーパスと日本語教育」『日本語教育』130，3-10. 日本語教育学会
鎌田倫子・古本裕子・笹原幸子・要門美規（2004）「日本語薬学会要旨集による薬学専門日本語の語彙調査」『研究紀要：富山医科薬科大学一般教育』32，51-59.
小宮千鶴子（2005）「理工系留学生のための物理の専門連語―高校教科書の調査に基づく選定―」『講座日本語教育』41,18-40.
小山照夫（2010）「日本語テキストからの複合語用語抽出」『情報知識学会誌』19(4)，306-315. 情報知識学会
菅谷有子・古市由美子・山﨑佳子（2008）「『工学系研究科日本語教室』におけるプログラム評価の一考察」『ヨーロッパ日本語教育』13，227-234.
菅谷有子・単娜・古市由美子・猪狩美保・村田晶子・山﨑佳子(2009)「工学系話し言葉コーパスにおける使用語彙の調査と分析－名詞を中心に－」『韓国日本語学会第20回国際学術発表会論文集－コーパスによる日本語・日本語教育の研究と応用－』37-45.
砂川有里子(2011)「日本語教育へのコーパスの活用に向けて」『日本語教育』150，4-18. 日本語教育学会
曹紅荃・仁科喜久子（2006）「中国人学習者の作文誤用例から見る共起表現の習得および教育への提言―名詞と形容詞および形容動詞の共起表現について―」『日本語教育』130，70－79. 日本語教育学会
単娜・猪狩美保・菅谷有子・古市由美子・山口真紀・山﨑佳子・岩﨑夕子(2009)「工学系話し言葉コーパスにおける日本語の使用実態－動詞を中心とした調査－」張威・山岡政紀編『日語動詞及其相関研究』378-389. 外語教学与研究出版社
単娜・猪狩美保・菅谷有子・村田晶子・古市由美子・山﨑佳子・山口真紀

(2010)「アカデミックな場面で使用される名詞のコロケーション－工学系話し言葉コーパスを用いた分析－」劉暁波他編『日語教育与日本学研究』203-207. 華東理工大学出版社
単娜・山口真紀・菅谷有子・古市由美子・村田晶子（2011）「ゼミ内発話における動詞の意味と用法のサンプル調査－特定分野に出現する名詞とのコロケーションにおける使用を中心に―」『言語文化と日本語教育』41，40-49. お茶の水女子大学日本言語文化学研究会
名古屋大学　国際化拠点整備事業（2010）『留学生のための専門講義の日本語』
野田尚史（2007）「文法的なコロケーションと意味的なコロケーション」『日本語学』26(12)，18-27.明治書院
林洋子(2004)「工学系修士論文口頭発表に用いられた語彙・表現」『専門日本語教育研究』6，25-32.
林洋子・国吉ニルソン・野口ジュディ・東條加寿子（2012）「日英の理工系口頭発表コーパスの構築と検索サイトJECPRESE」『第一回コーパス日本語学ワークショップ予稿集』273-282.
古市由美子・菅谷有子・岩﨑夕子・山﨑佳子(2008)「工学系Can-do Statementsの開発と実践―日本語能力評価基準の構築をめざして―」『二十一世紀における北東アジアの日本研究論文集』349-356. 北京日本学センター
三國純子・小森和子（2008）「コーパスを用いた論文作成のための慣用的共起表現の抽出」『小出記念日本語教育研究会論文集』16，55-68. 小出記念日本語教育研究会
三好裕子（2007）「連語による語彙指導の有効性の検討」『日本語教育』134，80-89. 日本語教育学会
村木新次郎(2007)「コロケーションとは何か」『日本語学』26(12)，4-17. 明治書院
山口真紀・菅谷有子・単娜・古市由美子・村田晶子(2010)「工学系話し言葉コーパスにおける和語動詞の使用実態－名詞との共起パターンの調査－」『専門日本語教育研究』12，41-46. 専門日本語教育学会
山﨑佳子・猪狩美保・岩﨑夕子・菅谷有子・単娜・古市由美子・村田晶子・山口真紀（2010）『工学系話し言葉コーパスの構築及びそれに基づく教材開発支援（研究代表者：山﨑佳子）』平成21年度科学研究費補助金挑

戦的萌芽研究　研究成果報告書（課題番号21652050）

Lewis, M. (ed.) (2000). *Teaching Collocation: Further developments in the lexical approach.* Language Teaching Publications.

Stubbs, M. (2002). *Words and phrases: Corpus studies of lexical semantics.* Oxford: Blackwell.

付記　本研究は平成23年度科学研究費補助金挑戦的萌芽研究（課題番号23652113）「研究支援を目指した『理工学系基本口頭表現用例学習辞典』の開発」を基に行っており、本論文は2012年3月15日、カナダのトロントで開催された春季AATJ（全米日本語教育学会）にて発表した内容に加筆修正を施したものである。