

「理工学系話し言葉コーパス」の構築とその応用
—理工学系7分野の語彙の実態調査—

**The development and its application of “the Science and Engineering
Spoken Japanese Corpus”**

-The analysis of vocabulary in the seven fields-

菅谷有子・伊藤夏実・遠藤直子・白鳥智美・関山聡之・成永淑・中村亜美・
古市由美子・宮部真由美・森幸穂・山口真紀, 東京大学

SUGAYA Yuko・ITO Natsumi・ENDO Naoko・SHIRATORI Tomomi・
SEKIYAMA Satoshi・SONG Yongsuk・NAKAMURA Ami・FURUICHI Yumiko
MIYABE Mayumi・MORI Sachiho・YAMAGUCHI Maki,
The University of Tokyo

1. はじめに

現在、来日する理工学系留学生（以下、留学生）は増加の傾向にあり、その多くは、来日直後より研究生生活を開始しなければならない環境にある（平成22年度文部科学省高等教育局学生・留学生課）。研究による時間的な制約、日本語のレベル差等、異なる学習条件や背景を持つ留学生に対して、日常生活と研究現場で役立つ日本語教育の内容と方法は何かということが、現在、我々教師にとっての課題となっている。

そこで、東京大学大学院工学系研究科に所属する日本語教師の有志によるコーパス研究チームは、理工学系留学生の研究生生活を包括的に支援するための目標言語調査として、研究室におけるゼミ発表や質疑応答などの自然発話を録音した。それを文字化し「理工学系話し言葉コーパス」として構築、さらにそのコーパスを資源とした教材開発に着手することにした。

本コーパスの構築は2007年に開始し、本学の理工学系7分野にわたる整備がこのたび完了した。本稿では、その概要と、その応用としての教材化の実践例を報告したい。

具体的には(1)コーパスの構築過程、(2)各分野における自立語（名詞、動詞、形容詞等）の分布の特徴、(3)分野にまたがって使用される自立語の特徴について報告する。さらに(4)本コーパスを資源とした教材、即ち教育実践の応用についてもあわせて紹介する。

2. 各分野の専門教員との連携によるコーパスの構築過程

大学院の理工学系ゼミを録音・データ化する過程には、公開には適さない研究

途中の情報、知的財産、個人情報など、多くのハードルが存在する。しかし、それを可能にし、その後の研究活動を促進させたのが以下の調査活動であった。

2-1. 目標言語調査の必要性を認識させた CDS 作成

大学院レベルの留学生に有用かつ必要な日本語教育の内容と方法を模索する中で、当研究科日本語教室独自の CDS (Can-do Statement) (古市ほか 2008) の開発が行われた。その結果、目標言語調査の必要性が認識された。

2-2. 研究室におけるゼミの録音の突破口となったプログラム評価

次に、当研究科日本語教室のプログラム評価を実施した。その結果、専門分野の教員と留学生の双方よりゼミ発表及びその質疑応答を理解できる日本語力が必要との声が上がった。この調査に応じた専門分野の教員の協力を得て、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の録音が可能となった (菅谷ほか 2008)。その結果、現在に至る 7 分野にわたる「理工学系話し言葉コーパス」の構築が実現したのである。

2-3. コーパス構築の必要性を裏付けた留学生の言語使用実態調査

上記 2-2. の調査・研究活動と並行して、留学生を対象とした日本語及び英語の使用実態調査を行った (『工学系研究科日本語教室報告書 2010, 2013 年度』)。調査の結果、理工学系の留学生にとって必要な日本語は、「生活のための日本語力」であると同時に「ゼミの発表及びその質疑応答を理解し参加を可能にする日本語力」であることが明らかになった。この言語使用実態調査によって、コーパスを構築し、それを留学生に対する日本語教育に応用する必要性が裏付けられた。

以上、上記のような一連の調査研究活動を通じて、専門分野の教員との協力体制を築き上げていくことで分野を広げての録音データの収集、コーパスの構築が実現していくこととなった。

3. 「理工学系話し言葉コーパス」の概要

3-1. コーパスの規模およびその語彙の分布

本コーパスは理工学系 7 分野の研究室のゼミを 5 年にわたって収録した中から、主に母語話者の自然発話を抽出したもので、その内訳は<表 1>に示す。7 分野の収録時間は 153 時間である。

また、延べ形態素出現頻度数は 1,602,972、異なり形態素数は 16,126 である。分野別の形態素数は<表 2>に示す。

形態素解析プログラムとしては、解析エンジンとして MeCab、解析辞書として UniDic を備えた「茶まめ」を使用した¹。なお、専攻分野名は、今後本稿では<表 1>の () で示した略称を用いることにする。

＜表 1＞7 分野の収録期間及び収録時間

専攻分野	収録期間	収録時間
電気系工学（電気）	2007年11月～2008年2月	約20時間
都市環境工学（環境）	2008年12月～2009年2月	約23時間
都市計画（計画）	2008年12月～2010年7月	約22時間
建築学（建築）	2009年6月～2009年12月	約26時間
社会基盤学（社基）	2010年4月～2010年7月	約23時間
化学システム工学（化シス）	2011年5月～2012年2月	約24時間
情報理工・電子情報学（情理）	2011年10月～2012年2月	約15時間
7分野まとめ	2007年11月～2012年2月	約153時間

＜表 2＞コーパスの規模・7 分野の延べ形態素数と異なり語数

	電気	環境	計画	建築学	社基	化シス	情理	7分野計
形態素延べ	162,154	171,963	205,235	398,456	217,451	247,846	199,867	1,602,972
異なり形態素数	3,650	4,319	5,544	7,269	5,788	5,013	3,457	16,126

次に、自立語及びその他の形態素（主として付属語）の全分野における分布を示すと＜表 3＞のようであった。なお、本稿では、自立語として名詞、動詞、形容詞、副詞、接続詞、連体詞を対象とし、感動詞はその他の形態素に含めて分析している。

＜表 3＞自立語及び付属語を含むその他の形態素の分布

分類 \ 分布	延べ		異なり	
	出現頻度	自立語に占める割合	語数	自立語に占める割合
主要な自立語	797,265	49.7%	15,208	94.3%
その他の形態素	805,707	50.3%	918	5.7%

＜表 3＞延べで見ると、自立語とその他の形態素の出現頻度は変わらないが、異なりで見ると、自立語は全体の 94 パーセント以上を占めており、多様な語彙が使用されていることが分かる。

そこで、本稿では、このように豊富な語彙が出現する自立語について、7 分野におけるその出現の分布を調査し、特徴を分析することにした。

3-2. 各分野の自立語（名詞・動詞・形容詞・副詞・接続詞・連体詞）の分布
各分野の自立語の出現頻度の延べ及び異なり語数を<表4>に示す。

<表4>7分野の自立語分布

品詞 \ 分野	電気	環境	計画	建築	社基	化シス	情理	7分野 統合
名詞延べ語数	42,813	48,436	60,159	111,997	61,303	71,814	58,919	455,441
分野総形態素数割合	26.4%	28.2%	29.3%	28.1%	28.2%	29.0%	29%	28.4%
名詞異なり	2,440	3,006	4,045	5,451	4,143	3,508	2,289	12,614
異なり語数割合	66.8%	69.6%	73.0%	75.0%	71.6%	70.0%	66.2%	78.2%
動詞延べ	18,307	21,632	25,876	52,047	26,635	29,128	22,642	196,267
分野総形態素数割合	11.3%	12.6%	12.6%	13.1%	12.2%	11.8%	11%	12.2%
動詞異なり	496	560	696	846	781	664	502	1,530
異なり語数割合	13.6%	13.0%	12.6%	11.6%	10.7%	13.2%	14.5%	9.5%
形容詞延べ	3,914	3,236	4,393	8,023	4,758	5,156	3,423	32,903
分野総形態素数割合	2.4%	1.9%	2.1%	2.0%	2.2%	2.1%	1.7%	2.1%
形容詞異なり	70	81	93	110	104	96	70	165
異なり語数割合	1.9%	1.9%	1.7%	1.5%	1.8%	1.9%	2.0%	1.0%
副詞延べ	6,861	5794	9560	16050	9,771	9,590	7,601	65,227
分野総形態素数割合	4.2%	3.4%	4.7%	4.0%	4.5%	3.9%	4%	4.1%
副詞異なり	171	192	187	258	225	226	173	470
異なり語数割合	4.7%	4.4%	3.4%	3.5%	3.9%	4.5%	5.0%	2.9%
接続詞述べ	1,264	1,420	2240	3,602	2,195	2353	1,635	14,709
分野総形態素数割合	0.8%	0.8%	1.1%	0.9%	1.0%	0.9%	1%	0.9%
接続詞異なり	19	20	19	22	19	18	17	23
異なり語数割合	0.5%	0.5%	0.3%	0.3%	0.3%	0.4%	0.5%	0.1%
連体詞述べ	2,897	3,089	4,523	8,044	4,723	4,750	4,692	32,718
分野総形態素数割合	1.8%	1.8%	2.2%	2.0%	2.2%	1.9%	2%	2.0%
連体詞異なり	22	17	18	24	19	20	17	28
異なり語数割合	0.6%	0.4%	0.3%	0.3%	0.3%	0.4%	0.5%	0.2%

7分野の自立語の分布については、各分野とも、延べ語数、異なり語数とも名詞が全体に占める割合が最も高く、次いで動詞が高くなっている。延べ語数、異なり語数とも最も低いのが接続詞であり、異なり語彙数も各分野ともあまり差はない。次に低いのが連体詞であり、両品詞とも、語彙の使用が限定されていると言える。一方、副詞と形容詞については、延べ語数に関しては副詞が形容詞のほぼ2倍となっている。しかし、異なり語数に関しては、情理以外の分野では形容詞の方がわずかに多いものの、副詞と形容詞に大きな差は見られない。即ち、副詞の方が限られた語彙を高頻度で使用しているということである。

形容詞をイ形容詞、ナ形容詞に分けて内訳を見ると、<表5>のようになる。延べの出現頻度はイ形容詞がナ形容詞の3倍となっているが、異なり語数に関しては、ナ形容詞がイ形容詞の2倍以上である。

＜表 5＞イ形容詞・ナ形容詞の分布

品詞 \ 分野	電気	環境	計画	建築	社基	化シ ス	情理	7分野 統合
イ形容詞延べ	3,159	2,384	3,082	5,834	3,700	3,905	2,663	24,727
分野総形態素数割合	1.9%	1.4%	1.8%	1.5%	1.7%	1.6%	1%	1.5%
イ形容詞異なり	70	81	93	110	104	96	70	165
異なり語数割合	1.9%	1.9%	1.7%	1.5%	1.8%	1.9%	2.0%	1.0%
ナ形容詞延べ	755	852	1,311	2,189	1,058	1,251	760	8176
分野総形態素数割合	0.5%	0.5%	0.8%	0.5%	0.5%	0.5%	0%	0.5%
ナ形容詞異なり	115	127	171	200	151	163	101	378
異なり語数割合	3.2%	2.9%	3.1%	2.8%	2.1%	3.3%	2.9%	2.3%

4. 分析

4-1. 7分野における名詞について

まず、語種に関しては、＜表 6＞に示した通り、自立語（延べ）については、和語の出現頻度が最も高く、次いで漢語、外来語の順となっている。ところが、名詞（延べ）では、漢語が 5.4 割、和語が約 3 割、外来語は約 1.4 割を占めている。さらに、名詞の異なり語を語種の分布で見ると、漢語が最多で、次いで外来語、和語となり、外来語は 2 割を占め、相対的に増加していることが分かる。

＜表 6＞自立語（延べ）・名詞（延べ）・名詞（異なり）語種別分布

語種	自立語（延べ）		名詞（延べ）		名詞（異なり）	
和語	435,990	64.30%	129,785	28.90%	2254	18.80%
漢語	183,072	27.00%	243,450	54.20%	5423	45.10%
外来語	37,709	5.60%	61,205	13.60%	2413	20.10%
混種語	8,447	1.20%	3,425	0.80%	218	1.80%
記号	1,015	0.10%	1,019	0.20%	641	5.30%
固有名詞	8,447	1.20%	10,123	2.30%	1028	8.60%
不明	3,446	0.50%	271	0.10%	40	0.30%
計	678,126	100.00%	449,278	100.00%	12017	100.00%

即ち、少なくとも調査対象の研究室ゼミにおいては、話し言葉を理解するうえで漢語、即ち漢字語彙の知識が重要であることを示唆している。また、外来語に関しても、延べの出現頻度は少ないものの、異なり語彙として和語以上に幅広い分布があることを明記したい。

以上のことから、延べ・異なりとも出現頻度が高い名詞においては、漢語を中

心とした多様な語彙を高頻度で使用している実態が観察される。従って、語彙教育の視点からは漢語を視野に入れた名詞語彙の理解が、研究室の話し言葉を理解するうえで大きな鍵となると言えよう。

4-3. 各分野における名詞の特徴

<表 7> 分野別高頻度出現名詞（上位 30 語）

分野	電気	頻度	環境	頻度	計画	頻度	建築	頻度	社基	頻度	化シス	頻度	情理	頻度
1	事	861	事	1115	事	1509	事	2837	事	1227	事	1387	事	1204
2	訳	754	物	498	間	828	物	1324	間	905	粒子	845	間	687
3	制御	619	方	410	人	573	所	870	所	541	物	797	数	596
4	場合	452	所	399	物	499	建築	858	後	524	間	781	物	580
5	図	384	間	388	後	434	今	818	今	459	後	615	音	544
6	計算	316	濃度	327	今	416	後	714	方	418	合成	610	音声	491
7	数	311	菌	319	所	390	方	711	波	381	時	492	話者	490
8	出力	310	処理	307	都市	388	研究	686	時	335	構造	403	単語	458
9	発電	305	後	288	方	333	中	622	風	333	今	391	時	398
10	変動	300	結果	268	地域	329	間	563	物	327	所	391	方	386
11	方	293	時	258	中	315	風	557	計算	319	方	390	発声	379
12	時	289	実験	249	時	310	住宅	478	研究	283	結晶	388	訳	342
13	結果	271	検出	241	風	290	人	476	人	274	ゼオライト	385	後	340
14	値	260	量	236	研究	270	解体	473	データ	256	吸着	313	実験	334
15	系統	249	今	231	問題	261	話	467	地形	217	水	304	認識	306
16	周波	248	水	219	訳	260	時	466	辺	210	シリカ	300	音節	306
17	容量	238	訳	214	コミュニティー	234	技術	412	話	208	結果	291	特徴	293
18	エリア	234	塩素	209	計画	231	問題	408	津波	202	研究	263	今	271
19	風	230	中	203	自分	225	訳	399	中	196	実験	258	母音	269
20	今	224	濃縮	199	話	200	他	377	感じ	188	シード	247	所	256
21	事故	216	風	192	実際	194	建物	344	奴	175	種	246	シラブル	252
22	蓄電	214	濾過	186	意味	189	改修	343	大体	171	中	242	音素	246
23	ピッチ	206	細菌	183	場合	175	設計	332	実験	166	条件	239	モデル	245
24	利益	200	浄水	182	関係	171	調査	329	砂	166	場合	235	周波	241
25	モデ	197	数	180	環境	149	必要	326	先	159	量	234	話	239
26	物	196	データ	178	地	148	論文	317	結果	159	ポリマー	233	量	233
27	後	194	添加	177	場所	146	情報	313	モデル	153	反応	197	ベクトル	232
28	量	192	図	174	スラム	143	先生	297	海岸	152	今回	187	風	228
29	電力	192	分解	173	住宅	140	空間	296	技術	150	為	181	場合	225
30	角	189	奴	169	高齢	136	構造	294	問題	145	時間	176	定義	175

7 分野全てにおいて、延べ語数、異なり語数とも最も出現頻度が高く、漢語の使用頻度の高い名詞だが、各分野に高頻度で出現する語彙にどのような特徴があるのか、上位 30 語を<表 7>で示す。なお、「茶まめ」による解析では、サ変動

詞は「名詞-サ変可能」として分類されるので、この表では灰色にして示した。各分野とも、その分野に特徴がある語彙が散見できる。電気では「制御・出力・発電」、環境では「濃度・処理・塩素・浄水」、計画では「都市・コミュニティー・スラム」、建築では「建築・住宅・解体・」、社基（旧土木工学）では「地形・津波・海岸」、化シスでは「粒子・合成・結晶」、情理では「音声・発声・音節」など、各研究分野の研究内容が反映された語彙の高頻度の出現が確認された。

分野に共通して上位に出現するのは、「事・間・人・所・物・時」等の基礎語彙である。

4-3. 名詞以外の自立語について

7 分野に出現する他の自立語であるが、紙面の関係上本稿では名詞以外は全分野に共通して出現する語彙のみを次節で紹介する。

4-4. 7 分野に共通する名詞以外の自立語について

それでは、分野間に共通する語彙の分布はどのようなものか、<表 8>に示す。分野にまたがって出現する異なり語は名詞が最も多く、次いで動詞、副詞、ナ形容詞、イ形容詞、接続詞、連体詞の順となっている。

また、7 分野に共通して出現する語彙を各々の品詞における出現率で見ると、高い順に、連体詞、接続詞、副詞、動詞、イ形容詞と、いずれも延べ語数の 90%以上をカバーしている。しかし、ナ形容詞は 69%、名詞に至っては 33%とカバー率が低くなっている。

<表 8> 7 分野共通語彙の品詞内出現頻度及びその比率

7 分野共通語	名詞	動詞	イ形容詞	ナ形容詞	副詞	接続詞	連体詞
異なり語数	454	216	37	40	88	16	15
出現頻度	151,839	184,548	23,145	5,645	63,071	14,568	32,609
品詞内出現率	33.34%	94.03%	93.60%	69.04%	96.69%	99.04%	99.67%

このことから、名詞はコーパス全体での出現頻度及び異なり語数は多いものの、7 分野全体で共通して使用されるというよりは、各分野にそれぞれ特徴のある語彙が分散している傾向がある。一方、ナ形容詞は別として、その他の品詞については、重複して出現する語彙がその品詞の異なり語のほぼ全体をカバーしている。従って、語彙教育の観点から、名詞に関しては各分野の特徴がある語彙教育を、また、動詞、イ形容詞、副詞、接続詞、連体詞に関しては 7 分野に共通する語彙

の教育を優先することが妥当であると考ええる。

なお、7分野に共通する自立語の一覧は上位30語までを<表9>に示す。

<表9>7分野に共通する自立語

	名詞	頻度合計	動詞	頻度合計	イ形容詞	頻度合計	ナ形容詞	頻度合計	副詞	頻度合計	接統詞	頻度合計	連体詞	頻度合計
1	事	10140	言う	32538	無い	8689	こんな	563	まあ	11373	で	9501	其の	14482
2	間	4231	為る	31677	良い	4536	そんな	517	そう	10635	では	1528	此の	8589
3	物	4221	居る	12028	大きい	1074	ばらばら	396	一寸	5517	又	1030	彼の	4759
4	後	3109	有る	11008	高い	944	意外	370	こう	5488	そして	541	同じ	1007
5	所	3034	成る	7792	凄い	726	可能	335	どう	3863	唯	519	或る	683
6	方	2941	遣る	5604	多い	676	確か	295	もう	3251	或いは	406	何の	600
7	今	2810	思う	5487	小さい	605	确实	265	例えば	1702	及び	180	色んな	546
8	時	2548	行く	4752	甘い	484	完全	241	先ず	1468	更に	180	こんな	541
9	訳	2200	見る	4294	難しい	476	簡単	218	矢張り	1279	然し	131	そんな	499
10	風	1865	出来る	4213	面白い	462	奇麗	199	多分	1248	が	114	どんな	351
11	研究	1793	来る	4154	少ない	423	結構	189	余り	1034	若しくは	106	大きな	279
12	中	1791	分かる	3592	長い	321	嫌	184	未だ	939	から	97	所謂	169
13	人	1643	考える	2894	新しい	313	懸命	183	一番	899	且つ	71	小さな	45
14	場合	1615	使う	2364	低い	296	厳密	172	結構	786	けれど	59	単なる	30
15	話	1472	出る	2283	悪い	281	最適	162	良く	757	但し	53	大した	29
16	結果	1342	抛る	2047	強い	251	自然	121	少し	710	一方	52		
17	数	1304	得る	1871	近い	226	十分	111	可成	660				
18	実験	1142	作る	1670	細かい	226	重要	98	色々	642				
19	量	1124	違う	1553	早い	204	色々	97	若し	563				
20	実際	1087	就く	1431	欲しい	195	新た	81	もっと	518				
21	問題	1074	書く	1362	短い	165	大雑把	71	全然	489				
22	データ	1051	入れる	1214	物凄い	160	大事	61	ちゃんと	460				
23	構造	1044	持つ	1125	広い	141	大丈夫	58	成る程	420				
24	感じ	999	入る	1114	可笑しい	137	大変	57	特に	387				
25	図	989	行う	1064	正しい	130	単純	54	又	329				
26	為	986	知れる	949	でかい	124	適正	51	元々	314				
27	他	984	出す	942	詳しい	115	適切	49	詰まり	312				
28	前	979	取る	933	古い	104	当たり前	45	ずっと	300				
29	奴	920	対する	903	深い	99	当然	42	何故	279				
30	先	905	仕舞う	872	宜しい	90	同じ	41	取り敢えず	277				

5. コーパスの日本語教育現場における応用実践例

以上の結果を踏まえ、現在当研究チームによる実践例を紹介したい。

5-1. 『理工学系語彙・用例学習支援システム レインボー』（以下レインボー）

まず、専門分野の特徴的な語彙が多く出現した名詞に関しては、当研究チームが、留学生のために作成したオンライン日本語学習支援システム『レインボー』（森ほか2012）（伊藤ほか2013）があり、その一部を試験的に公開している。出現頻度が高い名詞（頻度20以上）を見出し語とし、その名詞の訳語（英語・中国

語・韓国語) と、その名詞が含まれる用例を提示している。また、学習効果が高いと判断した共起表現や、専門用語が多く含まれる複合語も、専門分野の教員の協力を得て選定して提示するなど、各分野の特徴を踏まえた教材として開発している。

5-2. 漢字教材

語種として多く出現した漢字については、専攻の異なる留学生を対象とした「専門語彙・漢字クラス」における教材(遠藤 2013)である。本コーパスを資源とした単漢字や漢字語彙を用いて、短文作成をする際、学習者に文脈設定を明確にさせ、「読む・書く」だけでなく「聞く・話す」ことを意識化させる漢字語彙教育を行っている。

また、7分野のうち2分野、環境と計画の留学生を対象としたクラスでは、主教材の漢字提出順に従った漢字学習の際に、研究に必要な語彙や共起表現も導入する補助教材の作成を試みている。教材作成にあたっては、コーパスの用例を参照して、学習者の研究生活に直接役立つものを目指している。

6. まとめ

以上、本稿では、理工系という特定の分野におけるコーパスの構築及びそのコーパスの概要、そしてその教育への応用について紹介した。

極めてデータ収録が困難な研究室の話し言葉の録音が実現できたのは、長期にわたる実践研究がその背景にあり、その結果を各分野の専門教員と共有し、協力を得られたことが最も大きな要因であったと考える。

7分野の自立語は、延べ語数、異なり語数とも名詞が全体に占める割合が最も高く、研究室のゼミで話し言葉を理解するには、各専門分野の特徴的な漢語や外来語を含む名詞の教育が重要であることが示唆された。7分野に共通して出現する連体詞、接続詞、副詞、動詞、イ形容詞はいずれも延べ語数の9割以上をカバーしており、7分野に共通する語彙の教育が可能であろう。この傾向は、4分野を対象としたコーパス研究の成果(山崎 2010)と重なるものである。今後は、現在開発を進めているウェブ上の用例辞典『レインボー』に統合する形で、随時研究成果を公開していく予定である。また、漢字教材に関しては、コーパスデータを活かし学習者の使用領域に配慮したオーダーメイドの教材作成が可能となる。

日本語教師自らによる教育現場に即したコーパスの構築とその応用は、各専門分野で学ぶレベル差のある多様な学習者の環境に対応可能な日本語教育を可能にする。この試みは、教育内容の改善に大いに資するものと期待している。

今後の課題としては、コーパスデータを用い、語彙・漢字レベルにとどまらず

会話技術・表現の習得等も視野に入れた分析を行っていく予定である。また、市販されている教材との比較も行い、教材としての充実を図るべく、更なる検討を加えていきたいと考える。

<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」を基に行っている。>

参考文献：

1. 平成 22 年度文部科学省高等教育局学生・留学生課『我が国の留学制度の概要—受入れ及び派遣—』
http://www.mext.go.jp/component/a_menu/education/detail/_icsFiles/fieldfile/2011/12/12/1286521_4.pdf
2. 古市由美子・菅谷有子・岩崎夕子・山崎佳子（2008）「工学系 Can-do Statements の開発と実践—日本語能力基準の構築をめざして—」北京日本学研究中心編『二十一世紀東北亜日本研究論文集』349-356 学苑出版社
3. 菅谷有子・古市由美子・山崎佳子「『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 号 227-234
4. 山崎佳子・猪狩美保・岩崎夕子・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およびそれに基づく教材開発支援（研究代表者：山崎佳子）』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）
5. 森幸穂・伊藤夏実・遠藤直子・菅谷有子・成永淑・古市由美子（2012）「理工学系話し言葉コーパスによる学習語彙用例検索ツールの構築—学習者のモニター調査より—」『日本語教育国際研究大会』（於：名古屋）
6. 伊藤夏実・遠藤直子・菅谷有子・成永淑・古市由美子・森幸穂（2013）「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 Rainbow の開発』『横浜国立大学留学生センター 教育研究論集』21 号 115-157
7. 遠藤直子・伊藤夏実・菅谷有子・古市由美子・森幸穂（2013）「理工学系話し言葉コーパス」の構築と専門語彙・漢字教育への応用—文脈を明確にする文作成—」日本英語教育学会（於：早稲田大学）

¹今回の調査では、形態素解析プログラム「茶まめ」で使用されている形態素解析辞書 UniDic の品詞分類に従って語彙調査を行った。語彙の表記も同辞書の見出し「語彙素」の表記に従った。なお、今回の形態素解析においては、補助記号など書き言葉において用いられるものは解析対象から外した。