英語コーパス学会
第 39 回大会資料

日時：2013 年 10 月 5 日（土）– 6 日（日）
会場：東北大学（川内北キャンパス）マルチメディア教育研究棟
（http://www.tohoku.ac.jp/japanese/profile/campus/01/kawauchi/index.html）
〒980-8576 仙台市青葉区川内 41

英語コーパス学会 2013

■第 1 日目

ワークショップ **1**【統計解析環境 R による統計処理の基本 —検定と視覚化—（初級編）】
会　場：東北大学マルチメディア教育研究棟 2 階 M201 教室
日　時：10 月 5 日（土）10:00–12:00（9:30 受付開始）
講　師：阪上辰也（広島大学）
定　員：定員 60 名（先着順・要予約）
参加費：会員無料。非会員 2,000 円（当日会員としての大会参加費二日間共通）。
※予約申し込みは，参加を希望するワークショップ，氏名，所属，会員・非会員の別を記入の上，電子メールで jaecs.workshop@gmail.com まで（申込締切日：9 月 30 日）。

日　　時　　2013 年 10 月 5 日（土）
受付開始　　12:20（マルチメディア教育研究棟 1 階エントランスホール）
開 会 式　　13:20（マルチメディア教育研究棟 2 階 M206 マルチメディアホール）
司　会　田畑智司（大阪大学）
1. 会長挨拶　堀　正広（熊本学園大学）
2. 開催校挨拶　浅川照夫（東北大学）
3. 総会
4. 学会賞審査報告
5. 事務局からの連絡

〈研究発表第 1 室（マルチメディア教育研究棟 2 階 M206 マルチメディアホール）〉
司　会　仁科恭徳（明治学院大学）

研究発表 1　14:20–14:50
Parallel English-Japanese corpus with meaning representations as educational aid
Alastair Butler（東北大学）

研究発表 2　14:55–15:25
The design, development and research potential of the Kansai University Bilingual Essay Corpus
Miho Yamashita（関西大学大学院生）

研究発表 3　15:30–16:00
Contrastive lexical analysis of spoken and written interlanguage productions by Japanese learners of English: a study based on the newly added spoken module of the ICNALE
Shin'ichiro Ishikawa（神戸大学）

〈研究発表第 2 室（マルチメディア教育研究棟 2 階 M201 教室）〉
司　会　新井恭子（東洋大学）

研究発表 1　14:20–14:50
類義的な動詞不変化詞構文における不変化詞のアスペクト特性: *up* と *down* を中心に
大谷直輝（京都府立大学）

研究発表 2　14:55–15:25
Word, time and practice: a preliminary corpus-based comparative study on paired conversations of advanced-level learners and basic-level learners　Keiko Tsuchiya（東海大学）

研究発表 3　15:30–16:00
The NICT JLE (Japanese Learner English) Corpus の XML 整形式化と
それを使った習得段階別の語用論的言語特徴の分析　三浦愛香（東京経済大学）
佐野　洋（東京外国語大学）

〈休　憩　16:00–16:20〉

シンポジウム　16:20–18:20（マルチメディア教育研究棟 2 階 M206 マルチメディアホール）
《コーパスが語ること，コーパスが語らないこと》　司　会　金澤俊吾（高知県立大学）
講　師　浅川照夫（東北大学）
大室剛志（名古屋大学）
金澤俊吾（高知県立大学）

《懇 親 会　時間：18:30–20:30　場所：Bee ARENA Café；会費：5,000 円》

■第 2 日目

ワークショップ **2**【統計解析環境 R による言語データの分析（中級編）】
会　場：東北大学マルチメディア教育研究棟 2 階 M201 教室
日　時：10 月 6 日（日）9:30–11:00（9:10 受付開始）
講　師：阪上辰也（広島大学）
定　員：定員 60 名（先着順・要予約）
参加費：会員無料。非会員 2,000 円（当日会員としての大会参加費：5 日の参加費納入者は不要）。
※予約申し込みは，参加を希望するワークショップ，氏名，所属，会員・非会員の別を記入の上，
電子メールで jaecs.workshop@gmail.com まで（申込締切日：9 月 30 日）。

日　　時　　2013 年 10 月 6 日（日）
受付開始　　10:40（マルチメディア教育研究棟 1 階エントランスホール）

講演　　11:15–12:30（マルチメディア教育研究棟 2 階 M206 マルチメディアホール）
《計量的文献研究とコーパス》

司　会　高橋　薫（東京理科大学）
講　師　村上征勝（同志社大学）

〈昼休憩 12:30–13:30〉

〈研究発表第 1 室（マルチメディア教育研究棟 2 階 M206 マルチメディアホール）〉
司　会　Judy Noguchi（武庫川女子大学）

研究発表 1　　13:30–14:00
日本語母語話者の英語研究論文におけるヘッジ使用　　若松弘子（筑波大学大学院生）

研究発表 2　　14:05–14:35
教育利用可能なパラレルコーパス検索プラットフォームの構築に向けて
中條清美（日本大学）
アントニ・ローレンス（早稲田大学）
赤瀬川史朗（Lago 言語研究所代表）
西垣知佳子（千葉大学）
水本　篤（関西大学）
内山将夫（情報通信研究機構）

研究発表 3　　14:40–15:10
JEFLL Corpus Parallel Version の構築と活用　　投野由紀夫（東京外国語大学）

〈研究発表第 2 室（マルチメディア教育研究棟 2 階 M201 教室）〉
司　会　家口美智子（摂南大学）

研究発表 1　　13:30–14:00
心理言語学的語彙特性の語彙の豊かさ指標への応用可能性の検討
石井卓巳（筑波大学大学院生）

研究発表 2　　14:05–14:35
学習モデルとしての教科書用例の改善：使役動詞の記述例に基づく研究
井上　聡（環太平洋大学）

研究発表 3　　14:40–15:10
司法英語における類義語をどう活用発信型辞書に記述すべきか
鳥飼慎一郎（立教大学）
溜箭将之（立教大学）

〈研究発表第 3 室（マルチメディア教育研究棟 2 階 M204 教室）〉
司　会　石川慎一郎（神戸大学）

研究発表 1　　13:30–14:00
基本句を考慮した *n*-gram の計数　　田中省作（立命館大学）

研究発表 2　　14:05–14:35
品詞／機能語列の分布に基づいた非母語話者英語に潜在する言語系統樹の構築
永田　亮（甲南大学）

研究発表 3　　14:40–15:10
Personalised statistical writing analysis　　John Blake（北陸先端科学技術大学院大学）

閉 会 式　　15:20（マルチメディア教育研究棟 2 階 M206 マルチメディアホール）
閉会の辞　　岡田　毅（東北大学）

■1 日目

【ワークショップ 1】

統計解析環境 R による統計処理の基本 —検定と視覚化—（初級編）

阪上辰也（広島大学）

言語データの分析を行う際，用例数を集計したり，得られた数値の分布をグラフで視覚化したりすることが少なくない。多くの場合，Microsoft ExcelやSPSSといった商用ソフトを使うことになる。操作性はよく初めてでも手軽に扱えるという点ではよいが，その反面，ソフトが行う処理の信頼性やグラフ描画の美しさなどで劣る点もある。そこで，本ワークショップでは，無償で利用可能なオープンソースソフトウェアである統計処理環境Rを利用し，その基本的な操作，よく使われるであろう検定手法とデータを視覚化する方法について学ぶ。

具体的には，Rの導入方法からデータの読み込みといった基本操作から順を追って説明する（使用するのはR version 3.0.1）。続いて，コーパスから得られた数値データを集計した後で行う統計検定（t 検定やχ二乗検定など）およびデータの視覚化（ヒストグラム，箱ひげ図などのグラフ）について実習を通して学び，Rの有用性を紹介したい。なお，対象は，Rを本ワークショップで初めて使うという方を想定しており，Rそのものや統計手法に関する事前知識は必要としない。

サポート用サイト：http://sakaue.info/wiki.cgi?page=JAECS2013

【研究発表第 1 室】

【研究発表 1】

Parallel English-Japanese corpus with meaning representations as educational aid

Alastair Butler (Tohoku University / PRESTO, Japan Science and Technology Agency)

This talk introduces an initiative to build a parallel corpus for English and Japanese with human checked syntactic annotations that support a dynamic assembly of predicate logic meaning representations. The corpus has sentence and short discourse alignment. Possible educational (foreign language learning) uses of the corpus are sketched.

The syntactic annotation follows a modified Penn Treebank scheme (Bies et al 1995, Santorini 2010) representing syntactic structure with labelled parentheses. Labels are either word-level part-of-speech tags (N, P, ADJ, etc.), or phrase level labels with a basic tag to indicate constituent form (NP, PP, ADJP, etc.), while additional tags (separated by a hyphen) indicate function (NP-SBJ=subject, NP-OB1=object, ADVP-TMP=temporal adverb, etc.). Examples of the annotation are given below:

```
(IP-MAT (NP-SBJ (PRO I))
        (VBD sat)
        (PP (P on)
            (NP (D a)
                (N chair)))
        (. .))

(IP-MAT (NP-SBJ *pro*)
        (PP (NP (N いす))
            (P に))
        (VB 座り)
        (AX まし)
        (AXD た)
        (PU 。))
```

Such annotations contain sufficient information to allow for automatic generation of the following meaning representations, where e_2 is an existentially bound sitting event, x_1 is an existentially bound chair, while z_3 creates a binding linked to the speaker:

$\exists z_3 x_1 e_2($
$z_3 = i \wedge$
$chair(x_1) \wedge$
$past(e_2) \wedge sat(e_2, z_3) \wedge on(e_2) = x_1)$

$\exists z_3 x_1 e_2($
$z_3 = pro \wedge$
いす$(x_1) \wedge$
$past(e_2) \wedge$ 座り-まし$(e_2, z_3) \wedge$ に$(e_2) = x_1)$

For the language learner seeing and exploring such meaning representations holds the promise of crystallising the grammatical contributions of language elements, with verbs, nouns, prepositions, determiners, particles, etc., having clearly identifiable consequences. Thus verbs (e.g., *sat* and 座る above) are predicates with an event/state argument. Nouns (*chair* and いす) are predicates without an event/state argument. English prepositions (*on*) and Japanese postpositions (に) may create conditions to further modify an event, or may extend the argument range of a noun contribution (e.g., *of* and の). Determiners manifest all or some parts of quantificational structure.

With syntactic annotation, analysis is firmly rooted by the language data. Yet when syntactic analysis is coupled with the generation of meaning representations, many apparent language differences are eliminated, while significant language differences are exposed (e.g., presence of embedding in English vs. use of nominalisation in Japanese). Most notable is that the level of meaning analysis offers deep insight into the contribution(s) of functional elements (prepositions, determiners, coordinating conjunctions, particles, etc.) in terms of how aspects of the meaning representations are linked together by binding and quantificational structure, propositional connectives, etc., encouraging an appreciation and understanding of these crucial elements of language.

【研究発表 2】

The design, development and research potential of The Kansai University Bilingual Essay Corpus

Miho Yamashita (Kansai University)

Corpus-based studies of written English texts of ESL/EFL learners have been conducted by many researchers so far. None of them, however, investigated the target L2 texts in close reference to comparable texts written in learners' L1. We believe that a mono-lingual corpus tells us only part of the story, because ESL/EFL learners are by definition bilinguals.

Against this background, we set to construct a large scale of bilingual corpus of essays written in both English (L2) and Japanese (L1) at Kansai University, Osaka. The corpus project started in 2011, involving about 200 undergraduate students. They wrote narrative and argumentative essays both in English and Japanese for 13 different topics during an academic year, totaling about 1,900 essays each for both English and Japanese versions.

The final picture of this project is expected to have a corpus of 1.3 million English words and 3 million Japanese words by the end of 2013. The corpus will be the largest of its kind in Japan. Because of its bilingual nature, the corpus is expected to provide us with the evidence to infer possible cognitive interactions between students' L1 and L2 while writing and thinking. We can see, for example, why their English essays are as they are in terms of lexis or logical developments. Are they influenced by their L1 or not? Are there any logical patterns typical of Japanese students? The Kansai University Bilingual Essay Corpus will give us new insights into English / Japanese writing by Japanese students that we have never known from existing monolingual corpus analyses.

This corpus also has other features, which include: (1) corpus design similarity with NICE or ICNALE, thus making direct comparison with them possible, (2) more than 60 kinds of writers' background information (age, sex, academic major, overseas experience, English proficiencies etc.) as well as their texts' attributes (type, token, TTR, number of words and sentences per essay, vocabulary levels and distributions as measured by JACET8000 categories etc.).

In this presentation, the major design features of the Kansai University Bilingual Essay Corpus will be first reported. This will be followed by a summary report of one of our studies, i.e. a study of the lexical and syntactic features of the English essays written by our subjects. With comparison against the reference corpora, FLOB, FROWN, TIME Corpus, NICE and ICNALE, we have revealed a variety of vocabulary usages characteristic of our subjects in terms of logical connectors and linking adverbs. In addition, it has been found that most of our students, particularly those studying at the Department of Foreign Languages, are accustomed to the conventional styles of academic essay which we confirmed was influenced by the English education they received. For instance,

they tried to avoid the use of "I" as the subject of the sentences, although "I" has been identified in the past literature as one of the most frequently used subject pronoun in Japanese EFL essays.

We firmly believe that the Kansai University Bilingual Essay Corpus will offer many more interesting research opportunities from the perspectives of both L1 and L2 as well as the interaction thereof. In this presentation, implications for future research and pedagogy in the writing classes will be also discussed. (546 words including the title and author's name)

【研究発表 3】

Contrastive lexical analysis of spoken and written interlanguage productions by Japanese learners of English: a study based on the newly added spoken module of the ICNALE

Shin'ichiro Ishikawa (Kobe University)

Many of the previous studies have suggested that learners use L2 vocabulary differently in their spoken and written productions. For instance, Fordyce (2009) compared the use of epistemic lexical units by Japanese college students in speech and essays and revealed that learners have a general tendency to use more stance adverbials (especially "maybe") and fewer epistemic modal verbs in speech, though they do not necessarily show the difference in use of the major lexical verbs ("think" is overused in both of the production modes). Nomura (2012) analyzed the use of articles (definite, indefinite, zero, demonstrative, and quantifier) by Japanese secondary school students in speech and essays and exemplified that they tend to use more articles in essays than in speech and the production mode does not directly influence the error rate in article uses. These studies have revealed several noteworthy facts about learners' L2 vocabulary use in speeches and essays, but the lexical differences in the two production modes still remain largely as "a rather unclear picture" (Fordyce, 2009). This may be partly due to the limited control in data collection procedure and/or the limited size of the data, especially spoken one, used for comparison: the amount of speech analyzed in Fordyce (2009) and Nomura (2012) is approximately 6,000 tokens and 10,000 tokens respectively.

The author is currently engaged in the project to expand the International Corpus Network of Asian Learners of English (ICNALE), which holds 1.3 million tokens of controlled essays written by learners in ten Asian countries and regions as well as English native speakers and is now one of the world's largest learner corpora publicly available (Ishikawa, 2013), by adding the learners' speech collected in the same data elicitation scheme. The new ICNALE will enable us to conduct a more reliable contrastive interlanguage analysis focusing on the difference between spoken and written productions by learners in Asia.

In the current paper, the author illustrates the aim, the procedure, and the protocol of data collection in the ICNALE-Spoken project and discusses how Japanese learners use vocabulary in general in their spoken and written productions and how task types, task conditions, and learners' proficiency levels influence their speech. Preliminary data analysis has shown that the lexical difference in the two production modes does exist, but it is smaller than suggested before especially when comparing learners at the same proficiency level. It has also been suggested that the amount of spoken production is greatly influenced both by learners' proficiency levels and by data collection scheme. These findings will shed a new light on the study of learners' L2 vocabulary use in different production modes.

【研究発表第 2 室】

【研究発表 1】

類義的な動詞不変化詞構文における不変化詞のアスペクト特性: *up* と *down* を中心に

大谷直輝（京都府立大学）

本発表では，反義語である *up* と *down* を含む類義的な動詞不変化詞構文のアスペクト特性を考察することで，*up* には AGENT 指向的な，*down* には PATIENT 指向的な特徴がある点を明らかにする。動詞が表す事態が完了したことを表す不変化詞のアスペクト特性は先行研究でも論じられているが（Bolinger 1971, Tenny 1994），個々の不変化詞のアスペクト特性の共通点と相違点に関しては詳しい調査がなされていない。

本発表では，3 種類の類義的な動詞不変化詞構文（*burn up*/*down*, *drink up*/*down*, *shoot up*/*down*）を *BNC* から網羅的に収集して，不変化詞のアスペクト特性が持つ指向性の違いを考察した。類義的な動詞不変化詞構文は，He shot {*up*/*down*} the lion, He burned his house {*up*/*down*}のように，各不変化詞を含む構文が，動詞が表す事態や行為の完遂を表すという点で類義的である。*BNC* には，*burn up* (145 例)，*burn down* (319 例)，*drink up* (74 例)，*drink down* (26 例)，*shoot up* (282 例)，*shoot down* (485 例)が存在した。本研究では，

これらの例に対して，(i) 語順 (SV/SVO/受身)，(ii) 目的語の品詞，(iii) 目的語の定性，(iv) 自動詞のタイプ，(v) 意味拡張の 5 つの変数をタグ付けして，文法的・意味的な特徴を調査した。

本発表の結果，*up* と *down* には以下の傾向が見られた。第 1 に，他動詞用法では，*up* は VPO 型（*shoot up the lion*）で用いられる頻度が非常に高いが，*down* は *up* に比べて VOP 型（*shoot the lion down*）で現れる頻度が高い。第 2 に，自動詞用法では，*up* の場合，非能格動詞（*drink*）や再帰的な用法（*burn*, *shoot*）で現れる（He was *burning up* with fever）。一方，*down* は非対格動詞で用いられる傾向が非常に強い（My house *burned down*）。第 3 に，*up* は受動態で用いられる頻度が低いが，*down* は頻繁に用いられる。第 4 に，意味の面では，*up* は垂直的な意味が漂白化して，意味拡張が進んでいるが，*down* は垂直的な意味を保持し，物理的・抽象的な下方向への移動の意味を表す傾向がある。

1-3 の文法面での結果は，*up* は AGENT 指向的であり，*down* は PATIENT 指向的である点を示唆する。すなわち，*up* は，他動詞の VPO 型や非能格や再帰的な自動詞で用いられる頻度が高いことから，動詞との結びつきが強く，AGENT が行う行為が完了したことを表すと考えられる。一方，*down* は，受動態や非対格動詞で用いられる頻度が高いことから，動詞が表す行為や動作によって PATIENT に生じる結果状態を表すと考えられる。また，本発表では，文法面の変数(1-3)と意味面の変数(4)の相互関係を考察することで，不変化詞の意味拡張が進むほど，PATIENT 指向から AGENT 指向になる可能性についても考えていく。

【研究発表 2】

Word, time and practice: a preliminary corpus-based comparative study on paired conversations of advanced-level learners and basic-level learners

Keiko Tsuchiya (Tokai University)

This presentation reports a small-scale study comparing a pair conversation of advanced-level learners of English with that of basic-level learners from lexical and pragmatic perspectives. The focus is placed on the establishment of research methods to describe learners' speaking production in order to improve assessment schemes for the paired oral tests. An analysis of ratings is excluded here since its emphasis is on empirical descriptions of learners' speech production rather than assessments of their performance.

Two five-minutes-long pair conversations, one of which is a pair of advanced learners and the other is basic level learners, were analysed and compared in terms of lexical ranges, turns and pauses, and discourse-pragmatic strategies for explanations using a time-aligned corpus-based approach. The learners were grouped into the levels by the results of their placement test and term tests in the previous year. Dalton-Puffer (2007) studies learners' strategies for explanations in Content and Language Integrated Learning (CLIL) classrooms and categorises four strategies of explanations they used: (1) elaboration (e.g. exemplification), (2) addition (e.g. A and B, or A but B), (3) variation (e.g. A or B), and (4) connection (e.g. cause/consequence, evidence/conclusion, problem/solution, and action/motivation). I take a lexico-pragmatic approach, referring to the Dalton-Puffer's categorisations for explanation strategies in a qualitative analysis in this study. These methodologies were developed based on the time-aligned corpus-based analysis and the integrated research methods of corpus-based quantitative approaches with qualitative discourse-pragmatic and conversation analytic approaches in recent years (Adolphs, 2008; Tsuchiya, forthcoming; Walsh, Morton, & O'Keeffe, 2011).

The main research aims are: (1) to compare lexical ranges of the advanced-level learner pair with that of the basic-level pair, (2) to compare lengths of speaker turns and pauses between the two pairs, and (3) to examine learners' pragmatic strategies for explanations the two pairs used.

The results from this preliminary study show that the advanced-level learners used more words in B1 and B2 level in the English Vocabulary Profile (EVP), which is based on the Common European Framework (CEFR), than the basic learners, most of whose vocabulary are classified as A1 and A2. In terms of turn-taking structure, fewer but longer turns were observed in the advanced pair compared with the basic level pair. The qualitative analysis indicates that several pragmatic and functional expressions for connection were used by the advanced learners to maintain a longer floor of conversation to provide semantically complex ideas in the conversation. This differs their conversation from the basic level pair. Some feasible methods to describe and analyse speaking production of learners' English were suggested in this study, which could be implemented in future research with a larger data set.

Acknowledgements: This research report has made use of the English Vocabulary Profile. This resource is based on extensive research using the Cambridge Learner Corpus and is part of the English Profile programme, which aims to provide evidence about language use that helps to produce better language teaching materials. See http://www.englishprofile.org for more information. This project was funded by Tokai University Research Start-up Support Grant (No.2012-09).

【研究発表 3】

The NICT JLE (Japanese Learner English) Corpus の XML 整形式化とそれを使った習得段階別の語用論的言語特徴の分析

三浦愛香（東京経済大学）・佐野洋（東京外国語大学）

本発表では，The NICT JLE Corpus の XML 整形式化の手順を進め，その結果を基に語用論的能力である「要求」のスピーチアクトを示す言語特徴や言語使用を習得段階別に分析した結果を報告する。中間語用論では，外国語学習者による要求や謝罪などのスピーチアクトを談話完成タスク（DCT）で収集する手法が主流であり，Salgado (2011) は，要求のスピーチアクトの習得を観察し，母語話者に比べ学習者は直接的であり，熟達度が上がると間接的な表現が増えるとしている。しかし，Adolphs (2008) によると，コーパスを活用した外国語習得の研究では語彙やコンコーダンスラインに比重を置く傾向が強く，話し言葉コーパスに基づいた語用論的機能の研究はまだ数が少ない。

The NICT JLE Corpus は，イラスト描写やロールプレイなど異なるステージやタスクから構成されるインタビューテストの書き起しデータであり，試験官と受験者のやり取りから成り，全受験者は 9 段階の熟達度レベルに分けられている（和泉，内本，井佐原 2004）。本研究が目的とする語用論的な言語分析には，タスクごとのデータの分別や発話者同士の対話の文脈情報の参照が必須となるという背景から，当該コーパスの整形式化を行った。NICT が Web 上で提供するテキストデータは，超言語的な要素（「フィラー(F)」，「繰り返し (R / R?)」，「ポーズ(.)」，発話者同士の「オーバーラップ (OL)」など）のマークアップが厳密には XML 整形式ではない。本研究では，上記のタグ情報に対して，Perl スクリプトでの一括変換と目視確認による手作業の修正を施して整形式化したことで，多様なデータ処理が可能になった。専用の分析ツール(和泉他 2004) や AntConc などのコンコーダンサーでは不可能である試験官と受験者の発話の分別や特定のタスクや熟達度によるデータの選別が実現したことで，語用論的な言語分析における利便性が向上した。

語用論的言語特徴を抽出するにあたり，ロールプレイのタスクに限定し，Blum-Kulka, House & Kasper (1989)や Salgado (2011) の coding scheme を参考に，要求のスピーチアクトの言語使用を文脈から見出し，マニュアルでタグ付与をした。要求の中枢部を Head Act (HA)，Head Act を補助する部分を Supportive Move (SM)と特定し，Head Act の視点が話者か聞き手にあるか，また，表現の直接的な度合いを Direct (DR), Conventionally Indirect (CI), Indirect (ID)の 3 つに分類する。また”please”など politeness を示す語彙的な marker(LD mkr=”polite”)もタグ付与する。本発表では，語用論的なアノテーションを付与することで，タスクや習得段階別に要求のスピーチアクトの言語特徴の計量的な調査結果を報告する。

【シンポジウム】

コーパスが語ること，コーパスが語らないこと

司会　金澤俊吾（高知県立大学）

コーパスを使うことには，一度に多数のデータを扱うことができる，様々な言語現象を扱うことができるなど多くの利点がある。一方で，コーパスは，言語事実を観察するための道具に過ぎず，言語研究のためには，さらなる道具立てが必要とされるという見方もある。

近年，理論言語学の様々な分野で，コーパスが盛んに使われている。そして，新たな言語事実を発掘する取り組みがなされ，同時に，コーパスの使用が理論言語学の進展に何をもたらすのか，コーパス使用の有効性を探る研究が進められている。

コーパスが言語研究の様々な場面で使われている現在，コーパスを使うことの意義を改めて問う必要がある。コーパス資料から，何を読み取ることができるのだろうか。また，コーパスを使用する上で，どのような情報を補い，言語を分析する必要があるのだろうか。

本シンポジウムでは，コーパスの使用を通じて得られる結果に基づき，新たな言語分析の可能性を探ることを目的とする。とりわけ，頻度効果という視点からは扱われることのない言語事実，いわば英文法の周辺部に位置する言語事実に注目し，先行研究における言語表現に関する一般化に対して，新たな言語分析の可能性を提示することを試みる。

具体的には，3 名の発表者が，それぞれ，英語における前置詞や形容詞，動詞が関わる諸現象を取り上げて，語にみられる多義性や，諸構文にみられる意味的特徴について検証する。そして，本シンポジウムで扱う各言語現象には，いずれも，中核的意味から周辺的意味に至るまで，一定の原理に従って連続的に関連づけられることを明らかにする。

経路を表す前置詞の意味について

講師　浅川照夫（東北大学）

英語の経路前置詞 *across*, *along*, *around* の意味は多義であり，母語話者が直感的に基本的意味と感じているプロトタイプを核として複雑な意味のネットワークを形成している。例えば, *across* の場合, He swam *across* the river と He traveled *across* America から分かるように，人物が移動する経路の動線が線形でもよいし，方々に無方向に広がっていてもよい。また，Talmy (2000) によると，(a) *across* the river，(b) *across* the square field，(c) ?*across* the swimming pool，(d) **across* the pier のように，目的語の矩形の縦線と横線の比率によって容認可能性に差があるというが，A big snake is crawling/lying *across* the bridge に見られるように，目的語をどの方向に横断するかは決して一律に固定されているものではない。本発表では，経路前置詞の解釈原理に鳥瞰的視点と移動者視点の概念を導入することによって，それぞれの意味がプロトタイプから特殊へと一定の原理に従って連続的に関連付けられていることを説明してみたい。

周辺的な構文に見られる変種に関するコーパス資料とその解釈をめぐって

講師　大室剛志（名古屋大学）

文法は対等の資格からなる構文の単なる寄せ集めではなく，中核的な構文から周辺的な構文へと連続的に多重的に構成されている。また，１つの構文も対等の資格からなるメンバーの単なる寄せ集めではなく，構文を構成するメンバーには，その構文の基本形もあれば変種もある。このような連続的多重的な文法観に立ち，本発表では，英文法の周辺部に属すると思われる構文の変種を取り上げ，それらの変種に関する言語資料として BNC，BoE，Web 等から得られる資料を提示し，コーパス資料からそれらの変種の属性についてどこまで見えてくるのか，またコーパス資料が得られた時にその資料をどのように解釈することにより，それらの変種の理論的な分析につなげていくのかを見ることにする。具体的には，同族目的語構文で独立関係節を伴う変種，同族目的語構文と動作表現構文の受け身，*d'rather* が直接節を従える構文で補文標識の *that* を欠いた変種，同構文で補文の主語位置に対格の名詞を取る変種，等を取り上げる。

形容詞の限定用法にみられる修飾関係の多様性について

講師　金澤俊吾（高知県立大学）

英語において，Adj-N から構成される名詞句には，多様な修飾関係がみられる。例えば，*a quick drink* において，*quick* は，「飲む」事象を修飾する。また，*an ice-cold drink* において，*ice-cold* は，飲み物の冷たさを表す。この修飾関係の違いは，共起する動詞の分布の違いにも反映される。例えば，*a quick drink* は，軽動詞 *have* と共起するのに対し，*an ice-cold drink* は，*drink* と同義である主動詞 *have* と共起する。本発表では，BNC, BoE, Web 等のコーパス資料を用いて，英語における NP-*have-a /an*-Adj-N にみられる修飾関係について検証する。その中で，NP-*have-a /an*-Adj-N を 1 つの構文として捉え，この統語的連鎖から成る様々な用法は, 中核的な用法から周辺的な用法へと連続性を成すことを提案する。また, Dixon (1991: 342, 2005: 465)が混交によって形成されると指摘する *Have a quick whiskey.*は，この連続性の中で捉えられる用法であることを明らかにする。さらに，「飲む」を表す主動詞 *have* と，動詞 *drink* は同義であると指摘されてきた事実に対して，実際の使い方には違いがみられることを明らかにする。

■第 2 日目

【ワークショップ 2】

統計解析環境 R による言語データの分析（中級編）

講師　阪上辰也（広島大学）

本ワークショップでは，Rの基本的な機能を用いて，言語データの分析方法を学ぶ。すでに，AntConcやCasualConcなどのコンコーダンサーが利用でき，手軽にコーパスを分析できるが，どのような処理過程を経てその結果が得られるのかについて理解することは難しい。研究手法の適切さという観点からは，言語データの分析に関わる処理過程をブラックボックス化させたままにするよりも，可能な限り，その過程についての理解を深めておくほうがよい。

具体的な実習内容として，日本人英語学習者コーパスの「NICE」（Nagoya Interlanguage Corpus of English）を利用し，次の6つの処理を実習形式で学んでいく。1) データの読み込み，2) データの抽出，3) データの分解，4) データの整理，5) 数値データ（総語数など）の算出，6) データの書き出し，という手順を踏むことで，言語データ分析に関わる主たる処理を理解し，Rを介してその処理方法を学ぶことで，言語処理技術の基礎を押さえることができる。時間が許せば，日本語データの処理についても実演する予定である。

サポート用サイト：http://sakaue.info/wiki.cgi?page=JAECS2013

【講演】

計量的文献研究とコーパス

講師　村上征勝（同志社大学）

計量的文献研究におけるコーパスの利用の現状と課題，特に日本語コーパスの課題について紹介する。

文献の著者の推定，真贋判定，成立時期の推定などを行う計量的文献研究では，文章の数量的性質に注目し，単語の出現率を中心に書き手の文体の特徴を探り，問題解決の糸口を探ることが試みられてきた。計量的文献研究の嚆矢ともいうべき前世紀初頭のメンデンフォールによる「シェイクスピア＝ベーコン論争」を扱った研究では，60 万語の単語の長さが分析された。しかし，単語の出現率を用いた分析は行われておらず，出現率を用いた分析は，コンピュータと電子化されたコーパスの登場を待たなければならなかった。欧米における，『静かなドン』，『聖書』，『連邦主義者（*The Federalist*）』等の文献の分析では大なり小なりコーパスが用いられていたはずである。単語に品詞情報等を付加したコーパスを用いれば，文献の数量的性質の詳細な分析が可能となるため，著者推定，真贋判定等の具体的問題の解明のみならず，文章に関わる研究は飛躍的に進むと考えられる。

ところで日本語の場合はどうか。1975 年頃まで日本語を扱えるコンピュータはほとんどなかったこともあり，日本語文献の計量的分析は遅れた。しかし，現代日本語の形態素解析ソフトが開発されたことで，今日，コーパスを用いた計量的文献研究は盛んになりつつあるが，英語などと違い日本語の文献の場合には，文章が分かち書きされていないという大きな問題がある。そのためコーパスの作成に形態素解析ソフトを用いても，ソフトが異なると，形態素解析の結果も異なるという状況が見られる。加えて次の様な問題もある。たとえば，同一単語が，漢字，仮名と表記が異なって出現した場合に，そのままでは同一単語と認識されないし，逆に同音異語，たとえば，「こと（事）」，「こと（琴）」の区別が出来ない。更に古文ではもともと句読点がないため，文頭，文末の単語の分析に問題が生じる。

講演では，諸外国の代表的な計量的文献研究と，筆者がこれまで試みてきた「日蓮遺文の真贋判定」，『源氏物語』，「西鶴作品」のコーパスを用いた計量分析について紹介する。

【研究発表第 1 室】
【研究発表 1】

日本語母語話者の英語研究論文におけるヘッジ使用

若松弘子（筑波大学大学院生）

論文に用いられる英語は無味乾燥に事実のみを伝えてはいるわけではないという指摘は度々なされており，書き手の主張をより説得力あるものにしている証左として，Hyland (1998) は論文におけるヘッジの分析を行っている。ヘッジはLakoff (1973) を契機に言語学分野に導入され，例えば，*kind of* や*I think* のように，伝達する意味内容を間接的もしくは曖昧にする機能を持つ。論文に用いられるヘッジには *would*，*may*，*indicate*，*suggest*，*assume*，*seem*，*likely*，*possible*などがある。とはいえ，論文中のヘッジ使用については，論文は主観を排除し客観的であるべきで，ヘッジの使用は慎むべきという見方もあり，ヘッジの有用性を説く立場と対立する。この二つの異なる見解の存在は，論文においてヘッジを適材適所でバランスよく使用することの難しさをも示唆しているようだ。ESL学習者等を対象にした先行研究では，中国語母語話者，日本語母語話者，ドイツ語母語話者，チェコ語母語話者等による英文におけるヘッジの頻度，種類，特徴などが，対照とする英語母語話者（NSs）による英文のヘッジと有意に異なる場合があることを示しているが（Clyne, 1991; Hinkel, 2000; Hyland & Milton, 1997; Hyland, 1998b; Kranich, 2011; Yang, 2013; 小林, 2009），その差異は，学習者の熟達度が上がると消失するものなのか，高熟達度のESL学習者の英文においても残留しがちなものなのか，はっきりしない。本発表では，NSsによる英語論文と高熟達度の日本語母語話者による英語論文にみられるヘッジ使用の差異について明らかにする。なお，後者には日本人研究者が翻訳・校閲サービス等を利用して作成した論文も含まれると想定される。

分析に際しては，Thomson ISI databasesを利用し，過去10 年間の被引用数で上位にある注目度の高い論文のうち，化学分野において，日本語母語話者のみが著者である論文50本とNSsによる論文100本（米国50本，英国50本）を抽出し，コーパスを作成した。主要な語彙的ヘッジについてAntConc を用いて頻度を算出し，カイ二条検定を行ったところ，概してNSsによる化学論文においてより多くのヘッジがみられる傾向があった。法助動詞*must*, *may*, *could*, *would*, *can*や形容詞*possible*はNSsの論文よりも日本人研究者による論文に有意に少なかった（$p < .001$）。この結果は，例えば，日本人大学生による英文により多くのヘッジが散見されるというHinkel (2000) の指摘と矛盾する。一方，ESL学習者は論文の英語を直接的・断定的であるべきという誤った先入観を持つことが多いというHyland (2008)の指摘を支持する。

【研究発表 2】

教育利用可能なパラレルコーパス検索プラットフォームの構築に向けて

中條清美（日本大学）・アントニ・ローレンス（早稲田大学）・赤瀬川史朗（Lago 言語研究所代表）・
西垣知佳子（千葉大学）・水本　篤（関西大学）・内山将夫（情報通信研究機構））

Data-Driven Learning (DDL)では，英語学習者が検索ツールを使ってコーパスからターゲット語を検索し，豊富な言語使用例を見て，ことばの規則を発見して学ぶ。本研究グループは，中学 1 年生から大学生・院生という幅広い学習者を対象として DDL の指導実践を行い，DDL が語彙・文法学習に有効な手法であることを確認した。しかし，DDL は有効な手法でありながら，多くの日本人学習者が属する初級・中級レベルの指導では普及は遅れている。その理由には，DDL では検索用コーパスとして‘authentic’な言語データを使用するため，検索結果の英文が日本人学習者には難しすぎるということがある。さらに DDL 実践者から寄せられたフィードバックを見ると，①日本人の英語力にあった「簡潔で自然な」英文例，②簡便な検索ツール，③効果の高い DDL 教材に対する要望が高い。

現在，本研究グループは，上記の要望に応え，教育現場への DDL の普及に向けて，教育用日英パラレルコーパスと多彩な DDL 活動が可能な 4 種の検索ツールを搭載した DDL プラットフォームを構築中である。搭載する英語・日本語パラレルコーパスは，学習者の英語力に近く，かつ自然な例文を集めた著作権フリーの例文コーパスで，3,000 万語の易しい英語コーパスから抽出した例を参照して英文をおこし，それに日本語対訳を付けている。例文は「ケンブリッジ英文法」等の文法項目に対応している。DDL プラットフォームに搭載する 4 種の検索ツールのうち，1 種類は Web 検索が可能な，*WebParaNews*，もう 1 種類は自作の DIY コーパス等を検索可能な多言語検索ツール *AntPConc* である。両者は試用実践を経て無料公開した。語彙・文法項目の中には，これら 2 種のキーワード検索型ツールでは検索しづらいものがあるため，2 種類のプロファイリング例文表示型検索ツール，*GPPS (Grammatical Pattern Profiling System)*と *LWP　(LagoWordProfiler)*を開発中である。*GPPS* は見たい文法項目に合致した例文を難易度別に

表示し，*LWP* は見出し語別に品詞表示およびその例文を出力する。この DDL プラットフォームは 4 年間でオープン化を目指しており，今後の方向性として，これらのコーパスとツールを用いた DDL 教材の作成，効果検証，改善についても言及する。

【研究発表 3】

JEFLL Corpus Parallel Version の構築と活用

投野由紀夫（東京外国語大学）

JEFLL Corpus は日本人中高生 1 万人の英作文を収集した学習者コーパスである（投野 2007）。学習者コーパスの分析で最も特徴的なのは学習者の文法・語彙などに関する誤りの分析ができることにある。しかしながら，エラータグ付与は人手を介して行うのが通例で非常に時間と労力がかかり，たいていの学習者コーパスは部分的にエラータグ付きデータを作る（例：NICT JLE Corpus），あるいは研究者の目的に応じたエラータグ付与（problem-oriented error tagging）に終始する場合が多い。

本研究では，このエラータグ付与の問題を克服する 1 つの方法として，個々の英作文に対して添削バージョンを作り，その添削版と原文のパラレル・コーパスを構築する試みを報告する。さらに，オリジナルの作文と添削版の差分を自動でエラータグ付けする手法も紹介し，このようなデータの整備と利用が英語教育の改善にどのように資するかも論じたい。

JEFLL Corpus の各作文に対して，1 名の添削者（イギリス英語母語話者，日本での教授経験 10 年程度，日本人女性と結婚）に 1 万件の添削を依頼した。添削は local error を中心に文レベルを超えない範囲で，可能な限り原文の意味が伝わる最低限の誤り訂正を行うように指示し，エッセイ全体の理論構成などは考慮に入れなかった。添削文はすべて原文と対応付けする形でデータ整備が行われた。原文と添削文の対応付けデータを比較することで，その差分は何らかの修正を添削でほどこされた部分であるとわかる。この修正部分に適切にタグ付与を行えれば，エラータグを自動で付与できる可能性がある。原文と添削文の対比で差分として抽出できる可能性は以下の 3 つのパターンになる：

① 原文にあったものが削除された → 余剰エラー （addition error）
② 原文にないものが付加された → 脱落エラー （omission error）
③ 原文の形が変更された → 誤形成エラー（misformation error）

これを自動で抽出するために，編集距離（edit distance）と呼ばれるアルゴリズムを用い，Ruby でアノテーションを行うスクリプトを作成した。エラータグの自動付与の精度評価に関しても当日報告する。パラレルコーパスの作成は，対応付けデータを用意すれば，市販の ParaConc, MultiConcord, Sketch Engine, または最近公開された AntPConc などへの実装が可能である。現在，エラータグ付与をしたバージョンを整備中で，来年に向けて公開する予定である。

原文と添削文のパラレルデータの比較はエラータグが付与されていないデータだけでもさまざまな学習者の中間言語の様態を観察できる。実例を挙げながら，文法指導，教材開発，DDL などへの応用を論じる。

【研究発表第 2 室】
【研究発表 1】

心理言語学的語彙特性の語彙の豊かさ指標への応用可能性の検討

石井卓巳（筑波大学大学院生）

産出語彙に関する研究では，語彙知識の広さや深さに加えて，語彙の豊かさの測定が行われてきた。語彙の豊かさの指標は，語彙の多様性を示す異なり語の割合に基づく指標（TTR，Guiraud index，D 等），及び語彙の洗練性を示す語彙の一般的頻度に基づく指標（LFP，P_Lex，V_Size 等）に大別できる。

但し，どちらの指標も，語彙の処理，習得や学習，保持に影響を及ぼす語の質的側面をほぼ考慮していないため，低頻度な異なり語の産出のみで語彙の豊かさの向上や産出語彙の発達が測定されてしまう問題点が指摘されている。そこで，多面的な語彙の豊かさの測定を実現するため，近年の研究 (Salsbury, Crossley, & McNamara, 2011; 草薙, 2013)では，語の質的側面を構成する心理言語学的語彙特性を用いた測定が試みられている。しかしながら，研究や分析対象が限られている事に加え，これらの研究では既存の諸指標や他の語彙知識の側面と心理言語学的語彙特性との関係も不明瞭である。

従って，本研究では，(a) 産出語彙の発達段階の弁別と(b) 他の語彙の豊かさの指標，語彙サイズ，熟

達度との関係の観点より，MRC Psycholinguistic Database (Wilson, 1988)と WordNet (Princeton University, 2010)に基づく 8 種類の心理言語学的語彙特性の語彙の豊かさ指標への応用可能性を検討する。アジア人英語学習者コーパス ICNALE に採録されている英語母語話者（$N = 72$），及び TOEIC と Vocabulary Size Test (Nation & Beglar, 2007)の点数に基づく日本人 EFL 学習者 4 群（各群 $N = 18$）の 2 テーマ計 288 の課題英作文を対象とした分析の結果，以下 4 点が明らかになった。

1. 日本人 EFL 学習者と英語母語話者は 5 種類の語彙特性で有意（$p < .001$）に弁別された。
2. 日本人 EFL 学習者 4 群は全語彙特性で弁別されなかった。
3. 熟達度・語彙サイズとの相関は概ね非常に弱かった。
4. 既存の語彙の豊かさの諸指標との相関は弱～中程度であった。

即ち，英語母語話者と日本人 EFL 学習者は，複数の心理言語学的語彙特性により弁別された一方，日本人 EFL 学習者 4 群は全く弁別されない結果となった。また，心理言語学的語彙特性は語彙サイズや熟達度との関係は弱いものの，既存の語彙の豊かさ指標を包括する，或いは異なる側面を測定すると考えられる。発表では，関連する先行研究を踏まえて本結果を議論した上で，今後の研究方針や語彙の豊かさ増強のための指導方針を報告する。

【研究発表 2】

学習モデルとしての教科書用例の改善：使役動詞の記述例に基づく研究

井上　聡（環太平洋大学）

筆者は過去に複雑な補部構造を導く動詞の用法について研究を行った（井上，2010, 2011, 2012）。高校生に対して動詞の難易度についての調査を行ったところ，使役動詞に対する苦手意識が強いことが示された。COCA に基づいて使役動詞の用法を調査したところ，「*make / let*＋O＋原形不定詞」の典型性が高いことが明らかとなったが，学習者コーパスに基づく分析の結果，習熟度の低い学習者ほど，*make* を過剰使用し，*let* を過少使用する傾向が示された。

本研究では，教科書が学習者の運用能力に及ぼす影響に着目し，使役動詞の典型性がどの程度まで教科書に反映されているのかを調査するため，量的分析を行った。使用したデータは，2013 年に新たに採択された検定済み高校英語教科書（英語表現 I の 11 種と英語コミュニケーション I の 22 種）である。RQ1 では英語表現 I と COCA の関係について，RQ2 では英語コミュニケーション I と COCA の関係について調査を行い，RQ3 では 2 種の新教科書と COCA の関係について分析を行った。

まず，英語表現 I を見ると，母語話者による典型構文（*make / let*＋O＋原形不定詞）が過少使用されているのに対して（$p < .001$），比較的一般性の低い構文（*make / have*＋O＋過去分詞）が過剰使用されていた（$p < .001$）。また，過剰使用構文の大半が ‘*make oneself understood*’ や ‘*have*＋O＋*stolen*’ のような成句的・定型的なもので占められていた。次に，英語コミュニケーション I では，「*make / let*＋O＋原形不定詞」を中心として，使役動詞構文が全体的に過剰に使用されていた（$p < .01$）。最後に，多変量解析を援用したところ，英語コミュニケーション I における出現頻度が COCA に近接しているのに対して，英語表現 I においては，やや逸脱的な使用傾向が示された。

教科書の用例を選定する上で，コーパスから抽出される傾向を 100%反映させる必要はない。ただし，高頻度で使用される表現への理解を促進するためには，母語話者の典型性から過度にかけ離れた状況は改善されるべきである。本研究の結果，使役動詞構文について，まとまった談話構造の中では自然なインプットが豊富に含まれるものの，個別の用例には改善すべき課題が見出された。今後，英語表現 I の用例を選定する際には，コーパス知見を組み込むことによって，より適切な学習モデルを提示することが可能となるであろう。

【研究発表 3】

司法英語における類義語をどう活用発信型辞書に記述すべきか

鳥飼慎一郎（立教大学）・溜箭将之（立教大学）

アメリカやイギリスで司法英語を運用しようとする際に直面する大きな問題の一つが，専門用語の使い方である。専門用語の語義は専門用語辞典を引けば理解できるが，それらの専門用語がリーガル・ディスコースの中でどのように使用されているのかは分からないからである。この司法英語運用上の問題は，これまでは個々人の努力と試行錯誤の中で自らが克服すべきものとされてきたが，このプロセスは多くの時間とエネルギーを必要とする割に非効率的であり，そうして習得した語彙・文法表現が他の類

似表現とどのように異なるのかなどは個人で客観的な判断を下すことは難しかった。本発表者は，このような日本人学習者が抱える司法英語習得上の問題点を，活用発信型の司法英語辞典を編集することで少しでも軽減しようとするものである。その基礎研究のために，イギリス最高裁判例（1,451,263 語），イギリスロージャーナル（1,267,048 語），アメリカ連邦最高裁判例（1,574,403 語），アメリカロージャーナル（1,303,223 語）の 4 本の司法英語コーパスを構築した。

本研究発表では，*judgment*, *decision*, *decree*, *verdict*, *ruling*, *order* といった，司法当局が発する拘束力のある決定を意味する専門用語を例に取り，司法英語のディスコースの中でどのような使われ方をしているのかを例にとって，どう活用発信型の司法英語の辞書として記述してゆくべきかを論じてゆく。

judgment を例にとれば，この語は一般的には「判決」と訳されるが，『英米法辞典』（1991）では，「当事者間の権利義務について判断し紛争を解決する裁判所の最終判断。*judgment* と *decision* は，同じ意味で相互交換的に用いられることが多い。元来はコモン・ロー上の事件における判決のことで，エクイティ上の事件，海事事件，離婚事件における判決は *decree* とよばれたが，最近は，法律上または事実上，後者においても *judgment* という表現が用いられることが多い。」と定義され，解説が加えられている。*Black's Law Dictionary*（2009）もほぼ同様の記述である。これではこれらの語をどのように使ってよいのかわからない。

実際のアメリカ連邦最高裁の判例では，これらの専門用語を目的語として取る動詞は，*judgment* (*reverse* 46, *affirm* 44, *vacate* 25, *enter* 22, *grant* 22), *decision* (*make* 69, *review* 16, *affirm* 13, *issue* 13, *reach* 13), *order* (*enforce* 13, *enter* 13, *issue* 13, *seek* 9, *appeal* 6, *determine* 5), *verdict* (*reach* 17, *return* 8, *challenge* 5, *ground* 3), *ruling* (*make* 3, *appeal* 2, *compel* 2, *uphold* 2)となっており，各語の使用頻度，共起する動詞ともに大きく異なっているのである。

本研究発表では，上記 5 語の司法英語における本質的な意味，各語の句レベルでのコロケーション情報，談話レベルでの使用実態，判例の分野別，地域，ジャンル別における差異，などの要素をできる限り明らかにし，日本人学習者がこれらの専門用語をリーガル・ディスコースの中で使えるようになるための一助となる活用発信型の辞書の一例を示してゆきたい。

【研究発表第 3 室】
【研究発表 1】

基本句を考慮した *n*-gram の計数

田中省作（立命館大学）

コーパス研究において *n*-gram は，語を計数単位とした「語の *n* 連鎖」として考えられることが多く，頻繁に採用される言語素性の一つである。一方で，自由項を挟むような不連続な関係を捉えることが難しい，適切な *n* が明確ではない，といった問題がある。そこで，ギャップを許すような *n*-gram (Guthrie *et al*., 2006; 國吉・中西, 1997)や，定型表現の自動抽出を指向した松原他(2010)のような方法も提案されている。本発表は，コーパス研究における新しい分析ツールの提案として，松原他(2010)による基本句（句構造を内含しない句）を考慮した *n*-gram 計数の基本アイディアを紹介し，そのプログラムと利用法を報告する。なお，発表者は現在，松原他(2010)の方法も活用し，特殊データベースから言語知識の獲得を試みており，本プログラムはその一環で実装，公開している。

松原他(2010)の計数対象は，基本句に関する情報が付与された英文である。たとえば，“I take the book to the library”の基本句情報は，TreeTagger (Schmid, 1994)で付与すると，“[$_{\text{NC}}$ *I*] [$_{\text{VC}}$ *take*] [$_{\text{NC}}$ *the book*] *to* [$_{\text{NC}}$ *the library*]”となる。ただし，“[$_{X\text{C}}$ α]”はαが基本 *X* 句，NC, VC はそれぞれ基本名詞句・基本動詞句を表す。本プログラムでは，*n*-gram 生成時に基本句を跨ぐ場合，それら基本 *X* 句の語列αを〈*X*C〉という 1 計数単位に置換した列も組み合わせ的に考える。その結果，上記英文の，たとえば 3-gram には，“〈NC〉 *take the*”や“*I* 〈VC〉 *the*”，“〈NC〉 *take* 〈NC〉”なども含まれ，*n* を適当に動かしながら累積的に計数していくなかで，“*take* 〈NC〉 *to* 〈NC〉”といった自由項を含むような定型的な表現の一端も捉えることができる。さらに，各 *n*-gram には，松原他(2010)で提案されている頻度，表現の大きさ，左端・右端の予測可能性等を合算したスコアリングの他に，基本句を跨ぐ際の列の組み合わせ数を勘案したスコアリングもされ，適宜，フィルタリング等に活用できる。なお，本プログラムは本質的には基本句情報に対する計数上の記号操作であり，計数単位にどのような言語情報を反映させるべきか，そして言語データに対する具体的な前処理については，研究目的に応じて別途決定されるものである。

【研究発表 2】

品詞／機能語列の分布に基づいた非母語話者英語に潜在する言語系統樹の構築

永田　亮（甲南大学）

コーパス分析に基づく母語干渉の研究が盛んに行われている。従来研究で中心となるのは，誤り分析と過剰／過小使用に関する分析である。例えば，Aarts & Granger (1998)は，χ^2値を用いて，非母語話者の英文中で過剰使用される品詞列を特定している。

本発表では，母語干渉に起因して，非母語話者の英文に母語の親縁関係が保持されることを示す。具体的には，様々な非母語話者コーパスを自動分類すると，母語の親縁関係に対応した言語系統樹が得られることを示す。このことは，親縁関係を弁別する因子が，言語系統樹を構築できるほどに強く英文に転移することを意味する。

言語系統樹の構築には，確率的言語モデルに基づく手法（永田 & Whitakker，2012）を用いる。この手法では，まず，品詞解析を用いて各コーパス中の内容語を品詞に置き換える。次に，品詞／機能語の連鎖確率により各国語話者の英文をモデル化する。すなわち，品詞／機能語の使用傾向を確率として母語ごとにモデル化する。このモデルを用いると，各国語話者が各コーパを生成する確率を計算できる。更に，この生成確率の比によりモデル間の距離を計算する（連鎖確率が似ていれば，生成確率も似た値となり，結果的に比（距離）は小さくなる）。最後に，計算した距離を用いて階層型クラスタリングを行うことで言語系統樹を得る。

この手法を用いて，印欧語話者の英文（ICLE と LOCNESS）から構築した言語系統樹を図 1 に示す。図から，得られた言語系統樹が印欧語の言語系統樹に極めて類似することがわかる。なかでも，12 種類

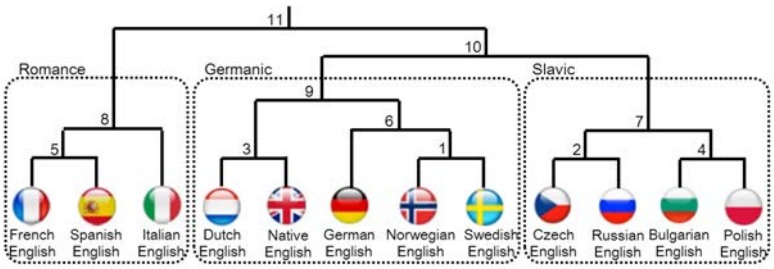


図1：印欧語話者の英文から自動構築された言語系統樹

の英文がロマンス語派，ゲルマン語派，スラブ語派に正しく分類されていることは特筆すべきである。更に，アジア圏の英語（ICNALE）においても同様の現象が見られることを示す。Inner / outer / expanding という 3 circle 11 種類の英文を収録した同コーパスの場合，母語の親縁関係と共に circle の関係性も保持される。また，英文のレベルとは独立して，親縁関係の保持が起こることについても議論する。

発表の後半では，親縁関係が保持される要因を考察する。まず，特徴的なフレーズ（クラスタリングに与える影響が大きいフレーズ）を抽出した結果について報告する。その後，抽出結果を基に，少なくとも，(a) 複合名詞の構成方法，(b) 冠詞の使用，(c) 副詞の位置が親縁関係の保持に寄与していることを明らかにする。例えば，複合名詞の構成方法が母語における複合名詞の構成方法と対応することを示す。

【研究発表 3】

Personalised statistical writing analysis

John Blake (Japan Advanced Institute for Science & Technology)

English is the *de facto* language for scientific journals with the highest impact factors. This means that non-native English speaking (NNES) researchers have to not only master their specialism but also English, which for speakers of dissimilar languages, such as Japanese, is a significant hurdle. The aim of this pedagogically-driven project is provide objective statistical evidence that NNES researchers can harness to make decisions on how to improve the generic integrity and lexicogrammatical accuracy of their draft manuscripts.

Selected features of each draft research article (approx. 2,500–10,000 words) are compared with a specially-created corpus of the target publication (approx. 250,000 words) and a tagged corpus of the appropriate

subject domain, namely information science, materials science or knowledge science (circa. 3 million words each). The subject domain corpora were annotated with parts of speech (POS) using GoTagger version 0.7, drawing upon rules from the Brill POS tagger. The generic integrity and lexical repertoire of drafts were then analyzed with respect to the target corpus and a subject-specific corpus using various textual analysis tools to generate data on vocabulary fit, readability, lexical profile, marked usage and grammatical accuracy.

The concept of keyness was used to identify the vocabulary fit with the respective target publication. The most frequently used unigrams, bigrams, trigrams, 4-grams and 5-grams were identified. Readability statistics, such as mean sentence length, lexical density, Gunning Fog index and Flesch reading ease, were also calculated and compared to the target corpus. The lexical profile was created using Tom Cobb's online vocab profiler (Web VP Classic v.4) to identify words listed in the general service and academic word lists. In addition, marked usage and lexicogrammatical errors were identified manually in the submitted draft. The keyword in context function was used to derive the statistical probability of the occurrence of marked forms. This was compared to an appropriate reference corpus, such as the untagged 100-million word British National Corpus or the tagged subject domain corpora. Grammatically incorrect usage was identified, corrected and commented on, and where relevant appropriate reference material was suggested.

Each researcher was presented with a personalized academic writing analysis showing a summary of the relevant statistics, followed by a more detailed analysis accompanied with an explanatory guide to help them interpret the statistics. Follow-up interviews were held with participating NNES researchers to identify the efficacy of the personalised statistical writing analysis. The future development and direction of this project will be discussed in light of this feedback.

《大会参加者へのご案内》

- 大会・ワークショップの受付：会場のマルチメディア教育研究棟 1 階エントランスホールで，1 日目は午前 9 時 30 分から，2 日目は午前 9 時 10 分から受付を行います。
- 東北大学川内キャンパスでの食事については当資料 21 ページをご参照ください。
- 構内での喫煙はできません。構内禁煙にご協力いただきますようお願いいたします。
- 当日会員について：会員でない方も，「当日会員」としてご参加いただけます（会費 2,000 円，二日間共通）。懇親会へもぜひご参加下さい。
- 懇親会（参加費￥5,000）では，今大会にてご講演いただく村上征勝先生（同志社大学）を囲んで和やかな雰囲気の中で，参加者同士普段はなかなかできない情報交換も可能になるのではないかと存じます。大会ご出席の方々には，ぜひ奮ってご参加いただけましたら幸いです。東北のお料理をはじめ，ソフトドリンクやビールに加え，美味しい地酒も用意していただいているとのことです。どうぞお楽しみに。なお，会場準備の都合で，ご参加予定の方には事前の予約をお願いしております。ご協力のほどよろしくお願い申し上げます。

    - 英語コーパス学会第 39 回大会・懇親会
    - 日時：10 月 5 日（土）18:30–20:30
    - 場所：東北大学川内北キャンパス・Bee ARENA Café
    - 会費：5,000 円

  ※懇親会参加予約申し込みは，**9 月 18 日（水）までに**
  1) ご氏名，
  2) ご氏名ふりがな，
  3) ご所属

  を記入の上，電子メールで懇親会受付（jaecs.reception@gmail.com）までお願いいたします。

◆会場案内図◆

⑪がマルチメディア教育研究棟，⑭が懇親会会場（川内サブアリーナ棟）です。

仙台駅からのアクセス（仙台市営バス）

仙台駅前のりば	行き先	下車停留所(所要時間・運賃)
9番のりば	宮教大・青葉台行 青葉通経由動物公園循環	東北大川内キャンパス・萩ホール前[バス停2-A]下車 （約15分、運賃180円）
	川内南キャンパス経由 （急行）東北大川内キャンパス	東北大川内キャンパス・萩ホール前[バス停2-A]下車 （約12分、運賃180円） ※平日午前の5便のみ
16番のりば	広瀬通経由交通公園・川内(営)行 広瀬通経由交通公園循環	川内郵便局前[バス停2-B]下車 （約15分、180円）

※所要時間は交通状況により異なります。

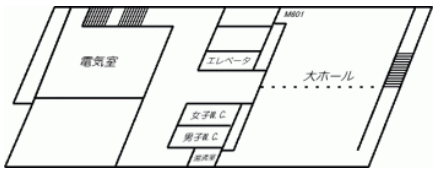
マルチメディア教育研究棟

平成15年3月，東北大学川内北キャンパスに新たな教育研究施設が完成しました。
21世紀の情報教育，語学教育，少人数教育，教養教育などに有効なマルチメディア対応機能を備えた教育研究施設です。主に全学教育に使用されています。
正規の教育プログラムに支障がない限り，大ホールは学術目的に使用可能です。

マルチメディア教育研究棟　各階のご案内

[6F][5F][4F][3F][2F][1F]

6F

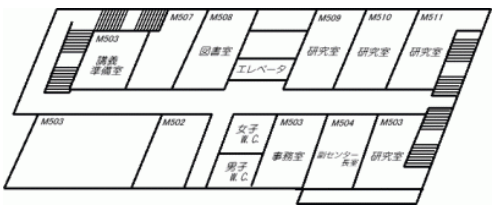


大ホール

研究発表会や会議用のホールで，通常座席数は126席。スクリーン1面と教材提示装置が設置されています。可動仕切り壁の操作により，2室に分割して利用可能です。

[ページの先頭へ]

5F

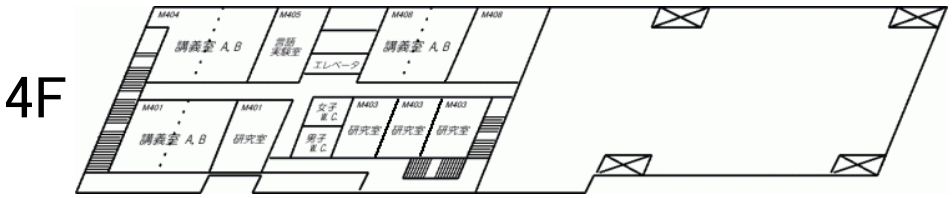


教育情報基盤センター情報教育部門

5Fには，教育情報基盤センター情報教育部門の教員研究室や事務室などが設置されています。

講義室

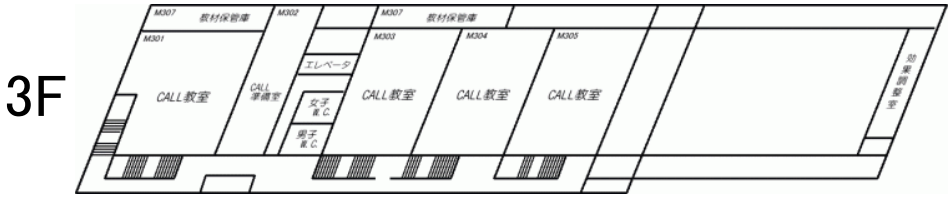
講義室は，各室とも床面積約100m^2，通常は46席。可動仕切り壁を操作することにより、少人数授業用教室2室に変わります。各室に，教材提示装置が設置されています。



[ページの先頭へ]

CALL教室

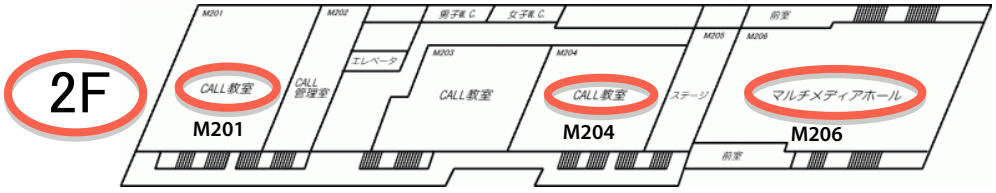
CALL(Computer Language Learning)による語学学習用教室が2Fと3Fに設置されています。学生用コンピュータ台数は，2Fと3Fで，あわせて324台設置されています。



[ページの先頭へ]

マルチメディアホール

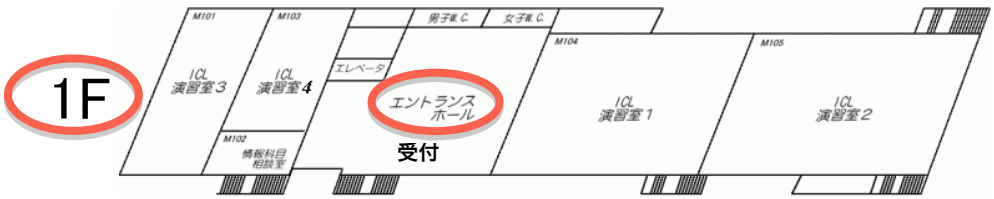
固定席431。大型スクリーン，電子黒板の機能をもつプラズマディスプレイ装置，教材提示装置，TVカメラ2台が設置され，照明や音響効果にも配慮した多目的ホールです。



[ページの先頭へ]

ICL演習室

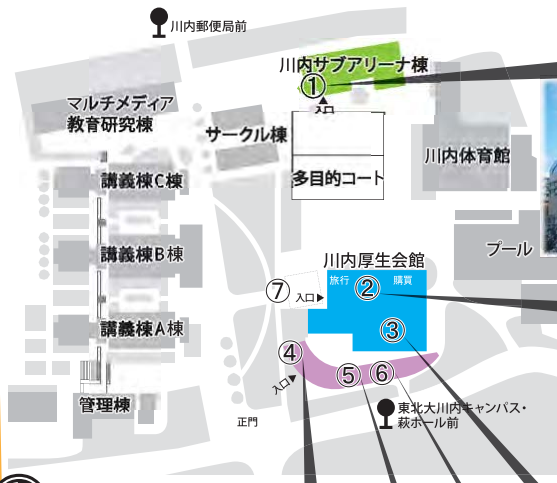
教育情報基盤センター情報教育部門のICL(Information and Computer Literacy)教育用施設として，演習室や相談室が設置され，端末機345台が設置されています。



[ページの先頭へ]

川内北キャンパスで「ゴハン」をたべるなら。

川内北キャンパス「ゴハンMAP」



① Bee ARENA Café ビーアリーナカフェ

営業時間　平日 11:00 ～ 20:00　土日祝 11:00 ～ 14：30

サブアリーナ棟にあり、２Ｆフロアもあります。
セットメニューがメインで、みなさんのお腹を満たしてくれます。

（たべ方）① 入口で食べたいメニューを選んで中へ。
② 手を洗い、トレイをとったらカウンターへ。
③「〇〇ください」と告げる。
④ セットの時はゴハンとみそ汁をとって先へ。
⑤ 惣菜、サラダをとりレジへ。
⑥ お会計後、箸・フォークをとってお席へどうぞ。

② 購買書籍店

営業時間　平日 8:30 ～ 18:30　土日祝 11:00 ～ 14:00

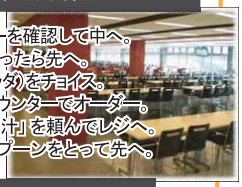
パン、おにぎり・弁当はもちろん、ドリンクやお菓子もあります。
学内で身近なコンビニ。食後にもぜひお立ち寄りください。

（かい方）① 買い物カゴを持って中へ。
② お店の右奥に「たべもの」はあります。
③ レジに並んだら財布を準備。組合員証（プリペイドカードに事前にお金を加金しておくと便利です。
④ レジ袋はお渡ししていないので、必要な場合、サービスカウンターへ。

③ 川内の杜ダイニング

営業時間　平日 8:00 ～ 20:00　土日祝 11:00 ～ 14:30

一日の始まりはバランスのとれた選べる３種の「定食」で。
昼夜は、一品ずつ選んで組み合わせて食べられます。



（たべ方）① 入口で本日のメニューを確認して中へ。
② 手を洗い、トレイをとったら先へ。
③ まずは小鉢（惣菜・サラダ）をチョイス。
④「メイン」を決めたらカウンターでオーダー。
⑤ 続けて「ゴハン・みそ汁」を頼んでレジへ。
⑥ お会計の前に箸やスプーンをとって先へ。

④ bush clover café ブッシュクローバーカフェ

営業時間　平日 7:50 ～ 8:50 ／ 11:00 ～ 17:00

朝は授業の前に「モーニングプレート」をどうぞ。（¥300）
パン工房で焼きたてのパンを淹れたてのコーヒーとともに。

（たべ方）① 入口で本日のメニューを確認して中へ。
② 焼きたてパンはトレイにのせてカウンターへ。
③ ベーグル・サンド、ドリンク等をオーダー。
④ ご注文後にお作りするので少々お待ちを…
⑤ お会計が済んだらお席へ。ごゆっくりお寛ぎください。

⑤ キッチンテラスCouleur クルール

麺コーナー

営業時間　平日 11:00 ～ 15:00　土日祝 閉店

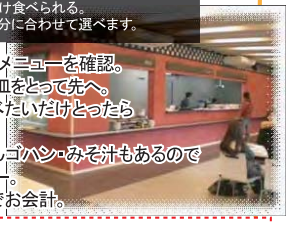
一日一麺。寒い日にはアツアツの麺を。
ラーメン・うどん・そばを食べられるのは平日だけです。

（たべ方）① メニューを見て食べたいものを確認。
② レジに並んでオーダー。先にお会計を済ませます。
③ 半券をカウンターへ順番に並べます。
④ ご注文後に作り始めるので少々お待ちを…
⑤ 番号を呼ばれたらトレイにのせて移動。
⑥ 箸とレンゲをとったらお席へどうぞ。

⑥ カレー丼・量り売りコーナー

営業時間　平日 8:00 ～ 20:00　土日祝 11:00 ～ 14:30

量り売り（100g＝126円）は食べたいだけ食べられる。
種類豊富なカレー・丼は、その日の気分に合わせて選べます。



（たべ方）① 並ぶ前にカレー・丼メニューを確認。
② トレーと量り売り用の皿をとって先へ。
③ 食べたいものを食べたいだけとったらコーナーを左へ。
④ カレー・丼、もちろんゴハン・みそ汁もあるのでカウンターでオーダー。
⑤ 最後にレジへ並んでお会計。

＊ご利用時のお願い

- 混雑時の席取りはご遠慮ください。譲りあってご利用ください。
- ゴミはゴミ箱へ。分別にもご協力ください。
- 食べ終わったら、それぞれの下膳コーナーへ食器をお戻しください。
- ミールカードご利用の方は、お会計時に「ミールで!」とお声掛けください。

＊朝ゴハン＊＊＊朝からたべられるのはコチラ

～ちょっと早起きにはなるけれど、一日の始まりは「朝ゴハン」から。

④ bush clover café	…7:50～8:50
③ 川内の杜ダイニング	…8:00～
⑥ カレー・丼・量り売りコーナー	…8:00～
② 購買書籍店	…8:30～

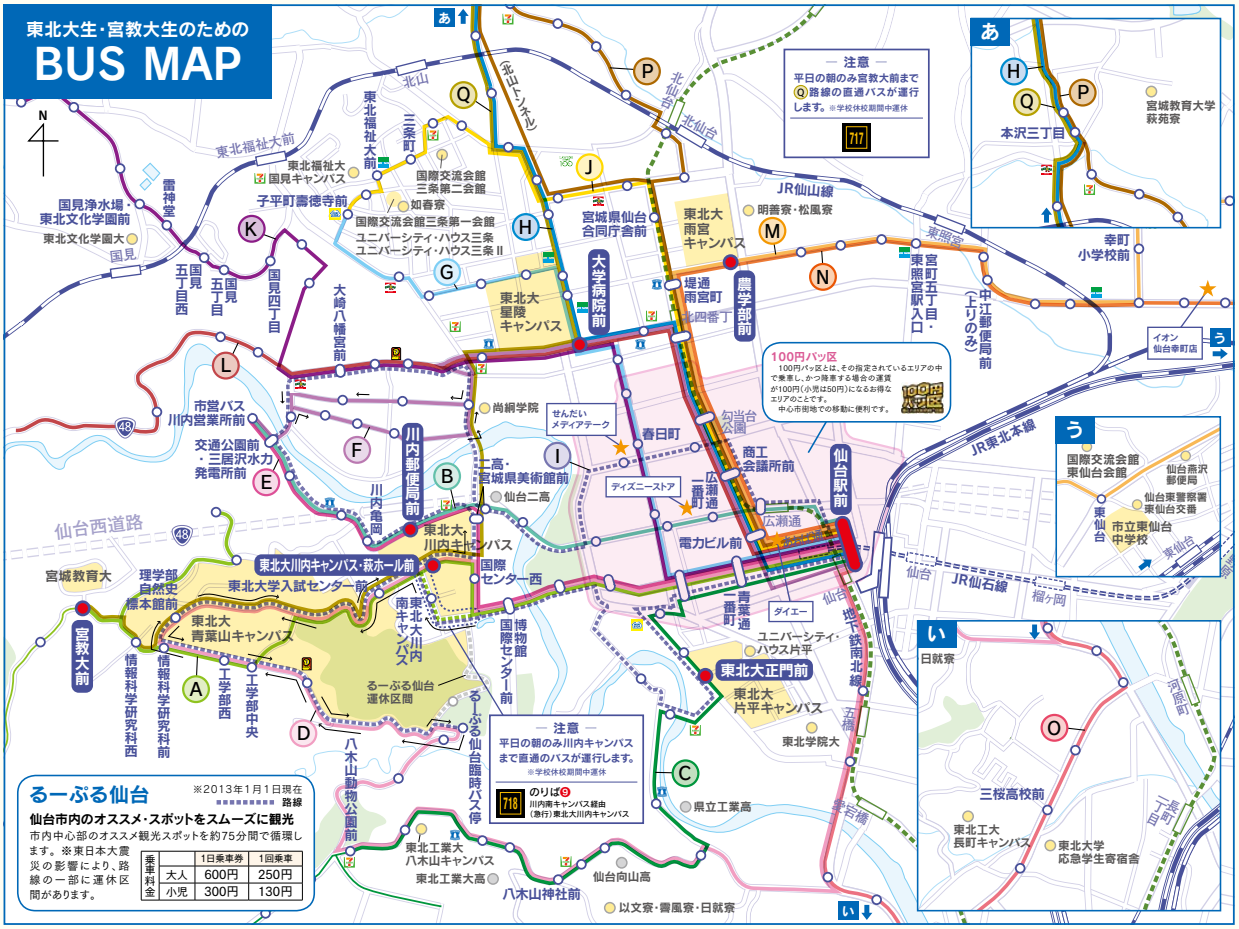
＊晩ゴハン＊＊＊夜でもたべられるのはコチラ

～授業やサークルが終わったらみんなで楽しく「晩ゴハン」を。

① Bee ARENA Café	…～20:00
③ 川内の杜ダイニング	…～20:00
⑥ カレー・丼・量り売りコーナー	…～20:00

univ. CO-OP 東北大学生活協同組合

東北大生・宮教大生のための BUS MAP



仙台駅のりば

路線	のりば	系統番号
A	❾ 下り	710 713 715
	国際センター、東北大学川内キャンパス、東北大学青葉山キャンパス、仙台市博物館、宮城教育大学	
B	⓰ 下り	730 739
	ディズニーストア、宮城県美術館、東北大学川内北キャンパス	
C	⓫ 下り	706 708 700 705
	⓬ 下り	707
	東北大学片平キャンパス、八木山動物公園	
D	⓫ 下り	609 ※時計回りの循環のため利用できますが遠回りです。
	ダイエー、東北大青葉山キャンパス、国際センター、仙台市博物館、八木山動物公園	
E	❾ 下り	720
	ダイエー、仙台市博物館、宮城県美術館、川内北キャンパス	
F	⓾ 下り	830 839
	※AM6～7時は⓰です。	
	ダイエー、県庁市役所、東北大学星陵キャンパス	
G	㉕ 下り	890 899
	ダイエー、東北大学星陵キャンパス、東北大学国際交流会館、ユニバーシティ・ハウス三条	
H	⓭ 下り	800
	ダイエー、東北大学星陵キャンパス、県庁市役所、宮城教育大学・萩苑寮	
I	15-3	この路線は、「るーぷる仙台」です
J	㉔ 下り	990 999
	ダイエー、県庁市役所、東北大学国際交流会館、ユニバーシティ・ハウス三条	
K	⓯ 下り	880
	ダイエー、東北大学星陵キャンパス、大崎八幡宮	
L	⓾ 下り	840 845 850 860
	ダイエー、東北大学星陵キャンパス、大崎八幡宮、文殊堂前	
M	⓳ 下り	110
N	⓲ 下り	120
	東北大学雨宮キャンパス、イオン仙台幸町店	
O	⓾ 下り	620
	東北大学応急学生寄宿舎	
P	⓭ 下り	900 ～ 905、923 925
	宮城教育大学・萩苑寮 ※923と925は北山トンネル経由	

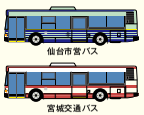
※掲載している情報は2013年1月1日現在のものです。変更になっている場合があります。ご注意下さい。

仙台駅のりば案内

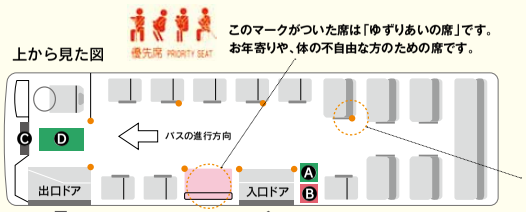
仙台市を走る路線バスの乗り方

車体中央にあるドアから乗って、
前のドアから降ります。
運賃は降りる時に払います。

乗り方はどちらも一緒です。



上から見た図



席にすわれない時は「手すり」につかまりましょう。

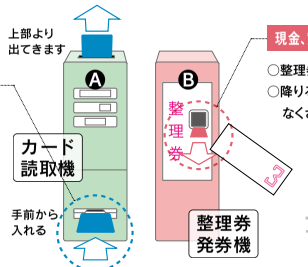
バスに乗る

1.乗車バス停の記録

バスカードの場合

○カード読取機にバスカードを入れて下さい。整理券を取る必要はありません。

バスカードについては、中面をご覧ください。



現金、定期券、フリーパスの場合

○整理券をお取り下さい。
○降りる時に必要になるので、なくさないでください。

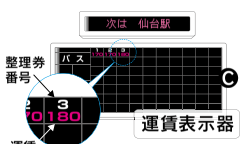
2.降車の意思表示

○次の停留所が、降りたいバス停になったら、車内にあるボタンを押します。

3.運賃の確認

○運賃はバス車内前方にある運賃表示器で確認します。

この場合は、自分の持っている整理券が3番なら、運賃は180円になります。



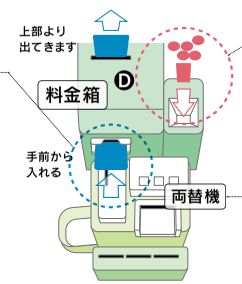
※上記金額は一例です。

降りるバス停が近づいてきたら…

4.運賃の支払い

バスカードの場合

○料金箱のカード読取機に、バスカードを入れてください。自動的に運賃が差し引かれます。



現金、定期券、フリーパスの場合

○運賃表示器に表示された金額と整理券を一緒に現金投入口に入れてください。
○釣銭は出ませんので、あらかじめ車内の両替機で両替してください。

1,000円札と硬貨が両替できます。両替は、バス停車時にしましょう。

○定期券、フリーパスの方は、整理券を入れて定期券、フリーパスを乗務員にお見せください。

バスを降りる

［制作］ 東北大学［学生支援課］ 022-795-7816
宮城教育大学［学生課］ 022-214-3595

※機器類のイラストは実際のものと異なる場合があります。

Sendai Bus Handbook

Sendai SMART

杜の都「仙台」の学生には、環境にやさしい「バス」が似合います。

「杜の都」と呼ばれる仙台だけど…

みなさんも知っての通り、日本では、地球温暖化の原因となる温室効果ガス排出量は年々増加…
地球温暖化の進行は、異常気象や自然災害などが頻発したりする可能性がある深刻な環境問題です。

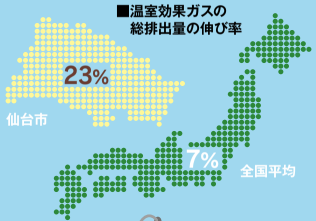
仙台市では…

1990年から2005年にかけての日本全体のCO_2排出量の伸び率約13%に対して、仙台市は約23%（約2倍）になってしまっています。特に仙台市の運輸部門のCO_2排出量に限っては、約31%も増加……

さらに同時期の日本全体の温室効果ガスの総排出量の伸び率約7%に対して、仙台市は約23%（約3倍）にもなっているのです。

【参考資料:仙台市環境局資料】

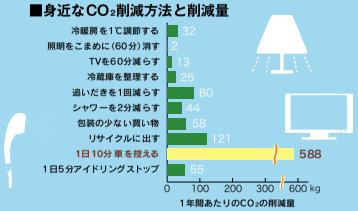
■温室効果ガスの総排出量の伸び率



なぜバスは環境にやさしいの？

仙台市ではCO_2排出量の削減に取り組んでいます。また、CO_2排出量削減は一人ひとりの行動が重要です。例えば下の表にある通り、日々の生活においてCO_2排出量を削減することができるのです。

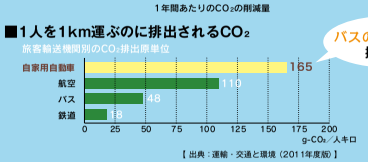
■身近なCO_2削減方法と削減量



さらに、この表を見て明かな通り、移動の際、車に乗るのを控え、バス等の公共交通機関を上手に利用することで、CO_2排出量を大幅に削減できるのです。

【参考資料:太田裕之、藤井聡:「環境配慮行動における客観的CO_2排出削減量事実情報提供の効果に関する実験研究」】

■1人を1km運ぶのに排出されるCO_2



歩くのも、自転車に乗るのも
公共交通に乗るのと一緒でエコ！
とっても健康にもいいしね！

仙台市にはどんな路線バスが走っているの？

バスの乗り方
概略バスマップ

2013年9月1日発行
編集・発行 英語コーパス学会
会長　　堀　正広
事務局　〒560–0043 大阪府豊中市待兼山町 1–8
大阪大学大学院言語文化研究科
田畑 智司研究室気付
電話：06–6850–5866
e-mail: jaecs.hq@gmail.com　twitter: @JAECS2012
URL: http://english.chs.nihon-u.ac.jp/jaecs/