

タイトル	英語ライティングクラスのための観光コーパス構築に関する研究
著者	上野，之江；尾田，智彦；佐々木，勝志；田中，洋也；森越，京子；UENO, Yukie；ODA, Tomohiko；SASAKI, Masashi；TANAKA, Hiroya；MORIKOSHI, Kyoko
引用	北海学園大学学園論集(151)：17-46
発行日	2012-03-25

英語ライティングクラスのための 観光コーパス構築に関する研究

Compiling a Corpus from Hokkaido Tourism Websites for English Writing Classes

上	野	之	江*
尾	田	智	彦**
佐々	木	勝	志***
田	中	洋	也****
森	越	京	子*****

This article reports how the authors created a corpus from official Hokkaido tourism organization websites. The Hokkaido Tourism Corpus data were first divided into 10 categories such as food, tourist spots, and festivals. Then using *Kensaku*, a concordancer developed by one of the authors, the corpus was analyzed qualitatively and quantitatively. Then, frequency lists on verbs and adjectives were extracted. Based on the lists, the authors developed teaching materials to facilitate students' writing, specifically introducing students' hometowns in Hokkaido. The authors believe that such vocabulary lists and expression lists are essential for students' writing on their local areas as well as improving their communication skills in writing.

1. 研究の背景

国際化の進展と共に、学生に求められる英語力の方向性も大きく変化しつつある。伝統的に我が国の英語教育は、欧米からの正確な知識・技術の導入のために文献の精確な読解力の養成に重点が置かれ、一般的な英語教育も英米の文化や文学等を題材にしたものが多かった。一方で、昨今の実用的な英語学習のブームとも言うべき状況の中では、英語のコミュニケーション能力と「英会話」の能力を半ば同一視し、「聞くこと」「話すこと」のみが重視されるような傾向もある。

しかしながら、2009年公示、2013年施行の文部科学省高等学校学習指導要領では「外国語科の目標」について

外国語を通じて、言語や文化に関する理解を深め、積極的にコミュニケーションを図ろうとする態度の育成を図り、情報や考えなどを的確に理解したり適切に伝えたりするコミュニ

ケーション能力を養う。

と定め、『学習指導要領解説』の中で、「この『コミュニケーション能力を養う』には、」という部分に、以下のように解説している。

今回の改訂により、中学校段階においても4技能を総合的に育成することとなっており、高等学校においては、中学校における学習の基礎の上に「聞くこと」、「話すこと」、「読むこと」及び「書くこと」の4技能を総合的に育成するための統合的な指導を行い、生徒のコミュニケーション能力を更に伸ばすことが大切である。

つまり、コミュニケーション能力は会話力（「聞くこと」、「話すこと」）だけでなく、全ての技能のバランス良い育成を目指すべきとしている。英語教育の現状を考え合わせれば、実践的コミュニケーション能力育成には、「書くこと」を含めた発信的な英語力の養成が急務である。

大学生や社会人のレベルでも、1979年の開発以来世界120カ国で実施され、年間約600万人が受験(2010年)するTOEIC®テストにおいて、従来のListeningとReadingのみのテストに加え、2007年よりSWテスト(TOEIC® Speaking and Writing Tests)を積極的に実施するようになっている。この事実も、Writingを含む発信的な英語力がますます重視されている事を示していると言えよう。

一方で、我が国の実際の英語教育の場面では、一部を除き、英語での自由な自己表現の機会が十分に与えられているとは言い難い。Kobayakawa (2011)は、高等学校の「英語Ⅰ」「英語Ⅱ」および「ライティング」の検定教科書各5冊について、「書くこと」の課題(task)の量的比較分析を行った。その結果、4技能の総合的な訓練を目指す「英語Ⅰ」および「英語Ⅱ」では、穴埋め(fill-in-the-blanks)問題をはじめ、制限作文や日本文を見て一文を埋める問題等が多く、「ライティング」では和文英訳や制限作文の課題が多く設定されていた。自由英作文の課題は、各5冊の課題の総数の中で「英語Ⅰ」では2.39%、「英語Ⅱ」では3.47%、そして「ライティング」においても3.35%を占めているに過ぎない。文部科学省の言う「実践的コミュニケーション能力」の育成のためにも、自由英作文を積極的に取り入れ、書く機会を増やす必要がある。更に、ここに示されたような英語教育を受けてきた高校生を受け入れる大学の英語教育においては、自由に自分たちのことや身近なことを英語で表現する機会を与えることは、その大きな柱となるべきであろう。

1.1. 先行研究

筆者らは1998年に、道内6つの大学の日本人学生と、海外の複数の大学の学生とのEメール交流を実施し、日本人学生の書いた英文Eメールから約10万語のコーパスを構築し、その語彙的な

特徴を分析した²⁾。その結果，学生が使用した語彙内容には，標準的な高頻度の基本的語彙や，学生・若者文化，現代性・時代性，日本文化などを表現する語彙に加え，北海道という地域性を表現する語彙にその特徴が見られた。つまり，北海道の大学で指導する筆者らにとっては，北海道という地域にかかわりの深い語彙・表現内容を教示することが，学生の英語での自己表現を促すための重要な手掛かりとなり，そのための研究および教育実践が強く求められている。

学生が住み育った地域を自己表現のテーマとすることは，近年社会的にも求められている。例えば，政府（国土交通省）の「観光立国推進基本計画」（2007年）の発表に象徴されるように，海外から日本により多くの旅行者（in-bound tourists）を招き，日本のことを英語で紹介し伝えるような英語力は，今後の日本のあり方を考えれば観光学専攻の学生に限らず多くの学生の取って不可欠なものとなって行くであろう。観光を今後の産業の1つの柱に据えようとする北海道にあっては，尚更のことである。

英語教育の観点から見ると，地域と関わる英語表現の指導や，学生の英語による自己表現を支援することに繋がる分野としては「観光英語」が存在する。しかしながら，大学での通常の観光英語などの教育が，日本のこと，とりわけ自分たちの地域のことを英語で発信するという方向で実施されているとは言えない。大学英語教育の教科書を例に挙げれば，14社の出版社が共同で運営するウェブサイト「大学英語教科書協会」のページ³⁾で「観光英語」のジャンルで調べると，47件の教科書がヒットするが，その大部分は日本人が外国へ出かけて行き，そこでの様々な会話を想定しているか，ホームステイに出かけての会話などを想定しているものである。明確に日本で外国人に日本のことを紹介・説明していると考えられるものは，僅かに3冊に過ぎない⁴⁾。

大学を中心とした英語教育の研究でも，観光英語あるいは Tourism English を視野に入れた調査・研究が行なわれつつある。日本のことを表現するための英語の研究として，Chujo, Utiyama & Oghigian(2006)は，885部の京都の観光案内資料から‘miru’(sight-seeing)，‘kau’(shopping)，‘taberu’(dining)，‘taikensuru’(hands-on activities)の4分野の42,025語から成る Kyoto tourism corpus を作成し，それを BNC (British National Corpus) の高頻度語と5種類の統計的方法で比較した。それにより，3段階のレベルでの，上記4分野別の学習目標語彙の抽出が可能となった。また，Fujita & Tsushima (2010)は，4分野の Tourism English⁵⁾ から各30,000語を抽出した Tourism Corpus を作成し，それを一般的な英語からの240,000語のコーパスと比較分析することによって，tourism (観光学) 専攻の大学生が学習すべき語彙リストの抽出を試みている。

しかしながら，京都という地域を題材にした英語表現に関しては，Chujo らは以下のように述べている。

This Kyoto tourism data covers various aspects of modern and traditional Japan, including its history, culture, current events, and local tourism attractions.

つまりそれは、現代の日本の状況に加え、伝統的な意味で日本的なものが色濃く反映されたものであり、北海道の状況を表現するためには、必ずしも適切なものとは言えない。

我々の身近な地域のことについて英語で紹介するための教育や、そのための研究、そして教材開発は、喫緊の課題である。

2. 研究の目的

これまで述べたような状況を踏まえ、北海道の大学で教える筆者らの学生に対して、自由な自己表現を促し、発信的な英語力を育成するためにも、北海道の地域性をうまく表現するための語彙や文章表現を教示する必要がある。それは、北海道の地域性を英語の表現において明らかにすることでもある。

従って、筆者らは北海道の地域性を表現する英語を研究し、また学生の自己表現を促すための教材を開発する基盤として、Hokkaido Tourism Corpusを作成することとした。更に、日本を表現するための英語のこれまでの研究は語彙のレベルに特化したものが多かったが、筆者らは、具体的な文章表現のレベルでも、適切な教材を提示することが不可欠と考えた。

本研究では、Hokkaido Tourism Corpusの作成の詳細を記述すると共に、その活用の方向性を示すことを目的とする。その研究成果の報告として本稿では、1) 北海道の観光ウェブサイトから抽出したHokkaido Tourism Corpusの作成過程を詳細に記述する、2) 作成したコーパスを分析し、使用語彙の特徴を考察する、3) その分析結果を基に学生が英語で自己表現するための手掛かりとなる語彙と表現の教材を作成した過程を報告する。

3. Hokkaido Tourism Corpus (HTC) の作成

3.1. 作成の手順

コーパス作成にあたっては、テキストデータの収集とコーパスデータの構築が必要となる。以下の点を考慮に入れグループで統一した基準を持ち作業をした。

3.1.1. テキストデータの収集について

北海道の英語版観光ウェブサイトから語彙と英文を抽出することにした。データ収集の基準を以下のように決めた。

1) データ収集地域の選定

どの地域の観光情報を収集するかグループ内で話し合い、『北海道観光の概況(H16)』(北海道経済部観光振興課(2004))、『北海道外客来訪促進計画：国際観光推進プログラム—ようこそ北海道—』(2005)から、訪問する外国人観光客数、宿泊者数上位20地域を抽出し選定した。将来学生用教材を作成する場合、必要な語彙が多く出てくる可能性があると考えたからだ。

（表1）『北海道観光の概要（H16）』より

外国人観光客数が多い市町村

観光客		宿泊客	
k 1	札幌市	s 1	札幌市
k 2	小樽市	s 2	函館市
k 3	旭川市	s 3	釧路市
k 4	千歳市	s 4	登別市
k 5	函館市	s 5	上川町
k 6	釧路市	s 6	帯広市
k 7	洞爺湖町	s 7	倶知安町
k 8	登別市	s 8	小樽市
k 9	喜茂別町	s 9	旭川市
k 10	上川町	s 10	洞爺湖町
k 11	帯広市	s 11	北見市
k 12	白老市	s 12	富良野市
k 13	伊達市	s 13	ニセコ町
k 14	石狩市	s 14	網走市
k 15	砂川町	s 15	音更
k 16	富良野市	s 16	斜里町
k 17	七飯町	s 17	稚内市
k 18	壮瞥町	s 18	伊達市
k 19	倶知安町	s 19	留寿都町
k 20	ニセコ町	s 20	弟子屈町

表1から，以下の5地域，6市町村を選定した。

（地域） 道央，道南，道北，十勝，オホーツク

（市町村）札幌，小樽，千歳，釧路，洞爺湖・登別，ニセコ・喜茂別・白老

市町村後半の「洞爺湖・登別」「ニセコ・喜茂別・白老」はそれぞれが小さな単位で，「洞爺湖・登別」は観光地である洞爺湖周辺としてひとつにくることができる。同様に「ニセコ・喜茂別・白老」も羊蹄山周辺，ニセコ周辺にくることができるために，2町村又は3町村でひとつとした。

2) データ取得基準

データ取得基準

Aランク：市町村のHP，観光協会のHP，公共性の高いもの

Bランク：NPO，商業関連サイト

Aがない場合はBを取る

どのような発信者のサイトなのかを確認し，公共性の高いホームページサイトから収集した。具体的には，第1候補として，収集対象の各市町村の公式ホームページの英語版観光案内からテ

キストデータの収集をするようにした。もし、市町村が英語版観光案内を持たない場合は、観光協会のホームページを第2候補とした。公的なものを収集判断の基準としたので、個人発信のサイトは除いた。WikiTravelも除くことにした。実際には、Aランクのデータを多く収集した。

ホームページを選んだのは、電子化が容易なためである。また、コピーライトの問題についてもインターネット上で公開している資料なら転記の問題はクリアできると判断した。英語版ホームページを持たない市町村の観光案内については、当該地市町村役場、観光協会が発行している印刷版より語彙収集をした。

3) データ収集の方法

ホームページに、数ページに及ぶ細かい情報がたくさん出てくる場合(ホテル情報、レストラン情報など)、又は情報が階層化されている場合は、第1画面だけを収集するようにした。ホームページからリンクがある特定施設のHP(動物園、ロープウェイなど)は入れなかった。

3.1.2. コーパスの作成

集まったテキストデータは、ホームページ1ページ分を1データとして、ひとつひとつにタグを付けた。出所情報を明らかにするため、タグ2行目<>の下に、[]を入れ、パンフ名、URLアドレスを付加した。

(コーパス1データの例)

<hp:道央:ニセコ町:ニセコ:act> ←データの内容を示すタグ

[Niseko Promotion Board <http://www.nisekotourism.com/yellow-page/index.php?lang=en&g1=02&g2=01>] ←データの出所を示す

ActivitiesTHINGS TO DO: What to do in Niseko (summer and winter)

Niseko is situated at the base of a dormant volcano Mount Yotei, and next to one the most beautiful rivers in Japan the Shiribetsu River and, as such, is well positioned to offer a variety of exciting summer and winter activities that can be enjoyed by visitors of all ages.

In spring, having lain under the icy snow for six months of the year, the Hokkaido landscape awakens and bursts in to life. While famous for a high annual snowfall and winter sports, Niseko is equally well-known for offering outdoor activities that take place against the backdrop of the beautiful summer landscape and firey autumn scenery. Although skiing and snowboarding is a popular choice for many visitors, the area offers a multitude of ways to enjoy Hokkaido's great outdoors.

1) タグの詳細

タグは< >で前後を囲み，この中にデータの取得情報，内容についての情報を書き入れた。タグを付けた理由は，コーパス内の語彙分析に加え内容について細かく分析するためである。以下の点について統一した。タグ内のコロンは必ず4つとし，コロンの数は変えない。複数のタグ項目を入れたい時は，スラッシュ (/) で分けた。

(タグのテンプレート) <hp/pa:圏域:市町村:観光地名:カテゴリ///>
① ② ③ ④ ⑤

①データはホームページから取ったものか，印刷から取ったものか明記した。

hp: homepage から収集したデータ

pa: 印刷物から収集したデータ

②圏域: 地域の名前（道央，道南，道北，十勝，オホーツク）のどれかになる。

③市町村: 札幌，小樽，千歳，釧路，洞爺湖・登別，ニセコ・喜茂別・白老のようになる。

④観光地名: 有名な観光地はその名前も記した。（定山渓，層雲峡など）

⑤カテゴリ: そのテキストデータを，記載内容を元に以下の 11 項目に分類した。複数のカテゴリに該当するページに関しては，いくついれても構わない。その場合は，カテゴリ間をスラッシュで区切った。

<food> food

<act> activity （参加するもの）

<acd> accommodation ホテル情報が多い時は最初のページだけ取る

<acs> access, transportation, going in and going around, 運賃

<ovw> overview, 概要, 歴史, 地理, 天気, what to see, what to do, agriculture

<spt> spot 見所 sightseeing, モデルコース

<wlf> wildlife

<pinfo> practical information ATM/Money/Costs/（滞在費用）

<faq> FAQ

<fes> festival, 祭り 町おこしのイベントを含む，〇〇祭り

<shop> 買い物，お土産

(Tag 見本) Tag Sample

1 行目	<hp/pa : 圏域 :	地町村 :	観光地名 :	カテゴリ / / />
	道央	ニセコ町	層雲峡	food/act/acd/acs/ovw
	道南	上川	十勝川温泉	spt/wlf/pinfo/faq/fes/shop
	道北	伊達	中山峠	
	十勝	釧路・根室		
	オホーツク	七飯		
		札幌		

各市町村の英語サイトはそれぞれの自治体によりばらつきがあった。お金をかけて十分に用意されたもの、素人が作成したと思われるもの等多様であった。北海道全域から満遍なく収集したいと思ったが、すべての市町村が英語サイトを持っているとは限らず、英語の質のよいものを選択しようとする、地域別収集量に偏りが出た。しかし、北海道という大きな枠で考えれば、北海道という地域性が反映された語彙が抽出されることに違いはないと考えた。

内容についてのタグ付けは研究グループの2名から3名で行った。意見が分かれたときは多数決で決めた。ウェブサイトのフロントページは、その地域、市町村の概観、紹介が書かれていることが多い。その部分をまとめて<ovw> : Overview としたが、そのフロントページの内容は多彩で、地域の概要、歴史、地理、天気、What to see, what to do, agriculture 等になった。この部分をもっと細かく仕分けする必要があったかもしれない。

タグをつけた理由は、第1に各データのインデックスとして利用できることが上げられる。タグを見ただけで、そのデータの出所、大まかな内容が一目で把握できるのは便利である。第2に、内容別に語彙分析をする時に便利と考えた。例えば、<wlf> (wildlife) というタグを手掛かりにすることにより、野生動物について学生が書きたい時にどのような語彙が役立つか、ソートをかけて抽出することが可能になると考えたからである。

このような手順で蓄積したコーパスデータを Hokkaido Tourism Corpus (以後 HTC と記述する) と名づけ、研究グループのメンバーが作成した分析ソフト *Kensaku*⁹⁾ を利用して分析を行った。

4. Hokkaido Tourism Corpus (HTC) の分析

4.1. 分析方法

本稿の筆者の1人が開発したコンコーダンサー *Kensaku* を用いることによって、作成したコーパスから頻度順の語彙表を作成したり、KWIC (Key Word in Context) 検索によって特定語彙を含む表現 (文) を検索することが可能になる。また、*Kensaku* のカウント機能を利用して以下

の数量的分析を行った。

4.2. 分析結果と考察

1) Hokkaido Tourism Corpus (HTC) の量的分析結果

総データ数：	161	データ
総語数：	58340	語
文数：	4037	文
異なり語数：	6703	語
TTR：	11%	

収集したテキストデータは 161 データであった。北海道の市町村や観光協会が主に作成した公的な観光案内 HP サイトから 161 ページ分が収集蓄積されたことになる。*Kensaku* でカウントする場合に支障となる「*」などの記号は取り除いたが、テキストはそのまま編集はしていない。

総語数は、58,340 語で、前出の Kyoto tourism corpus が 885 部の京都観光案内資料から 42,025 語を収集したのと比較しても、同じような規模となった。

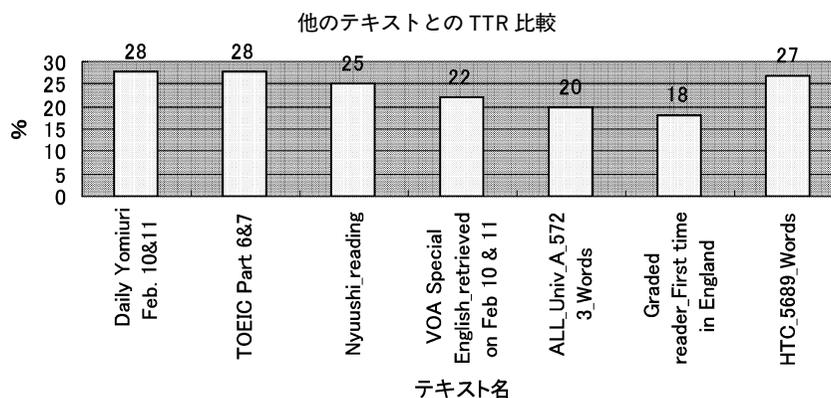
文数は 4,037 文で、1 文平均 14.45 語となる。異なり語数⁷⁾は 6,703 語である。TTR は 11% であった。⁸⁾

TTR が示すのはテキスト中の語彙の密度である。同じ語数のテキストならば、同じ単語を繰り返し使うほど、テキストは水で薄めたように意味的に粗いものになる。このようなテキストは TTR が低くなる。これに対し TTR が高いテキストはそこに使われている単語の種類が多く、意味的に密なものとなる。(斉藤他, 1998) Spoken テキストと Written テキストを比較すると Spoken テキストの方が TTR は低くなる。

斉藤他 (1998) は総語彙数を統一してテキストの比較をしている。本稿でも総語彙数をほぼ統一して比較した。こうすることで HTC の語彙使用のレベルがわかる。

比較したテキストと総語彙数・TTR

	Tokens	Types	TTR
Daily Yomiuri Feb. 10 & 11	5643	1592	28%
TOEIC Part 6 & 7	5289	1521	28%
Nyuushi_reading	5544	1400	25%
VOA Special English_retrieved on Feb 10 & 11	5761	1284	22%
ALL_Univ_A_5723_Words	5723	1158	20%
Graded reader_First time in England	5544	1016	18%
HTC_5689_Words	5680	1584	27%



比較したテキスト⁹⁾は、英字新聞(Daily Yomiuri)、TOEIC 練習問題のリーディング部門 Part6 と Part7、VOA (Voice of America) Special English サイトのスクリプト、初級者用の Graded Readers、大学入試問題の長文問題である。これらを 5,700 語前後収集し、HTC から任意に編集した 5,700 語を比較した。語彙密度が高い、リーディングの難易度が高いと予想されるのは、Daily Yomiuri と TOEIC 練習問題、入試のリーディング問題であった。結果は予想通りであった。一番やさしく TTR が低かったのは、初級者用の Graded Readers であった。VOA は Special English サイトのニューススクリプトなのでこれも初級者用に語彙、文法の程度を抑えた平易な英語に徹している。

ALL_Univ_A_5723_Words は筆者らが 1999 年に調査した大学生の E-mail コーパスから編集した約 5,700 語である。今から 20 年前の大学生が書いたメール文章は初級者用の Graded Readers よりは語彙密度が高いが VOA Special English よりは低いのは納得がいく結果である。

これらのテキストと HTC の語彙を比較すると、このコーパスの内容は英字新聞や TOEIC 問題よりは低く VOA Special English より高いということになる。内容は観光についてなので、英文としては理解しやすいものになっていると考えられる。

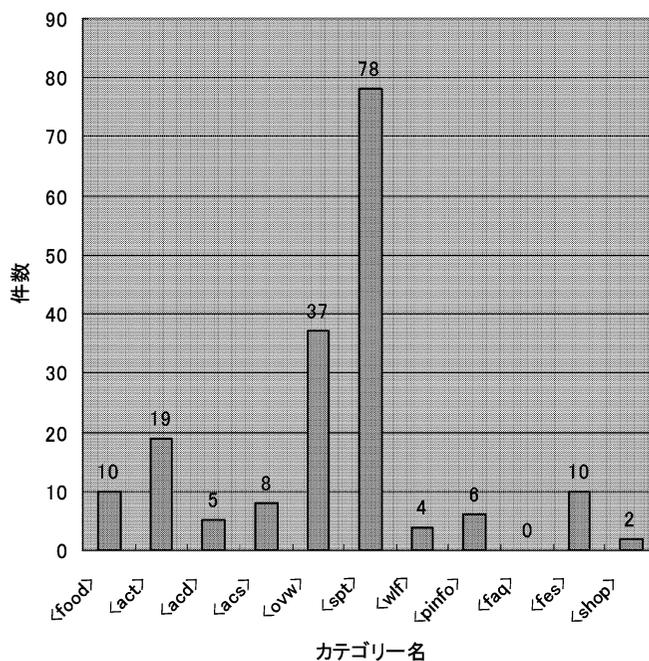
2) カテゴリー別データ数

次にコーパス内の内容について調べた。コーパス作成時に内容を読みひとつひとつのデータに付記したカテゴリーのタグを数えた。これにより HTC にはどんな内容が多く含まれているかがわかる。

カテゴリー別データ数

<food>	10
<act>	19
<acd>	5
<acs>	8
<ovw>	37
<spt>	78
<wlf>	4
<pinfo>	6
<faq>	0
<fes>	10
<shop>	2
合計	179

HTC カテゴリー別データ数



総データ 161 につけたカテゴリータグは 179 であった。ひとつのデータが 2 つ以上のカテゴリーを持つ場合がある。数が一番多かったのは、78 データあった <spt>spot で観光地の見所、観光スポット紹介、モデルコースがこれにあたる。次に 37 データあった <ovw>overview が続く。このデータの大半は各 HP のフロントページである。その地域、観光地の概要、歴史、地理、天気、見所などが大まかに紹介されているページである。様々な情報が小さく散りばめられている。農業に関する紹介もこの中に入れた。今後は、この overview を精査してさらに内容を細かく分けることが必要になるだろう。

3) 使用頻度別語彙リスト

Kensaku で使用頻度数別に語彙リストを作成すると以下ようになった。以下に示すのは、使用頻度が高い順に 1 位から 250 位まで並べた語彙である。1 列に 50 の語彙が示されている。

Kensaku 使用頻度数別語彙リスト (1位~250位)
(2010_Analysis_10724_CorpusDataConnected_Freq_lemma)

the	3520	shop	126	most	81	local	60	product	47
of	1830	was	125	car	80	their	60	monument	46
and	1731	Nemuro	122	may	79	great	59	bath	46
in	1394	make	122	close	79	April	59	hold	46
a	1296	many	121	locate	79	start	59	Goryokaku	46
is	1023	time	119	place	79	like	59	Abashiri	46
to	1015	about	119	spt	78	know	58	through	45
it	472	also	118	bus	78	but	58	marsh	45
from	467	town	118	Akan	78	world	58	available	45
as	429	its	116	Otaru	77	visitor	58	snow	45
are	425	There	112	large	76	over	58	drift	45
for	420	sea	112	your	76	high	57	train	45
you	400	year	111	beautiful	75	road	57	Sounkyo	44
on	395	open	110	season	75	late	57	house	44
Hakodate	389	day	109	style	75	spot	57	tram	44
at	366	mountain	108	Okhotsk	74	surround	57	street	44
can	363	min	106	html	74	index	56	event	44
have	340	where	105	com	72	along	56	kind	44
with	338	no	103	natural	71	find	56	e	43
lake	304	Kushiro	102	so	70	new	55	after	43
by	300	get	101	such	70	Mt	55	red	43
be	298	Obihiro	101	JR	70	htm	55	location	42
this	274	tourist	100	old	69	facility	55	visit	42
park	274	fish	99	summer	69	come	54	village	42
that	271	nature	99	room	68	site	54	include	42
Hokkaido	257	winter	99	Ainu	68	km	52	become	42
city	228	course	98	internation	68	route	52	call	41
area	224	water	97	famous	68	Shiretoko	52	history	41
spring	189	will	97	they	66	than	52	green	41
Asahikawa	169	all	95	restaurant	66	bird	52	Tokachi	41
station	166	only	92	association	66	some	52	temperatur	41
http	162	map	91	offer	66	center	52	range	40
or	161	walk	91	food	66	adult	52	way	40
enjoy	159	more	91	Motomachi	66	fresh	52	very	40
www	159	build	91	english	65	night	51	between	40
hot	157	there	90	we	65	here	50	Tel	40
hp	157	flower	89	forest	65	popular	50	Center	40
take	151	use	88	off	64	crane	50	country	40
which	149	jp	87	tour	64	autumn	49	please	40
Japan	145	museum	85	tree	64	end	49	farm	40
one	138	around	85	Shikotsu	63	into	49	Tourism	40
see	137	river	85	first	63	kankou	49	number	39
festival	136	other	84	hour	62	if	49	chitose	39
ice	133	foot	84	experience	62	every	48	good	39
cho	133	hotel	84	not	62	October	48	main	39
minute	133	when	83	well	61	name	48	guide	39
Japanese	131	Chitose	83	during	61	hall	48	full	39
yen	131	people	82	up	61	free	48	small	39
view	127	early	82	stop	61	ski	47	two	39
Onuma	126	national	81	go	61	Sta	47	white	38

一番左側の 1 列は、頻度別 1 位から 50 位までの語である。そのうち内容語（Content words：固有名詞，名詞，動詞，形容詞）は 21 語であった。例えば 49 番目の view に関しては、*Kensaku* の KWIC_A_Search にかけるとその前後関係（コンテキスト）より、動詞として使われている場合が 51 例，その他は固有名詞や名詞として使われていることがわかる。

（固有名詞） Hakodate, Hokkaido, Asahikawa, Japan, Onuma, view (10)

（名詞） lake, park, city, area, spring, station, festival, ice, minute, yen, view(66),

（動詞） enjoy, take, see, view (51)

（形容詞） hot, Japanese

（ ）内の数字は出現回数を示す

これらの語を鳥瞰することで、HTC の内容が推察できる。Hakodate, Asahikawa, Onuma のデータが多くある。Lake, park, ice, view など自然の紹介，季節は spring の言及が多く，festival などの行事や city の情報も多く含まれる。観光客向けにアクセス情報として station, minute, 費用情報として yen が含まれていることが推測できる。Hot は hot spring（温泉）を連想させる。これらの情報は北海道に住む執筆者らの印象と違和感はない。

4) Wordle による視覚化

次に、コーパスデータの語彙全体を Wordle¹⁰⁾ でイメージ化する。作成した観光コーパスを Wordle により視覚的に表現することで、その全体像，出現頻度の多い語彙を把握する一助とした。この作成については HTC に 2 次的加工を加えた。統合されたコーパスデータからタグ情報を取り除いたテキストデータを作成した。その後，Wordle を用いて初期イメージを「Wordle イメージ 1」として作成した。文字の大きさは出現頻度を反映している。Hokkaido, Niseko, Hakodate はコーパスでは 300 回前後の頻度となっている。文字の大きな語は出現頻度の高い語であることから，HTC にはどのような記載が多いのかをより視覚的に想像することができる。

コーパスデータの Wordle イメージ 1



次に「Wordle イメージ 2」を作成した。機能語を取り除くとともに、コーパスデータ内に見られた Tel (電話番号), 地名 (Niseko, Otaru) などの観光コーパスそのものの分析には不要と考えられる語も個別に取り除いた。この加工は Wordle イメージ作成のためだけに行われたもので、Kensaku による語彙分析に影響を与えるものではない。

コーパスデータの Wordle イメージ 2



固有名詞を抜いたことで、より内容に焦点をあてて見ることができる。文字の大きさが出現頻度の高さを反映している。area, park, festival, enjoy, visitors, tour, resort など観光にふさわしい語が目立っている。これらに加え、snow, ice, ski, food, sea, nature 等で北海道らしさが出ている。また観光情報であろうと思われる語として、information, facilities, available が並ぶ。内容は Kensaku で作成された頻度別語彙リストと同じであるが、より視覚的に想像しやすい。onsen, ramen など日本語がそのまま使われている例もある。

5) JACET8000 との比較

Kensaku で HTC を『大学英語教育学会基本語リスト *JACET List of 8000 Basic Words*』¹¹⁾ (以後 JACET8000 と記述する) と比較した。JACET8000 の 8 レベルの語彙データと比較照合を行い、token と type のカバー率を出力した。また、それぞれのレベルでどのような語彙が出現したのかを見るために頻度別に上位 20 語を以下に示す。参考までに 4.1.2 2) に示した出現頻度第 50 位までの語彙を JACET8000 で仕分けした結果は注 12 に示す。¹²⁾

	JACET Lv1	JACET Lv2	JACET Lv3	JACET Lv4	JACET Lv5	JACET Lv6	JACET Lv7	JACET Lv8	Other
token	37481	4941	1875	1043	846	675	527	589	10321
% of token	64.29%	8.47%	3.21%	1.78%	1.45%	1.15%	0.90%	1.01%	17.70%
type	1037	676	437	281	223	206	172	146	2605
% of type	15.47%	10.08%	6.52%	4.19%	3.32%	3.07%	2.56%	2.17%	38.87%

Level_1]		Level_2]		Level_3]		Level_4]		Level_5]	
the	3520	lake	304	index	56	Convention	34	km	52
of	1830	festival	136	facility	55	mount	26	Tourism	40
and	1731	tourist	100	monument	46	exhibit	25	Mid	31
in	1394	map	91	drift	45	admission	24	designate	30
a	1296	museum	85	golf	38	accommodation	23	salmon	30
is	1023	locate	79	outdoor	29	addition	23	copyright	28
to	1015	restaurant	66	beer	29	fee	23	grind	25
it	472	association	66	cafe	28	resort	22	warehouse	23
from	467	tour	64	capacity	28	inquiry	17	harvest	17
as	429	visitor	58	outline	26	attraction	16	viewpoint	17
for	420	spot	57	promotion	25	spectacular	14	spectacle	15
you	400	surround	57	cycle	24	clinic	14	Salmon	14
on	395	site	54	memorial	21	freeze	14	mid	14
at	366	fresh	52	castle	21	inn	14	ferry	13
can	363	route	52	Golf	20	internal	14	owl	12
have	340	autumn	49	delicious	20	terminal	12	relaxation	12
with	338	hall	48	eagle	20	gallery	12	dedicate	11
by	300	bath	46	recommend	19	highlight	11	brew	10
be	298	available	45	summit	18	option	11	beautifully	9

Level_6]		Level_7]		Level_8]		Other]	
ski	47	crane	50	yen	131	Hakodate	389
marsh	45	blossom	23	cape	34	Hokkaido	257
cherry	33	waterfall	23	peninsula	29	Asahikawa	169
bloom	31	canoe	18	lighthouse	21	http	162
scenery	26	crab	17	herring	13	www	159
brewery	16	abundant	14	cabbage	12	hp	157
stroll	15	rental	14	Downtown	10	cho	133
enjoyable	14	trout	12	gorge	10	Onuma	126
aquarium	12	volcanic	10	affiliate	9	Nemuro	122
dentist	11	clan	9	cuisine	9	min	106
dome	11	firework	9	downtown	9	Kushiro	102
mint	11	fishery	8	railroad	8	Obihiro	101
seasonal	10	vicinity	7	violet	8	jp	87
hectare	9	volcano	7	shrimp	7	Chitose	83

stun	9	dam	6	skate	7	Akan	78
retrieve	8	grill	6	splendor	7	spt	78
civic	7	alley	5	squirrel	7	Otaru	77
convenience	7	inhabit	5	auto	6	Okhotsk	74
dine	7	recreational	5	backdrop	6	html	74

(2010Analysis_JACET8000_120125)

HTCの64.25%は、JACET8000の最も高頻度語である最初の1000語(レベル1)にレベル分けされている。JACET基本語改訂委員会によると、2000語程度の語彙力があれば最低限の読み書きはでき、上位2,000語で英字新聞の75%程度、平易な読み物であれば90%近くがカバーされるとある。レベル1の1,000語内の語彙が半分以上占めるHTCは平易な読み物であると予想される。

レベル1の頻出回数上位20語を見るとわかるように、ほとんどが機能語で中学生の早い時期に学習する語になっている。レベル2の語彙が全体の8.5%、それ以降は5%未満となる。レベル8までに指定されている8,000語の外にある語を*Kensaku*では「Other」として仕分けする。このOtherに含まれたのは、17.7%で10,354語あった。これらの多くは、北海道の地名を表わす固有名詞であった。

学生が北海道の紹介、自分の故郷の紹介を英語で表現する場合、どのような語彙が必要になるかと考えると、まずレベル1の1,000語を必ず身につけなければならない。Appendixの表を見てわかるとおり中学、高校で必ず遭遇する機能語、英語の基本語が多い。

次に、北海道という地域性が出ている語彙は、レベル2以降のグループに多く見られると想像できる。出現頻度は少ないが、その一語があることにより北海道という地域性が色濃くでる語彙が高レベルの語彙の中にあるはずである。レベル2から8の語彙を観察すると、すべてが内容語である。レベルが上がるにつれて、徐々に出現頻度も落ちていく。その中で、学生のライティングに役に立つ語彙として、一般的な英語の使用状況(Generalなfrequency)とは異なるような地域性(Locality)が反映されている語彙を別個に教える必要があると考え、このHTCを土台に語彙教材を作成した。

5. 教材作成

まず手始めに、日本について発信したり地域を語るために必要とされる語彙、地域紹介に必要な基本的な語彙教材作成を目指すことになった。*Kensaku*で作成した使用頻度数を基本とした語彙リストを利用して作成した。

5.1. 教材作成のための語彙リスト作成手順

1) 異なり語の総数は6,703語である。その中から使用頻度順位1750位まで(使用頻度数が5回まで)の語の中から動詞及び形容詞と思われるものを抽出した。名詞ではなく動詞・形容詞を選

んだ理由としては、まず学生の自己表現の力を伸ばすためには、より応用の可能性の高い事項を教示したいと考えたからである。名詞の場合は、表現したい事項の分だけ多くの語が必要となる半面、和英辞典などで調べて比較的容易に目的の語を探することができる。一方、動詞は英文の中心であり、動詞を中心として幾つかの文を学ぶことにより、様々な表現の応用可能性が広がる。また動詞に関しては、文法・語法の面でも個々の語について学ぶべき事項も多い。同様に形容詞も、その有効な使用によって表現の可能性を大いに広げることになる語であり、応用の可能性が高い。以上のような理由で、動詞・形容詞を中心に教材作成を行うこととした。

2) その中から、地域紹介の際に使いそうな動詞 13 語、形容詞 20 語を選択した。選択の根拠は、サンプルの中に含まれそうな語句、教師が教えたい語句、学生にぜひ使えるようになってほしい動詞と形容詞を主観で選んだ。

動詞

	Words	JACET8000 Level	Frequency
1	enjoy	1	159
2	locate	2	79
3	offer	1	66
4	experience	1	62
5	surround	2	57
6	View	1	51
7	include	1	42
8	designate	5	30
9	produce	1	30
10	provide	1	23
11	recommend	3	19
12	require	1	17
13	establish	2	14

(100724_CorpusDataConnected_Freq_lemma)

選んだ動詞を頻度別に並べた。JACET8000 のレベル別で見ると、ほとんどがレベル 1 であった。一番レベルの高い語は designate でレベル 5 であった。その他は、recommend がレベル 3、surround と establish がレベル 2 であった。

形容詞

	Words	JACET8000 Level	Frequency
1	local	1	60
2	various	1	13
3	available	2	45
4	traditional	1	27
5	unique	2	26
6	memorial	3	21
7	average	1	19
8	seasonal	6	10
9	enjoyable	6	14
10	spectacular	4	14
11	abundant	7	14
12	magnificent	3	11
13	volcanic	7	10
14	convenient	3	7
15	typical	2	8
16	panoramic	Other	9
17	historic	2	9
18	agricultural	3	8
19	industrial	2	6
20	annual	2	7

(100724_CorpusDataConnected_Freq_lemma)

形容詞は動詞より頻出頻度は低かったが、多様なものが出てきた。

local, various, available, annual, abundant, agricultural, industrial は地域紹介の基本語として使えると考えた。北海道の自然や文化を紹介すると出てきそうな以下の語も取り入れた。traditional, unique, seasonal, spectacular, magnificent, volcanic, panoramic

雄大な自然を記述する時に使われるであろう, spectacular, magnificent, panoramic のうち、最後の panoramic が Other に含まれていた。JACET8000 は現代英語の諸相に対応すべく作られたリストである。(JACET8000, p.121) このため各地域の特性を表現する時に必要な語彙は多く含まれていないことがある。このような語彙は、レベルが高い語彙として位置づけられていたとしても、教育に関わるものが自ら収集して学生に積極的に教えていかなければならない。

5.2. 観光コーパス例文集の作成

HTC から抽出した動詞と形容詞に意味と例文をつけ『観光コーパス動詞例文集』『観光コーパス形容詞例文集』を作成した。例文は HTC から *Kensaku* の KWIC 機能を利用して直接テキストデータに当たり選んだ。13 語の動詞については 1 語に対して 5 個の例文を、20 語の形容詞については各 3 つの例文を用意した。以下に、作成した教材の一部を例として示す。

designate 指定する，指名する <be designated as ～ に指定される>
<p>The Chishima cherry trees in Nemuro are more than 100 years old, and it has been designated as the Trees of Nemuro. 根室の千島桜は（樹齢）100年以上で，根室の木に<u>指定</u>されている。</p> <p>There are many kinds of birds that are designated as a natural monument and a rare species. 天然記念物や希少種として<u>指定</u>されているたくさんの種類の鳥がいる。</p> <p>Kushiro Shitsugen-Akan-Mashu Highway was designated as Scenic Byway. 釧路湿原－阿寒－摩周ハイウェイはシーニック・バイウェイ*（景勝間道）に指定されている。</p> <p>*シーニック・バイウェイ：◆1989年の米国シーニック・バイウェイ法に基づき，景観性や歴史性にすぐれており，観光を活性させるための一連の施策が適用される公道とその周囲の文化遺産</p> <p>Red Crested Cranes are designated as national special natural monument. 丹頂鶴は国の特別天然記念物に<u>指定</u>されている。</p> <p>The marsh itself is an off-limits area, being designated as a special zone of the national park. 湿原は立ち入り禁止区域で，国立公園の特別区に指定されている。</p>
enjoy 楽しむ，楽しく過ごす，満喫する，味わう <enjoy + 事，物，食べ物，～ing すること>
<p>Visitors can enjoy hot springs, camping, sail boating and more. 訪れた人々は温泉，キャンプ，ボート等を<u>満喫</u>できる。</p> <p>A sea kayak is a great way to fully enjoy the natural splendor of Shiretoko. シーカヤックは知床の自然の素晴らしさを十分<u>楽しむ</u>のによい方法である。</p> <p>Visitors can enjoy Sake tasting and seasonal foods in the restaurant and shop. 訪れた人々はレストランで利き酒をしたり，旬の食材を<u>味わ</u>える。</p> <p>You can enjoy camping in the forest near Tomisato Lake. 富里湖近くの森でキャンプを<u>して楽しむ</u>ことができる。</p> <p>Other activities to be enjoyed in Onuma Park include canoeing, tennis, golf, fishing and camping. 大沼公園で<u>楽しめる</u>他の活動としては，カヌー，テニス，ゴルフ，釣り，そしてキャンプがある。</p>

available (物が) 利用・使用できる, 入手できる, 得られる

Horse trekking courses are **available** and are very popular with beginners who want to try horse riding for the first time.

ホース・トレッキングのコースが利用でき, 初めて乗馬を試したい初心者に人気があります。

The variety of exciting sports is **available** in Chitose and Lake Shikotsu.

様々な面白いスポーツが千歳や支笏湖で利用できます。

Mineral water is **available** in shops and hotels in downtown Obihiro.

ミネラル・ウォーターが帯広市外のお店やホテルで入手できます。

panoramic 全景の, 広大な, 全体を見渡す

The top floor offers a splendid **panoramic** view of the city.

最上階からは, 街のすばらしい全景を眺めることができます。

The **panoramic** view of Tokachi from the scenic viewpoint is impressive.

その展望ポイントからの十勝の広大な眺望は感動的です。

Along these scenic routes, you can witness **panoramic** views of Hokkaido.

これらの景色の良い観光ルートに沿って, あなたは北海道の広大な景色を目の当たりにすることができます。

local 地元の, その土地の, 現地の, 各駅停車の

Seasonal delicacies such as seafood and fresh **local** produce are available at a good price. 海産物や新鮮な地元の農産物といった四季折々のおいしいものが, 手頃なお値段で手に入ります。

※ “delicacies < delicacy” は, ここでは「美味, 珍味」の意味。“produce” は名詞で「農産物」。

The rainbow trout and cherry salmon that thrive in the calm, pure waters feature in many of the delicious **local** specialties of Kamikawa.

静かな清流に育つニジマスやサクラマスは, 上川の美味しい地元特産品の中でも目玉となるものです。

The school was closed in 1974, and Mrs. Watanabe, a **local** resident, continued to feed The Japanese Cranes.

その学校は1974年に閉校になりましたが, 当地の住人であった渡邊さんが丹頂鶴に餌を与え続けたのです。

6. 結 語

本論では、北海道で学ぶ大学生が自分たちのこと、自分たちが生まれ育ち、あるいは今自分たちが暮らす地域のことを表現することによって、英語での自己表現の経験を積み、発信的なコミュニケーション能力の育成につながることを目指してきた。そのためにも、北海道の実際の公的な英語での発信情報から Hokkaido Tourism Corpus を作成し、コーパスに基づく研究の手法を援用して、語彙的な研究を行った。

大学生を対象とした学習語彙の研究や、特にそれらを語彙表のような形で学生に提供する試みは、高校生や中学生と比較して十分とは言い難い。また、園田（1996）が述べるように、語彙表の選定や作成の過程が詳細に開示されることはほとんど例がない。本論は、大学生の学習目標の手掛かりとなるべき基礎データの作成過程を詳細に述べると共に、北海道という地域を、英語の語彙という視点から改めて考察した。

次に、実際に学生に与える自己表現の教材の提示を行った。本研究は語彙分野を基礎としたものであるが、学生の自己表現の応用可能性を重視し、動詞・形容詞を中心とした例文を含む教材を作成した。語彙表のように語を単独で提示するのではなく、HTC から採取した実際の生きた例文を提示することで、北海道で学ぶ学生がより身近な自己表現の機会を得て欲しいと考えた。

今後は、本論で例示したような教材を用いて実際に授業を展開し、そこで学生がどのような学習行動を行い、またどのような反応や感想、評価を示すのか等について、更に研究を進めたい。

資料

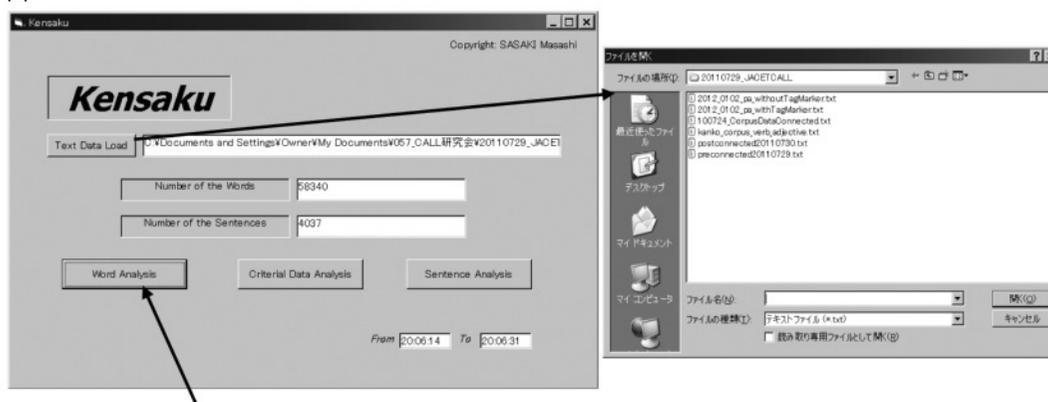
コンコーダンスー *Kensaku* による分析例

作成したコーパスから頻度順の語彙表を作成したり、KWIC によって特定語彙を含む表現(文)を検索するために本稿の著者の1人が開発したコンコーダンスー*Kensaku*¹³⁾を用いた。ここでは本稿での分析のための活用の具体例を示す。

1. 頻度順の語彙表

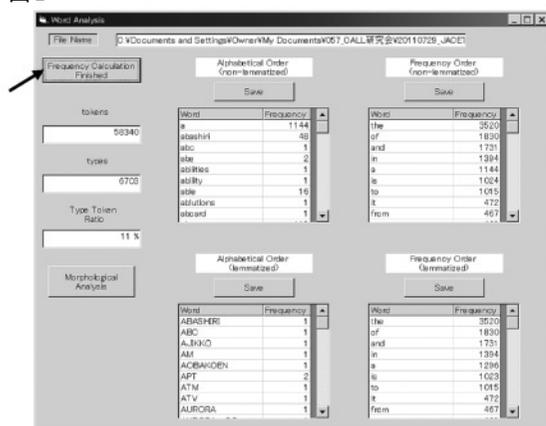
Kensaku を起動し次の画面(図1)で Text Data Load ボタンをクリックしコーパス・データを読み込むと総語数とセンテンス数が計算され表示される。

図1



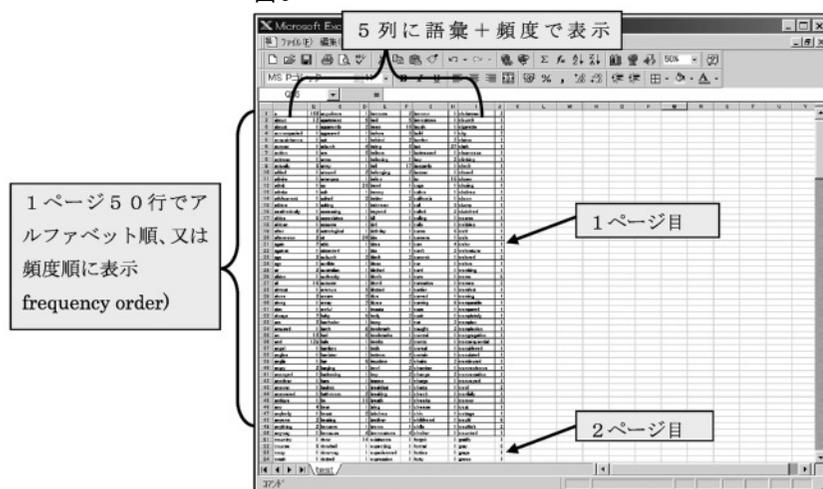
さらに、Word Analysis ボタンをクリックすると次のように頻度を計算するフォーム(図2)が現れるので Frequency Calculation ボタンをクリックすると計算を開始する。データ量が多い場合には、若干時間がかかるが、レンマ化されたデータとレンマ化されないデータをそれぞれアルファベット順、頻度順に計算する。

図 2



計算されたデータはマイクロソフト・エクセルなどで読み込むことの出来る csv ファイル形式で保存することが出来る。さらに、保存したファイルを再びエクセルで開くと図3のように、1ページ当たり縦50行、横5列で出力されるので、そのまま罫線を引くことで語彙表を作成できる。本稿ではこのように作成した語彙表頻度を元に動詞と形容詞を抽出したり、学生の作文の分析を行った。

図 3



2. KWIC による例文の抽出

語彙表から頻度を元に抽出した動詞と形容詞から例文集を作成するために、*Kensaku* の KWIC 機能を使用した。図1に示したフォーム上にある Sentence Analysis ボタンをクリックすると図

4のようなフォームが現れる。

このフォームには、全部で5つボタンが表示されるが、ここでは、センテンス中の1語を検索するため、KWIC_A_Searchというボタンをクリックする。この操作によって、例えば、動詞 include を含んだセンテンスだけを抽出しようとする場合、原形の他に includes / included / including が活用形として考えられるが、これらを含んだ表現を抽出することができる。KWIC_[A]_Searchというボタンは、活用形を含まない include のみを含んだ表現を抽出するが、KWIC_A_Search ボタンは include の活用形に共通の includ という文字列を入力すると、図5のようにセンテンス単位で表示する¹⁴⁾。

図4

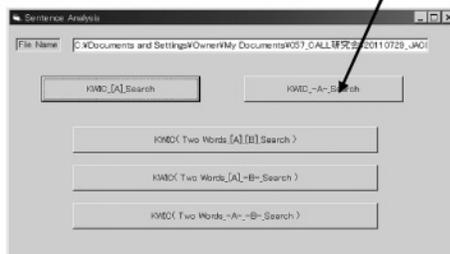
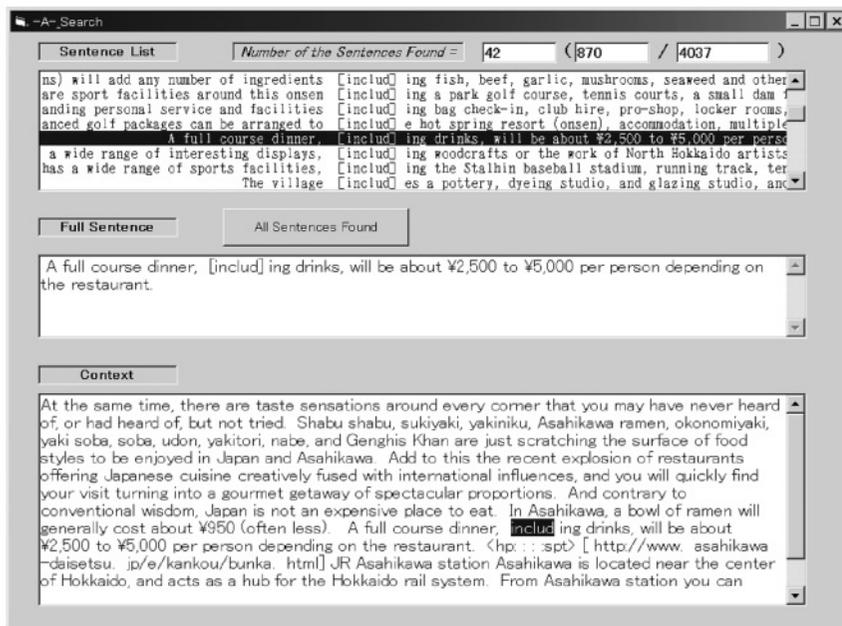
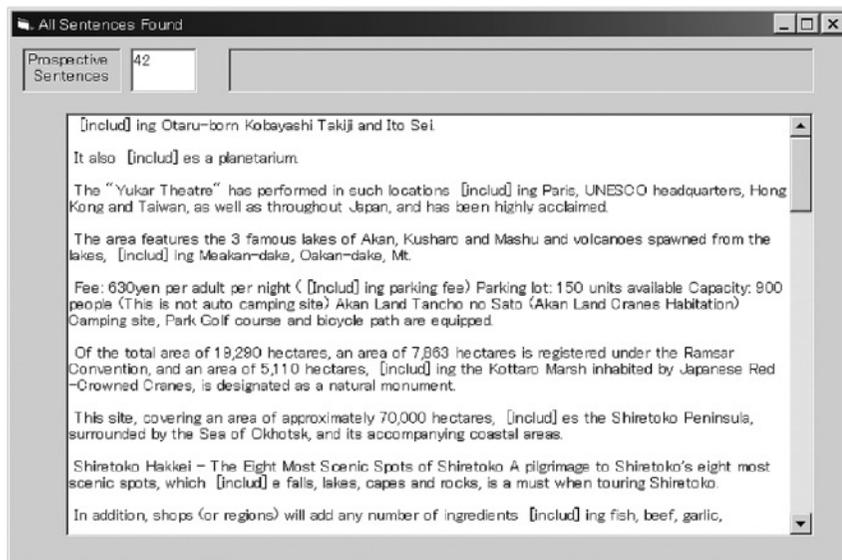


図5



また、該当のセンテンスのどれかにマウス・クリックでフォーカスをかけると図5の下欄にコンテキストが示されるとともに、All Sentences Found ボタンを押すと、図6のように該当する全てのセンテンスがコピー&ペースト可能なデータとして表示される。

図 6



学生指導用の例文はこれらからふさわしいものを選択した¹⁵⁾。

注

本稿は大学英語教育学会 (JACET) 北海道支部 CALL 研究会が第50回 JACET 国際大会 (2011年, 福岡市, 西南学院大学) で発表した内容をまとめたものである。

上野 之江* (北海学園大学, 教授), 尾田 智彦** (札幌大学, 教授), 佐々木 勝志*** (武蔵女子短期大学, 教授), 田中 洋也**** (北海学園大学, 講師), 森越 京子***** (北星学園大学短期大学部, 教授)

1) TOEIC の公式 Web ページには, 以下のような記述がある。

TOEIC テストを開発・制作する非営利テスト開発機関である Educational Testing Service (ETS : 米国ニュージャージー州プリンストン) はこれまで国際的な環境におけるコミュニケーションで使われる英語について数々の調査・検証を重ねてきました。その中で, 英語の利用が職場や日常生活の場でますます拡大していること, それに伴いスピーキングとライティングという能動的な能力を直接的に測定・評価する必要性が高まっている現状を認識し, 2つのテスト (Speaking と Writing のテスト) の開発に着手しました。

<http://www.toEIC.or.jp/toEIC/about/what/index.html> (2011年12月17日アクセス)

2) 上野他 (1999a), 上野他 (1999b), Yukie Ueno et al. (2000)

3) <http://daieikyo.jp/aetp/> (2011年12月18日アクセス)

4) 五十嵐昭人『英語で巡る日本の世界遺産』(南雲堂), 河原俊昭・榎木蘭哲也・岡戸浩子『観光英語で日本発見』(英宝社), 河原俊昭『観光英語で日本案内』(英宝社)

5) Air travel, Accommodations, Travel-related firms & associations, Tourism Academic の4分野。

6) *Kensaku* の詳しい説明, 機能, 使用方法は後半に記述する。

7) 「異なり語数」とは, 同じ単語が何度用いられていてもこれを一語とし, 全体で異なる単語がいくつあるかをかぞえた数。

8) Type-token Ratio の略。Lexical density (語彙密度) を示す数値。総語彙数 (token) の中の異なる語 (type) の比率をしめす数値。テキスト, 文章の難易度を比較する時に使われる。通常パーセントで示され, パーセントが高いほど語彙密度が高いとされる。

TTR = 異なり語数 / 総語数 × 100 の公式から計算される。(例) 29 異なり語数 / 57 総語数 × 100 = 50.88 となる。以下に定義の原文を示す。

a measure of the ratio of different words to the total number of words in a text, sometimes used as a measure of the difficulty of a passage or text. Lexical density (= type-token ratio) is normally expressed as a percentage and is calculated by a formula:

Lexical density = number of separate words / total number of words in the text × 100

For example, the lexical density of this definition is:

29 separate words / 57 total words × 100 = 50.88

(Richard & Schmilt, 2002, p.305)

9) 今回比較したテキストの出典は以下の通り。

Daily Yomiuri : <http://www.yomiuri.co.jp/dy/> (2012年2月10日, 11日アクセス)

TOEIC Part 6&7 : TOEIC 練習問題

Nyuushi_reading : 大学入試問題

VOA Special English : <http://www.voanews.com/learningenglish/home/>

(2012年2月10日, 11日アクセス)

ALL_Univ_A_5723_Words : 1999年作成大学生のE-mail コーパスより抽出

Graded reader_First time in England : <http://www.btinternet.com/~ted.power/first.html>

（2012年2月10日アクセス）

10) *Wordle* (<http://www.wordle.net/>) は、語の集合が雲の形状に表示される“word cloud”をテキストから作成するウェブアプリケーションである。2008年にIBMの研究者、Jonathan Feinbergによって個人的に開発された。テキスト内のより高頻度の語が、よりフォントサイズが大きくなるよう設定されており、a, theなどの一般的な語を取り除いた状態で表示できる。一般的な、タグクラウドを作成するアプリケーション（Tag Crowd, The Tag Cloud等）との違いは、フォントの種類、レイアウト、色などの視覚情報をユーザの選択により、より効果的に表示できることである。

直感的な操作性と視覚効果による表現力を持つことで、教育者をはじめとして、主要新聞紙などのマスコミ関係者、研究者、一般ユーザなど、幅広く使用されている。主要な使用者は、学校の教員などの教育関係者である（Viegas, Wattenberg, & Feinberg, 2009）。

使用方法も、綴り方や作文の教育、政党支持者による政治的なブログ記事の比較、個人的な手紙や考えの視覚化、Tシャツのデザイン等と多岐に渡っている。最近では、外国語ライティング指導などの教育手段としての活用（Baralt, Pennestri, & Selvandin, 2011）や研究の補助手段としての活用も報告されている（McNaught & Lam, 2010）。Viegas 他(2009)。その活用の広がりには、Jenkins(2006)が定義する「参加型文化（participatory culture）」によるものであると指摘している。）

11) 『*JACET8000 Basic Words*』は大学英語教育学会基本語改訂委員会が2003年に発表した8,000語の語彙リストである。以後日本の英語教育現場、研究に幅広く利用されている。各レベルの記述は以下のように記載されている。

Level 1（順位1000位まで）

中学校の英語教科書などに頻出する基本的な単語です。一般の英文の約70%以上は、このレベルの単語で書かれています。英語学習の最も早い段階で身につけたい単語です。

Level 2（順位1001～2000位）

高校初級レベルの単語です。このレベルまでの単語で、英字新聞の75%、平易な読み物であれば90%近くをカバーします。英検ではおおよそ準2級のレベルに対応します。

Level 3（順位2001～3000位）

高等学校の英語教科書レベルの単語です。大学入試センター試験は、おおよそこのレベルの単語までで作成されています。英検の2級に合格するには、このレベルをクリアする必要があります。仕事や研究などで特に英語を必要としている人はもちろんですが、一般の社会人も教養として身につけておきたいレベルの単語です。

Level 4〔順位3001～4000位〕

大学受験レベル、及び大学一般供給の初級レベルに相当します。日本人学習者で、英語の単語力があるかどうか問われるのは、このレベル以上の単語をしっているかどうかで決まります。

Level 5〔順位4001～5000位〕

難関大学受験レベル、及び大学一般教養に相当します。英検準1級合格のためには、このレベルの単語を習得する必要があります。TOEICテストでは、おおよそ400点から500点前後が目安となります。

Level 6〔順位5001～6000位〕

英語専門としない大学生やビジネスマンが目指すべきレベルです。英検では準1級、TOEICテストでは600点レベルに相当します。

Level 7〔順位6001～7000位〕

英語専攻の大学生や仕事で英語を使うビジネスマンの到達目標とするレベルです。このレベルがクリアできれば、英検1級やTOEICテストでは95%以上の単語をカバーしていることになります。

Level 8〔順位7001～8000位〕

日本人英語学習者の一般的な単語学習の最終目標です。この目標に到達できれば、あとは関連領域の専門用語を増やすだけです。

以上のレベル別順位表のほか別に別表として「Plus250」があります。これは国名、曜日名、数詞名などを順位をつけずに一覧表にした250語です。これはもっとも基本的な単語なので、Level1といっしょに使う事ができます。『JACET8000 英単語』(pp.5-6)

- 12) 4.2. 2)に示した出現頻度第50位までの語彙をJACET8000で仕分けすると Lake, festival がレベル2, yen がレベル7, 残りはすべてレベル1に含まれる語であった。

Level 1	Level 2	Level 3, 4, 5, 6	Level 7	Level 8	Other
Park City Area Spring Station Ice Minute View (n.v)* Enjoy Take See Hot Japanese **	Lake festival		yen		

* JACET8000 基本語彙表では、動詞、名詞一緒に表記している。

** レベル1を補う250語に入る重要語。

- 13) 本稿の著者である JACET Hokkaido CALL SIG (JACET Hokkaido Special Interest Group on CALL) は、10年ほど前に、“Corpus-Based Analyses of E-mail by Japanese College Students”を『大学英语教育学会 紀要』第32号に寄稿したが、そのプロジェクトの必要から既に学習者用に語彙表と concordance 又は KWIC (キーワード索引) 機能をもつツールとして本稿の筆者の一人が開発していた試作ソフトをテキストの語彙レベルを評価するために JACET4000 との照合が可能となるように改良を行い、ソフト名を *Kensaku Ver 2.0* として使用した。当時この種のソフトとしては DOS ベースで作動する grep, Windows ベースの Wordsmith や Monoconc などがあった。しかし、grep の場合はパターン・マッチングによる KWIC 機能が主であるため語彙表作成などができず、また DOS レベルで、パラメーターを入力して扱うにはそれなりの慣れを必要としていた。また、Wordsmith や Monoconc の場合は Windows ベースで作動し GUI によって操作できるため使い勝手は良いが、様々な機能があってそれをある程度知って使いこなすための手間や購入、インストールなどの煩雑さを考えたとき、当面の「限られた機能」が実現可能であれば試作ソフトを改良して使用することが合理的と判断したのである。今回のリサーチではこの *Kensaku* の最新バージョン(関西大学の染谷泰正氏の lemma 化データにより改良)を用いた。
- 14) 既に図1の説明において、コーパス・データを読み込む際に、総センテンス数を計算していることを示したが、それは、コーパス・データをピリオド(.), クエスチョン・マーク(?), エクスクラメーション・マーク(!)をマーカーとしてセンテンスに切り分けたものをデータ化し、その数を計算したものであり、ここでの操作はこのデータを活用するようになっている。
- 15) *Kensaku* のもう一つ大きな機能として JACET8000 と 8 レベルの語彙データとの比較照合を行い、token および type のカバー率を出力および csv ファイル形式で保存することができる。今回のリサーチで作成した観光コーパスをこの機能で分析すると次のように出力される。

	JACET 1	JACET 2	JACET 3	JACET 4	JACET 5	JACET 6	JACET 7	JACET 8	Other
token	37487	4041	1878	1042	846	675	527	559	10354
% of token	64.25%	6.98%	3.21%	1.78%	1.45%	1.15%	0.9%	1%	17.14%
type	1007	676	437	281	223	206	172	146	2607
% of type	15.47%	10.08%	6.51%	4.18%	3.32%	3.07%	2.56%	2.17%	38.89%

No.	JACET 1	Frequ.	JACET 2	Frequ.	JACET 3	Frequ.	JACET 4	Frequ.	JACET 5	Frequ.	JACET 6	Frequ.
1	Actually	1	Agriculture	4	Agricultural	1	Administrative	6	Collaboration	1	Ample	
2	Almost	1	Attraction	1	Architecture	1	Approximately	2	Fitness	2	Aviation	
3	Also	9	Civil	1	Artificial	1	Availability	5	MM	1	Civic	
4	Although	5	Comfortable	1	DELICIOUS	1	CM	1	Orthodox	8	Conversation	
5	American	2	Cultural	3	Deer	4	Commerce	3	Photography	1	Edible	
6	April	59	Data	1	Definitely	1	Convention	34	Privacy	1	FAX	
7	Ade	1	Defence	4	Delicious	1	Currently	1	Saltman	14	Fax	
8	August	38	Debate	2	Electrical	2	E	9	Tourism	40	Goodwill	
9	Australia	3	EAST	1	Fantastic	1	Imperial	1	Vivid	1	Livestock	
10	BEAUTIFUL	1	East	10	Furniture	1	Locally	1	aboard	1	Ma	
11	BEST	1	Eastern	12	Furberness	1	M	6	accumulate	1	Municipal	
12	Beautiful	2	Effective	5	GOLF	1	Maximum	3	ache	1	Par	
13	Best	2	Equipment	1	GoF	20	Minimum	1	actively	1	Parade	
14	Britain	1	Excellent	1	Harbor	1	P	3	adjacent	2	Parade	
15	British	5	Financial	1	Heritage	14	Reasonably	1	administrator	1	Scenery	
16	C	15	Fun	1	H	3	Recreation	2	aesthetic	3	Seasons	
17	Canada	6	Generally	2	Leisure	2	Strictly	1	amusement	4	advent	
18	Canadian	1	Golden	1	Library	4	acceptable	1	annually	1	alcoholic	
19	Catholic	3	Gray	1	Magnificent	1	accessible	1	apology	1	assembly	
20	Center	40	Historical	1	Museum	2	accommodate	2	appreciation	3	artistic	
21	Central	3	Industrial	1	Mutual	3	accommodation	1	array	3	aspirin	
22	Certain	1	Literature	2	Mysterious	1	addition	23	avarian	2	archaeology	

参考文献

- Baralt, M., Pennestri, S., & Selvandin, M. (2011). Using Wordles to teaching foreign language writing. *Language Learning & Technology*, 15(2), 12-22.
- Chujo, K, Utiyama, M, & Oghigian, K. (2006) Selecting Level-Specific Kyoto Tourism Vocabulary Using Statistical Measures. *New Aspects of English Language Teaching and Learning* (pp.126-138). Taipei: Crane Publishing Company Ltd.
- Fujita, R. & Tsushima, T. (2010). Toward Creating a Specialized Vocabulary List for Tourism Majors: Analysis of its Profile and Receptive Knowledge Among University Students. *JACET Journal*, 51, 1-13.
- Jenkins, H. (2006). *Convergence culture: Where old and new media collide*. New York University Press.
- Kobayakawa, M. (2011). Analyzing Writing Tasks in Japanese High School English Textbooks: English I, II, and Writing. *JALT Journal*, 33(1), 27-48.
- McNaught, C. & Lam, P. (2010). Using Wordle as a supplementary research tool. *The Qualitative Report*, 15(3), 630-643.
- Richards, J. C. & Schmidt, R. (2002). *Longman Dictionary of Language Teaching and Applied Linguistics, 3rd Edition*. Harlow, UK: Pearson Education limited.
- Ueno, Y., Hayasaka, K., Sasaki, M., Yoshida, M., Abe, A., ODA, T., & Yokoyama, K. (2000). Corpus-Based Analyses of E-mail by Japanese College Students. *JACET Bulletin*, 32, 137-149.
- Viegas, F. B., Wattenberg, M., & Feinberg, J. (2009). Participatory visualization with Wordle. *Visualization and Computer Graphics, Vol.15*, pp. 1137-1144.
- 相澤一美・石川慎一郎・村田年（編）。（2005）。『JACET8000 英単語』。東京：桐原書店。
- 石川慎一郎。（2008）。『英語コースと言語教育：データとしてのテキスト』。東京：大修館書店。
- 上野之江・阿部晃夫・尾田智彦・早坂慶子・吉田翠。（1999 a）。「語学教育に生かす E-メール交流(1)」。『北海学園大学学園論集』第 100 号，215-259。
- 上野之江・阿部晃夫・尾田智彦・早坂慶子・吉田翠。（1999 b）。「語学教育に生かす E-メール交流(2)」。『北海学園大学学園論集』第 101 号，145-183。
- 齋藤俊雄・中村純作・赤野一郎。（1998）。『英語コース言語学：基礎の実践』。東京：研究社。
- 園田勝英。（1996）。『大学生用英語語彙表のための基礎的研究』。言語文化部研究報告叢書 7。

- 北海道大学.
- 大学英語教育学会 (JACET) 基本語改訂委員会 (編). (2003). 『大学英語教育学会基本語リスト JACET8000』. 大学英語教育学会 (JACET).
- 北海道経済部観光振興課. (2004). 『北海道観光の概況 (H16)』.
- 北海道経済部観光振興課. (2005). 『北海道外客来訪促進計画: 国際観光推進プログラム—ようこそ北海道—』.
- 文部科学省. (2010). 『高等学校学習指導要領解説外国語編・英語編』. 東京: 開隆堂.