

コンピュータ用語のデータベース作成と特徴の分析

—留学生の情報活用能力の養成を目指して—

後藤寛樹・深澤のぞみ・濱田美和

要 旨

留学生が日本語環境でコンピュータを十分に使いこなすためには、コンピュータ画面に現れるコンピュータ用語や指示の文を自分で読んで判断しながら操作していかなければならない。

筆者らは、留学生のコンピュータ操作を支援するための用語集を作成することを目指して、コンピュータ用語のデータベースを作成し、使われている語彙表現の特徴の分析を行った。本稿では、データベースの作成方法を紹介し、その上で、コンピュータ画面に現れる語彙表現の特徴として、できるだけ短く表示しようとする傾向があること、語・句・文、そして長い複合語や省略のある形などが多種多様に混在していることを挙げ、さらに語彙レベルの分析からは、コンピュータ用語には留学生が苦手としている漢語や外来語の占める割合が多いこと、日本語学習の過程では中級以上のレベルと見なされる語彙が多くを占めていることなどを述べる。

【キーワード】 コンピュータ用語、留学生、日本語環境でのコンピュータ操作、データベース

1 はじめに

IT革命に伴い、大学でも学生向けに情報活用能力養成のための授業が開講されることが一般的になってきた。大学で学習や研究を進めていくのに、情報活用能力は欠かせないものである。この情報活用能力には多くの要素が含まれるが、とりわけコンピュータを使いこなす能力は、学習・研究活動だけでなく、学生生活のいろいろな局面と関わりがあり、大学生活全般で重要な位置を占めている。留学生もこの例外ではなく、彼等の留学生生活の成否にコンピュータを使いこなす能力の有無は大きく関係している。

コンピュータは今や全世界に普及し、留学生も来日前に何らかの形でコンピュータを使ったことがあるケースがほとんどである。しかし、日本語環境でのコンピュータ操作経験を有する学生は少なく、そこには大きな困難が伴う。大学で開講されている情報処理の授業などに出席しても、日本人学生向けの内容にはついていけないという声を聞くこともよくある。

したがって、留学生が日本の大学で情報活用能力を発揮して学習・研究を進めていくためには、何らかの支援が必要であると思われる。そこで、筆者らは留学生がコンピュータを使用する際の問題点などについて基礎調査を行い、それをもとにした教材の開発を行ってきた。さらにそれに続く段階として、留学生がより深くコンピュータを使いこなすことができるよう、参照しやすいコンピュータ用語集の作成を試みている。本稿では、これまでの調査や作成した教材の概要を簡単に紹介し、コンピュータ用語集の基礎となるデータベース作成の必要性や方法、そしてデータベースを分析して明らかになったことを論じていく。

1.1 留学生がコンピュータを使用する際の問題点

筆者らはこれまで、留学生が日本語環境でコンピュータを使用する際に問題となる点や、よく見られる誤用などについて調査を行った⁽¹⁾。その過程で明らかになったのは、次の2点である。

(1) コンピュータ用語の難解さ

コンピュータを操作するには、画面上に出てくる用語や指示などを理解しなければならないが、

それらの用語や指示には、専門語であるために通常の日本語教育では扱われないようなことばや、難解な漢字で書かれた語、さらには留学生が苦手とするカタカナで書かれた語が多い。

(2) 日本語入力の問題

留学生が日本語環境のコンピュータで文書を入力していく場合、ローマ字入力を行うことが多いが、日本語のローマ字表記を習得していないため、長音や促音、撥音、拗音などの入力ミスをしてしまい、その後の漢字変換がうまくいかないことがよくある。留学生は手書きでもこれらの間違いをすることはよくあるが、重要なのは、コンピュータ上での入力の場合、最初の入力が間違っていると、次のステップである漢字変換の際に、正しい漢字が探し当てられなかったり、他の漢字に変換されてしまったりすることがあり、元の文意と全く違うものになってしまうなど、問題が顕著になりやすいということである。

1.2 留学生のためのコンピュータ入門テキスト

筆者らは、1.1で述べた調査結果を踏まえながら、留学生のためのコンピュータ入門テキスト『留学生のための日本語コンピュータ』⁽²⁾を開発した。このテキストは、簡単なコンピュータ用語を導入し、日本語入力の方法を練習問題とともに提示した上で、よく用いられるソフトウェア（Microsoft社のWord, Excel, PowerPointなど）の使い方を留学生が課題を作成しながら習得できるようにデザインされたものである。現在、富山大学で行われている留学生のためのコンピュータ授業の中で活用しており、一定の効果を上げている⁽³⁾。

1.3 コンピュータ用語調査の必要性

留学生はコンピュータの使い方を一通り学んでも、まだ決して十分というわけではない。大学での学習・研究活動の中では、単なるワープロ文書の作成やインターネットでの情報検索といった比較的単純な操作だけではなく、いろいろなソフトウェアや操作を組み合わせながら、もっと専門的な使い方をしなければならないこともあるからである。その場合、留学生は複雑な操作を、自分でコンピュータ画面に出てくる項目や指示のことばを読んで、判断し、進めていかなければならない。そこで、筆者らは、留学生にとって調べやすいコンピュータ用語集を作成し、留学生が自分でコンピュータの複雑な操作をしなければならないときの助けとなればと考えるに至った。

しかし、コンピュータ分野は新しく発展してきたものであり、また進歩の速度も速いことなどから、コンピュータ用語についての研究はまだあまり進んでいない。ましてや教育のためのコンピュータ用語集やその基礎となる研究は、ほとんど見受けられない状況である。コンピュータ画面に出てくる語彙表現を見ると、一見しただけでも、極めて多様だということがわかる。難解な専門語から日常的な一般語までが同じ画面上に現れるし、また前述の調査結果にもあったように、通常の日本語教育の中では扱われないことばが多いこと⁽⁴⁾、漢字やカタカナで書かれた語が多いことなど、様々な特徴が見て取れる。したがって、留学生のコンピュータ用語集を作成するためには、この多種多様なコンピュータ用語の実体を把握し、分析を行った上で整理することが必要となる。そこで、筆者らはまず、コンピュータ用語のデータベースを作成して、それを分析することにした。

1.4 本稿で扱うコンピュータ用語の範囲

コンピュータ用語データベースについて論ずる前に、本稿で扱うコンピュータ用語の範囲について述べておきたい。

コンピュータ用語ということばは、通常コンピュータやアプリケーションの概念や原理を説明するための基礎用語、機能や操作を説明するための用語、コンピュータ関連用語など、広い範囲を表している。

しかし、筆者らの最終目的は、留学生のコンピュータ操作を支援するための用語集作成ということであるため、ここではひとまずコンピュータの画面上に現れる語彙表現だけを扱い、コンピュータ用語と呼ぶことにする。

第2節では、これらのコンピュータ用語のデータベース作成の基本方針を述べ、さらに第3節でこのデータベースの分析によって明らかになったコンピュータ用語の特徴について論じる。

2 データベース作成の基本方針

コンピュータの画面上には、コンピュータ用語辞典に掲載されているような、いわゆるコンピュータ用語から日本語教育でも扱われるような日常的語彙まで、実に様々な語句が現れる。日本人であれば、それらの語句が専門的な語句かどうかを判断したり、操作に大きく関わってくるような語彙かそうでないかの見当をつけたりすることは容易であろうが、既に筆者らの一連の研究^⑤でも分析してきた通り、留学生、特に非漢字圏からの学生にとって、画面上に現れる漢字あるいはカタカナで書かれた語などの文字情報は非常にわかりづらく、それがコンピュータを操作する上での困難さを生み出している。データベースを作成するにあたって、これらの語をどこまで、どのように取り扱うかを検討した。なお、データベースの作成には、FileMaker社のFileMaker Pro 5.5を使用した。以下、2.1および2.2で、データベース作成の基本方針について詳しく述べ、2.3および2.4で作成したデータベースについて述べる。

2.1 データベース化の単位

データベース作成にあたって、まず検討したのは、コンピュータ画面上に現れる語をどのような単位でデータベース化していくかということである。コンピュータの画面に出てくる語句には、「終了」「保存」「ファイル」などのように語単位で表示されるものもあれば、「別名で保存」「ヘッダーとフッター」のように句単位で表示されるものや「テキストのドラッグアンドドロップ編集を行う」「保存時にプロパティを確認する」のように文単位で表示されるものもある。コンピュータ画面上の語句を分析する際に、これらを一括りに語単位に分割してデータベース化しただけでは、その特徴は見えてこない。また、「スタイルシート」のようにそれを構成するそれぞれの語の意味がわかっている、全体が何を指すのか理解していないと実際の操作に結びつかないものもある。逆に、語・句・文の各単位のまま入力しただけでも、語彙レベルの特徴を見出すことは難しく、コンピュータを操作する上での基本的語彙として、最低限押さえておかなければならない語彙を抽出するのも困難だろう。そこで、画面上の語句を語・句・文のそれぞれの単位のまま抜き出してデータベース化した「まるごとデータベース」と、それらを語単位に分割してデータベース化した「一語データベース」の2種類のデータベースを作成することにした。

2.2 データベース化の対象と範囲

次に検討しなければならなかったのは、どのアプリケーションに出てくる語句を、どこまでデータベース化するかということである。コンピュータには様々なソフトウェアが組み込まれているが、留学生が実際に専門分野の学習や研究でコンピュータを利用するときのことを考えると、利用頻度の少ないソフトに使用される語彙を調査することはあまり意味がないと言える。そこで、データベース化の対象とするソフトは、一般的に使用されているワープロ、表計算、プレゼンテーションソフトとEメールソフト、Webブラウザに限ることにした。また、OSや日本語入力に関するものはすべてのソフトに共通して関わってくるものなので、これもデータベース化の対象に加えることにした。なお、今回の調査ではまずMacintoshを対象としてデータベースの作成を行った。ワープロ、表計算、プレゼンテーションソフトについては、広く利用されていると思われ、筆者らが開発した留学生向けコンピュータ教材^⑥でも取り上げている、Microsoft社のWord、Excel、PowerPoint（バージョンはいずれも2001）を扱うことにし、

データベース化の対象ソフトウェアを決定した次の段階として、これらのソフトで使われる語をどの範囲までデータベース化するかが問題となる。例えば、図1のように、Wordのメニューバー「挿入」を選択すると、「改ページ」「ページ番号」などのメニューが表示され、その中の「定型句」を開くと、「定型句」「新規」などのサブメニューが表示される。場合によってはこのようなサブメニューが何層にも階層化されている。また、メニューによっては例えば、図1の「ページ番号」を選択すると図2のウィンドウが表示されるといったように、操作の詳細を決定するウィンドウが現れ、その詳細ウィンドウの中にさらに次のウィンドウへと続くボタンがあったり、ドロップダウンリストボックスで具体的な操作項目を決定する箇所があった

- (1) メニューバーの用語および各メニューを開いたときに表示されるメニュー項目（メニュー項目からサブメニューへと続く場合はそれも含める）はすべて対象とする。
- (2) メニュー項目から別の操作ウィンドウが開く場合は、その最初に関開くウィンドウに出てくる語句までを対象とする。
- (3) 操作ウィンドウ内にタブで数枚のシートが含まれている場合は、そのすべてのシートをデータベース化の範囲に含める。
- (4) 操作ウィンドウ内の語句は、コマンドボタン、チェックボックスやオプションボタンで選択する操作の語句と、それらの見出しとなっている語句および指示や説明文のみを対象とし、リストボックスに含まれる語句は対象としない。

挿入 書式 フォント ツール 整理 ウィンドウ 作

改ページ ▶

ページ番号...

日付と時刻...

定型句 ▶

フィールド...

記号と特殊文字...

コメント

脚注...

図表番号...

クロスリファレンス...

索引と目次...

図 ▶

HTML オブジェクト ▶

テキストボックス ▶

ムービー...

ファイル...

オブジェクト...

ブックマーク...

ハイパーリンク... 98K

定型句...
新規...
挨拶文ウィザード...
挨拶文 ▶
起し言葉 ▶
結び言葉 ▶
ヘッダー/フッター
件名行
会社名 ▶
経路 ▶
署名 ▶

ページ番号

位置:
ページの下

横位置:
右

☒ 最初のページにページ番号を挿入する

プレビュー

書式

キャンセル

OK

— 6 —

ンドウ内の語句であるが、この操作ウィンドウは、基準(2)にしたがってメニュー項目を選択したときに最初に開くウィンドウのみを対象とした。例えば、図5で二重線の吹出しで示した「用紙設定」ボタンをクリックすると、さらに別の操作ウィンドウが開くが、そのウィンドウ内の語句は対象外ということになる。また、(e)のタブをクリックすると、同じウィンドウ内に別の操作設定シートが表示されるが、これらのシートについては基準(3)にしたがい、すべてのタブシートを開いた最初のシートまでを対象とした。(f)の値一覧にはチェックボックスやオプションボタンの形式で表示されるものが該当し、リストボックスで表示されるものは対象外とした。なお、実際のデータベース上では(a)~(g)に「その他」を加えた8項目を、それぞれの語句の出現場所を示す分類項目として用いている。

このようにして、データベース作成の方針を定め、2種類のデータベースの作成にとりかかったわけだが、ここでそれぞれのデータベースの特徴を簡単に説明しておきたい。

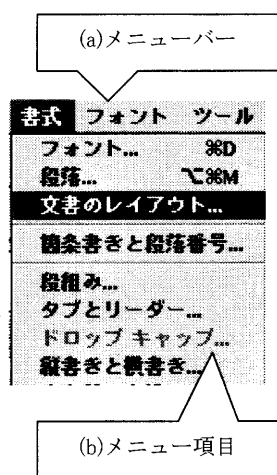


図3

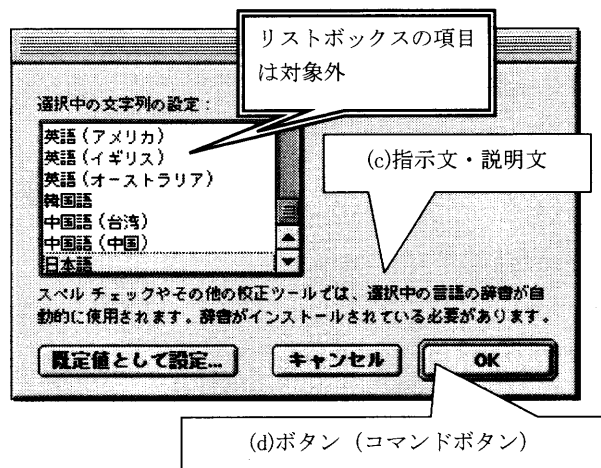


図4

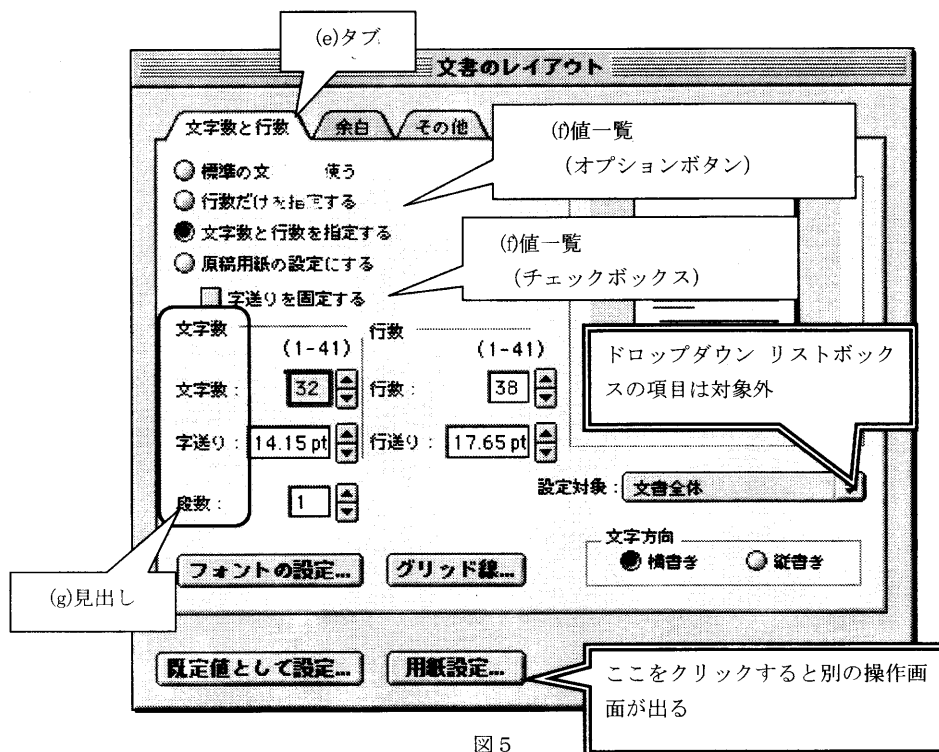


図5

2.3 「まるごとデータベース」

「まるごとデータベース」はコンピュータの画面上に現れる語句を、語・句・文のレベルに関係なく、そのままの形で入力したものである。このデータベースでは該当する語句の出現ソフトおよび場所、ショートカットキーの有無、語構成を入力項目として挙げた。図6は「まるごとデータベース」のカードの例である。

まず出現ソフトおよび場所であるが、語句によってある特定のソフトでしか用いられないようなものがあったり、特定の場所に頻繁に出現するものがあったりするので、それを分析できるように入力項目として取り上げた。例えば、「CC」「送受信」はEntourageにしか用いられない語句であり、「集計」「昇順」はExcelにしか出てこない。なお、今後Windowsについても同様の調査をする予定なので、Macintosh、Windowsのどちらに使用される語かをチェックする欄も設けてある。

図6 まるごとデータベース

ショートカットキーの有無は、この調査が留学生のためのコンピュータ用語集の作成を最終的な目標として始められたものであり、用語集を作成する段階で、ショートカットキーがあるものについては、それもあわせて掲載できることを狙いとして取り上げた項目である。ショートカットキーが存在する操作については、そのショートカットキー操作もあわせて記載した。

語構成については、既に触れたように、コンピュータの画面上には語・句・文の様々なレベルの語句が出現するので、どの単位かを分類できるように項目を立てた。

2.4 「一語データベース」

「一語データベース」は「まるごとデータベース」に入力した語句を語単位に分割して入力したものである。このデータベースでは、該当する語の出現頻度、日本語教育における難易度のレベル、語種、出現ソフト・場所、辞書での記載の有無の各項目を扱った。図7は「一語データベース」のカードの例である。

まず語の出現頻度であるが、これは実際のコンピュータ画面でどの語がより多く用いられているのかを調べるために取り上げた項目である。出現頻度の高い語については、コンピュータを操作する上で必要な語だと言えるだろう。

語のレベルについては、その語が留学生にとって日本語を学習する上でも重要度の高い語かどうかを調べるために取り上げた。

また、漢字表記のものについては、その漢字のレベルもあわせて考慮した。レベルの設定には、『1万語語彙分類集』⁷⁾を基準にした。

語種については、和語、漢語、外来語、混種語のうち、どのような語種の語彙が多く用いられるのかを調べるための項目である。

出現ソフト・場所については、「まるごとデータベース」と同様に、語によってある特定のソフトでしか用いられないものがあったり、特定の場所に頻繁に出現するものがあったりするので、それを分析できるように入力項目として取り上げた。

辞書での記載については、「まるごとデータベース」に収録されているものを、どこまで語の単位に分割するかと大きく関わってくる。例えば、「頭文字」「禁則処理」はそれぞれ「頭」と「文字」、「禁則」と「処

図7 一語データベース

理」からなる複合語であると考えられるが、どちらも『広辞苑』（第5版）⁽⁸⁾に記載されている。このような語は全体で1語（1項目）と見なし、『広辞苑』に記載があることがわかるように、カード上でチェックを入れた。また、コンピュータの専門用語の中には、「ドラッグアンドドロップ」のように全体で1つの操作を表すものや、「コントロールパネル」のように全体でコンピュータ内の特定の部分を指すものなどがある。これらは全体を理解していなければ実際の操作に結びつかないものと考えられ、これを構成要素に分割して考えることは、用語集作成という目的を考えるとあまり意味がないと思われるので、全体で1語とみなした。ただし、例えば別の箇所「ドラッグ」が単独で用いられている場合は、別の項目として「ドラッグ」のカードを作成している。このとき、どこまでが一まとまりか考える際に参考にしたのが、インターネット上で公開されている、『アスキーデジタル用語辞典』⁽⁹⁾である。この用語辞典に掲載のある語については、全体で1語とみなし、『広辞苑』と同様に記載があったことがわかるよう、カード上でチェックを入れた。また、Microsoft社のソフトに関する用語については、『アスキーデジタル用語辞典』ではあまり扱われていなかったため、『Microsoft単語帳』⁽¹⁰⁾もあわせて参考にした。

3 日本語のコンピュータ用語の特徴

第3節では、データベースから見えてくる日本語のコンピュータ用語の特徴を明らかにするとともに、日本語学習者にとってどのような点が難しいかについて述べたい。まず、3.1で「まるごとデータベース」に基づく分析結果、そして、3.2で「一語データベース」に基づく分析結果を見ていく。

3.1 「まるごとデータベース」の分析結果

「まるごとデータベース」は全部で5,418項目となった。2.3でコンピュータの画面上には語・句・文といった様々なレベルの語句が出現すると述べたが、具体的にどのようなものが出現するかを調べると、大きく分けて、表1のようなバリエーションがあることがわかった。

表1 コンピュータ画面に出現することばの語構成

語構成	例
一語	ウィンドウ、ログイン、HTML、分類、…個、解像度、片付ける、お気に入り
複合語 ⁽¹¹⁾	Webオプション、フォントサイズ、更新日時、二重取り消し線、移動先ブック名
NのN	データの取り出し、印刷の向き、英数字用のフォント、既定のファイルの場所
NとN	タブとリーダー、簡条書きと段落番号、線種とページ罫線と網かけの設定
N/N	ドライブイン/ドライブアウト、拡大/縮小率、マイナス/長音/ダッシュ
N N	脚注 ページの最後、作成者 ページ 日付
NまたはN	Netscape Communicator 4.5または4.7、シートの削除またはコピー
形容詞+N	新しいスライド、古いパスワード、アクティブな文字コード、高度な検索
連体修飾節+N	ペーストする形式、変化させるセル、印刷する範囲、最後にアクセスしたページ
N+格助詞	次へ、ファイルから、スキャナまたはカメラから、直前の動作から、左から右へ
N+のみ	この単語のみ、半角英字のみ、未読メッセージのみ、手動の更新のみ
N+について	Netscape6について、このメニューについて、バルーンヘルプについて
一部省略（助詞、スル、動詞）	プレビュー（を）隠す、形式を選択してペースト（する）、別名で保存（する）、下へ検索（する）、条件を詳しく（する）、ゴミ箱を空に（する）、後で投稿（する）、ファイルからサウンド（を）挿入する、すべて非表示（にする）
従属節のみ	新しいメッセージが届いたとき、開いているメッセージを移動/省略した後
一文	拡張子を追加する、インラインで日本語を入力する、大きいアイコンを使用する
二文以上	選択したシートを削除します。一度削除すると元に戻すことはできません。

この中で最も多いのは一語で、「まるごとデータベース」全体の半数近くを占めている。

一語に次いで多いのが複合語で全体の2割弱を占めている。その多くは2語で構成された語であるが、「コピー元ファイル名」「文字列表示範囲枠」のように、4語、5語で構成されたような語もある。特に「全画面表示」「文字列表示範囲枠」のように漢字ばかりの語や、「ディスクスペース」「ワークシートメニューバー」のようにカタカナばかりの語は、留学生が辞書を使ってことばの意味を調べたい場合、どこでどう区切っていいかさえないといった問題が生じやすい。

三番目に多いのが一文で、割合は12%となっている。ただし、一文と言っても、目的語と動詞だけの短いものが多く、4語以下の語で構成されている文が6割以上を占めている。二文以上のものを含めても13%にしかならず、コンピュータ用語には文単位のもの割合は非常に少ないと言える。また、一文の長い文、それから二文以上のもののほとんどは指示文・説明文であり、コンピュータ操作をする場合の重要性は低いと思われる。

その他には、名詞と名詞を「の」「と」「または」などを使って結びつけたもの、形容詞、副詞、連体修飾節で名詞を修飾したもの、名詞に格助詞、「～について」「～のみ」をつけたもの、従属節だけのものなどがあるが、共通して言えることは、動詞などは省略し、できるだけ短く表示しようとしていることである。

また、できるだけ短く表示するために、「プレビュー隠す」のように助詞を省略したり、「形式を選択してペースト」のように「する」を省略したりする傾向が見られる。省略の中でも特に「する」が省略されたものが多く、この中には、「すべてクリア」や「今すぐ送信」のようなものもある。このように、助詞などが省略された場合、省略されたことばを類推できるだけの日本語力が必要ということにもなり、留学生にとっては難しいものと考えられる。

以上から、コンピュータ用語の一つ目の特徴として、できるだけ短く表示しようとする傾向があることが挙げられる。これは、コンピュータ画面に表示する場合、その操作のためのことばはできるだけ短い方がスペースも取らないし、コンピュータを使う者にとっても瞬時に判断できたほうが操作時間の短縮につながるためと思われる。しかし、短く簡潔に表すことが留学生にとって常にわかりやすいとは限らない。先に述べたように、漢字やカタカナばかりで書かれた語をどこでどう切っていいかわからないといった問題の他に、文であれば、ある一語がわからなくても、その前後の文脈から意味を類推することもできるが、一語しか表示されていない場合は、その語がわからなければどうすることもできない。留学生のみならず、コンピュータ操作に不馴れな日本語母語話者にも同じことが言えるだろう。つまり、コンピュータ用語では、操作の鍵となる語を知っているかどうかということが非常に大切であると言える。

それから、二つ目の特徴として、多種多様な表現が混在していて、統一性がないことが挙げられる。例えば、名詞と名詞が並列関係にある場合は、並列助詞の「と」以外に、スラッシュやスペースを利用して並列関係を表しているものもある。他に、表1には載せなかったが、「ノート、配布資料、アウトライン」のように読点で区切ったものもある。さらに、「印刷」と「プリント」という同じ操作を表すことばがどちらも出現していたり、「インターネット」と「Internet」、「コントロール」と「CONTROL」のようなカタカナを使ったりアルファベットのまま表記したりといったように、使用するソフトによって、同一操作を表す用語が異なったり、中には同じソフト内でも違う表現が用いられていることもあり、留学生にとっての混乱の原因となる。

3.2 「一語データベース」の分析結果

「一語データベース」は、延べ語数で11,207語、異なり語数は1,410語となった。3.2では語種と『1万語語彙分類集』での日本語学習における難易度レベル設定を中心にコンピュータ用語の特徴を見てい

く。まず3.2.1で「一語データベース」全体の特徴を述べた後、3.2.2では高頻度の語に的を絞って見ていくことにしたい。

3.2.1 「一語データベース」全体の語種とレベル設定

異なり語数1,410語のうち、延べ語数で見ても、異なり語数で見ても、漢語と外来語の割合が高い。漢語が全体の4割強、外来語が3割強を占めている。留学生の苦手とする漢語と外来語がそのコンピュータ用語の中で重要な位置を占めていることがわかる。また、外来語と言ってもその中身も様々で、コンピュータ用語には、「HTML」「Startup」「JavaScript」のようなアルファベット表記のままの語や「Webページ」「ユーザーID」「GIFアニメーション」などアルファベットとカタカナが混ざった語が多く出現する。加えて、カタカナ表記の語の中にも、「コンフリクト」「ソルバー」「ディゾルブ」「スピリットワイプアウト」などのように単に英語をそのままカタカナで表示しただけのような語もあり、日本語母語話者でも理解しにくいことばが多い。

それでは、コンピュータ用語は具体的に日本語教育ではどのぐらいのレベルに設定されているのだろうか。『1万語語彙分類集』に基づくレベル設定の結果を表したのが表2である。『1万語語彙分類集』では、レベルA、B、C、Dの4つのレベルが設定され、それぞれ日本語能力試験1級、2級、3級、4級レベルとされている。

表2 レベル別の比率

	異なり語数(1,410語)		延べ語数(11,207語)	
レベルA (日能試1級)	122語	8.7%	1,406語	12.5%
レベルB (日能試2級)	376語	26.7%	3,916語	34.9%
レベルC (日能試3級)	132語	9.4%	954語	8.5%
レベルD (日能試4級)	104語	7.4%	812語	7.2%
設定なし	676語	48.0%	4,119語	36.8%

『1万語語彙分類集』で設定がない語の割合は、異なり語数では48.1%、延べ語数では36.8%となっていることから、コンピュータ画面に出てくる語の半分以上は、上級レベルの学生であれば理解可能な語ということになる。異なり語数ではレベルBのことばの割合が26.7%で、全体の4分の1を占めている。延べ語数で見ると、さらにレベルBの比率が35.0%と増えている。また、レベルAの比率も12.6%となり、全体の1割強を占めている。つまり、コンピュータ画面に出てくることばを理解するには、コンピュータの専門語のみならず、中上級レベルの語彙も必要となると言える。そのため、日本語学習者は、コンピュータ画面に出ていることばがわからない場合は、コンピュータ用語辞典と一般の国語辞典との両方で調べなければならず、彼等にとってかなりの負担となる。留学生にとって日本語環境でコンピュータ操作をする場合には、コンピュータの専門語を学ぶだけでなく、中上級レベルの語彙力も要求されることがわかる。

コンピュータの専門語の大半はレベルの設定なしに分類され、この割合は大きい。ただ逆に、設定なしの語がすべてコンピュータの専門語であるとは言えない。設定なしの語には、「制御」「最新」「配色」「枠内」「カウント」「ズーム」などの一般語も多く含まれる。

3.2.2 高頻度の語の語種とレベル設定

次に高頻度の語に焦点を当てて同じように語種とレベル設定を見ていきたい。表3は「一語データベース」の中での高頻度の語、上位30語を順に並べたものである。

表3を見てまず気付くのが、漢語と外来語の占める割合の高さである。30語中、漢語が16語、外来語が8語、和語が6語となっている。一般的な書きことばに関する語彙調査では高頻度の語で和語の占める割合が高いが⁽¹²⁾、コンピュータ画面に出てくることばではこのことは当てはまらない。その理由としては、画面に出てくることばは、先に述べたように文単位で出現するものが少ないことが考えられる。

表3 コンピュータ画面に現れる高頻度の語

順位	見出し語	頻度	順位	見出し語	頻度	順位	見出し語	頻度
1	表示	267	11	列	85	20	メッセージ	68
2	キャンセル	227	12	リンク	84	22	オプション	65
3	設定	196	13	使用	83	22	作成	65
4	OK	168	13	入力	83	22	…名	65
5	文字	122	15	検索	82	22	～てください	65
6	保存	108	15	ページ	82	26	行	62
7	削除	105	17	ファイル	79	27	指定	59
8	名前	103	18	プレビュー	77	27	追加	59
9	変更	98	19	すべて	76	29	修正	57
10	選択	86	20	する	68	30	開く	55

それでは、これらの語は、『1万語語彙分類集』ではどのぐらいのレベルに設定されているであろうか。

表4から、レベルDが2語、レベルCが2語、レベルBが14語、レベルAが6語、設定なしが6語となっており、初級レベルで理解できる語は4語しかなく、中級レベルでも18語しか理解できないということがわかる。高頻度の語に限って見ても、コンピュータ用語は、留学生、特に初級レベルの日本語学習者にとってはかなり難しい語彙であると言える。

表4 『1万語語彙分類集』に基づくレベル設定

順位	見出し語	レベル設定	順位	見出し語	レベル設定
1	表示	設定なし	15	ページ	レベルC
2	キャンセル	レベルB	17	ファイル	レベルA
3	設定	レベルA	18	プレビュー	設定なし
4	OK	レベルA	19	すべて	レベルB
5	文字	レベルA	20	する	レベルD
6	保存	レベルB	20	メッセージ	レベルA
7	削除	レベルB	22	オプション	設定なし
8	名前	レベルD	22	作成	レベルB
9	変更	レベルB	22	…名	レベルB
10	選択	レベルB	25	～てください	設定なし
11	列	レベルB	26	行	レベルB
12	リンク	設定なし	27	指定	レベルB
13	使用	レベルB	27	追加	レベルB
13	入力	レベルA	29	修正	レベルB
15	検索	設定なし	30	開く	レベルC

3.3 分析結果のまとめ

「まるごとデータベース」「一語データベース」、いずれの分析結果からも、留学生にとってコンピュータ用語は非常に難しいものであることがわかった。コンピュータの専門語だけでなく、中上級レベルの語彙力も必要とされること、そして、その語彙の多くは留学生の苦手とする漢語、外来語であること、さらに短い表現にするために、3語以上で構成された複合語、助詞や動詞を省略した語もあること、また、同じ操作でも違う語を使ったり、同じ語でもカタカナで表記したり、アルファベットで表記したりして統一されていないことなど、様々な問題がある。ただ、文単位のことばが少なく、一語、複合語などの語単位のことばが多いコンピュータ用語では、ある語を知っているかどうかでその操作ができるかどうかの鍵になるということについては、裏を返せば、ある鍵となる語を知っていれば、操作ができるということにもなる。

4 おわりに

コンピュータ用語のデータベースを作成する過程で、1種類のデータベースでは不十分だということがわかり、その結果、「まるごとデータベース」と「一語データベース」の二つを作成することになった。これは、日本語のコンピュータ用語が多種多様で、語彙表現の構成や語種、そして表記の点からかなり混沌としたものであることを反映している。したがって、留学生にとってコンピュータ画面に出現する語は、彼等が留学生活で接する日本語の中でも非常に難しいものの一つであると考えられる。

今回の分析結果より、留学生のためのコンピュータ用語集の開発の一つの指針としてコンピュータ操作の鍵となる語の抽出が必要であることがわかった。今後の作業としては、Windowsで使用される語についてもデータベースに取り入れ、コンピュータ操作で鍵となる語をさらに詳しく分析し、用語集の作成を行う予定である。

注

- (1) 後藤・深澤・濱田(2001)を参照のこと。
- (2) 深澤・濱田・後藤(2000)を参照されたい。
- (3) 富山大学留学生センターでは、センターで開講している大学院入学前予備教育日本語研修コースと、日韓共同理工系学部留学生プログラムの中で、このテキストを用いたコンピュータ授業を行っている。また、学部留学生を対象とした「日本事情」の授業の中でも、数回、留学生の情報活用能力養成のための授業および演習を行っており、部分的にこのテキストを用いている。詳しくは、富山大学留学生センター(2000, 2001a, b, 2002)を参照のこと。また、学部留学生に対する授業については、濱田・深澤・後藤(2001)を参照されたい。
- (4) 中山(2001)によると、新聞に多用されている外来語であるにもかかわらず日本語教科書の中で全く扱われていないものに「ホームページ」「ネット」「メール」があったという。コンピュータに関連した語の中でもこれらのものは極めて基本的なものであるが、日本語教育の中では重視されていないということがわかる。
- (5) 詳しくは、深澤・後藤(2000)、濱田・深澤・後藤(2001)、後藤・深澤・濱田(2001)を参照されたい。
- (6) 深澤・濱田・後藤(2000)を参照されたい。
- (7) 専門教育出版『日本語学力テスト』運営委員会編(1998)『一万語語彙分類集 改訂』専門教育出版
- (8) 新村出編(1998)『広辞苑(第5版)』岩波書店
- (9) 『アスキーデジタル用語辞典』はインターネット上で公開されている辞典で、URLは<http://yougo.ascii24.com/>である。
- (10) 『Microsoft単語帳』は、マイクロソフト社がオンラインサポートサービスで公開しているサポート技術情報の中の用語を説明したものである。
URLは<http://www.microsoft.com/japan/Terminology/>である。
- (11) 本稿では、一語データベース作成の際に、一語と認定した語が複数結合した語を複合語と呼ぶ。
- (12) 林大監修(1982)を参照のこと。

参考文献

- (1) 国立国語研究所 (1981)『専門語の諸問題』秀英出版
- (2) 国立国語研究所 (1984)『日本語教育指導参考書12 語彙の研究と教育(上)』大蔵省印刷局
- (3) 国立国語研究所 (1985)『日本語教育指導参考書13 語彙の研究と教育(下)』大蔵省印刷局
- (4) 後藤寛樹・深澤のぞみ・濱田美和 (2001)「留学生向けコンピュータ教材の開発とその使用」『日本語教育』110号, 日本語教育学会, pp.150-159
- (5) 富山大学留学生センター (2000)『日本語研修コース第1期報告書 らいちょう』
- (6) 富山大学留学生センター (2001a)『日本語研修コース第2期報告書 らいちょう』
- (7) 富山大学留学生センター (2001b)『日本語研修コース第3期報告書 らいちょう』
- (8) 富山大学留学生センター (2002)『日本語研修コース第4期報告書 らいちょう』
- (9) 中山恵利子 (2001)「日本語教科書の外来語と新聞の外来語」『日本語教育』109号, 日本語教育学会, pp.90-99
- (10) 濱田美和・深澤のぞみ・後藤寛樹 (2001)「留学生センターにおけるコンピュータ教育について」『富山大学総合情報処理センター広報』Vol.5, pp.24-32
- (11) 林大監修 (1982)『図説日本語』角川書店
- (12) 深澤のぞみ・後藤寛樹 (2000)「留学生に対するコンピュータ授業－専門への橋渡しとしての役割とその実践」専門日本語教育研究討論会資料
- (13) 深澤のぞみ・濱田美和・後藤寛樹 (2000)『留学生のための日本語コンピュータ』(独自開発教材)