

〔論 説〕

スマートフォン用ラテン語学習辞書の編纂

箕原辰夫

1. はじめに

ラテン語は、フランス語、ロマンス諸語といったその系譜に繋がる現代の言語だけではなく、ドイツ語や英語のようなゲルマン系の言語も含めて、西洋の言語の基礎をなしている。また、その発音方式は日本語に極めて近いものであり、多くの日本人は意識していないが、日本語のローマ字の綴りの方式はラテン語の綴りになっている。これは、英語圏の人々が日本人の氏名のローマ字表記を見て、まったく違う発音で名前を呼ばれる所以となっている。加えて、全世界的にみても、新しい学術的な用語などは、ラテン語の造語規則に基づいて作られている。現代まで活着しているラテン語について、日本ではあまり多くの教育がなされていない状況にある。英語やフランス語と同様に、そこに多大な影響を及ぼしているラテン語について知識や教養は、言語学的に見ても広く普及させる必要があると思われる。この研究は、潜在的に必要とされるそのようなラテン語教育を、狭い言語学の教室の中で行なうのではなく、もっと一般の人々に対して、開かれた形で、かつ自律的に学べるような形で提供できる言語学習のための辞書システムを編纂することを目的としている。そのために、このシステムの実装の対象をスマートフォンあるいはタブレット式のコンピュータに絞っている。また、学習に必要な辞書のデータの形態、提示の仕方についても考慮する。この辞書システムは、単なる検索のための辞書だけではなく、それを利用して持続的に学習者が学習していけるような教材としての位置を占めることを願い、実現していくものである。

辞書については、この紀要でもエジプト文字の入力システムを開発した際に、その必要性を実感してきた。翻訳で使うため、あるいは「かな漢字変換」のような入力支援システムを作成するため、あるいは広く学習用教材を作成するために必要な辞書の構造を考える必要がある。博士課程の学生のとくに、京都大学の長尾眞先生と岩波情報科学辞書のオンライン化の作業に従事したことがあるが、そのときより意味データベースの基幹情報としての辞書システムのモデリングを行ってきた。今後、シュメール、エジプトといった古代文字を使用する言語から、日本ではあまり知られていないタミル語、およびロマンス諸語の基盤となっているラテン語について、辞書を編纂していく予定にしている。そのためのデータのモデリングをいかに行なうか、外部スキーマであるビューをどのように形成するかについて、議論を行なう。

2. 辞書の選定

対象者は、学習者であり、作成する辞書は学習者が古典ラテン語あるいは現代ラテン語を読んだり、書いたりするために使いたい。特にラテン語で綴るための辞書として用意したいと考えている。そのような意味においては、ラテン語・英語の双方向からの単なる検索ができる辞書よりも、実際の使用例なども含めた総合的な辞書にしたいと考えている。特に、名詞や形容詞の格や、動詞の活用形について、検索の際に、各格や活用形から主たる形がわかるようにしたい。そのための基盤となる辞書データを探すことにした。これとは別に、単語列が網羅されているラテン語の入門書（「ラテン語入門」河底尚吾，泰流社，1985年など）から単語を抜き出して、手作業で辞書データを使用する方法もあるだろう。しかしながら、体系性を持たせるためにも、基盤となる辞書データは、フリーライセンスで探せる羅英・英羅辞典をベースにすることにした。なお、以降の本文では既存の辞書については、「辞典」という言葉を用いることにする。

iPhone などのスマートフォンのラテン語辞典としては、図1に示す Paul Hudson 氏の「Latin English Dictionary +」や、図2に示す Harry Schmidt 社の「Lexidium Latin Dictionary」を始めとして、10種類程度の辞典がある。いずれも有料になっている。これらの辞書は、いずれも1879年に出版された「Lewis and Short Latin Dictionary⁽¹⁾」を原典としている。また、それに加えて、以下に述べる William Whitaker 氏の「Words」のデータも採り入れている。

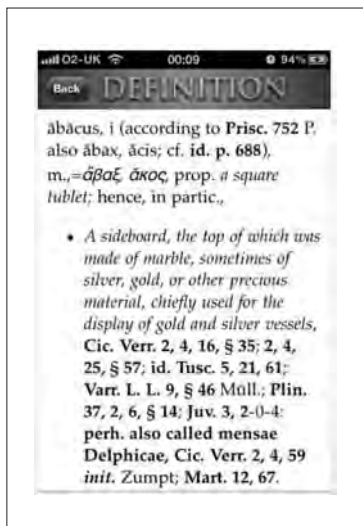


図1 Latin English Dictionary +



図2 Lexidium Latin Dictionary

(1) L&S と略されることが多い, A Latin Dictionary, Founded on Andrews' edition of Freund's Latin dictionary revised, enlarged, and in great part rewritten by Charlton T. Lewis, Ph.D. and Charles Short, LL.D. Oxford, Clarendon Press, 1879.

Web上にアップロードされている辞典の中には、原典からのスキャナによる読み取りからOCRでテキストに起こしたもので、形式を整えるためにさまざまな変換をしなければならないものもある。また、OCRからテキストを起こしたものは、当然ながら、文字の読み取りの間違いも含まれている。そのため、テキストとして非常に整理された羅英辞書を探すことにした。「Oxford Latin Dictionary⁽²⁾」は、古典ラテン語の定評のある辞典であるが、1968年版のPDF形式のデータは主に有償でダウンロードできるが、テキストデータが分離されていない。また、1968年以降の新版のものでは、母音の「u」と半母音の「v」および、母音の「i」と半母音「j」の区別がなくなっているとのことである。基本的には、母音と半母音は区別した方が入門としてはわかりやすいと思われる。Web上からライセンス的にもダウンロード可能で、純粋にテキストベースで整備されて用いることができるのは、「Lewis and Short Latin Dictionary」と「William Whitaker's Words」ぐらいになっている。

「Lewis and Short Latin Dictionary」については、TUFTS大学のPerseusデジタルライブラリのプロジェクトのWebページ⁽³⁾から単語ごと、あるいはオフラインソフト⁽⁴⁾などを介して一括してXMLの形式でダウンロードできる。また、archive.orgからは1891年にHarper & Brothers Publishersから出された新版のPDF版⁽⁵⁾もみることができる。Perseusプロジェクトでは、ギリシャ・ローマ文明を始め、アラビア文字で書かれた文書、ドイツ語で書かれた文書、19世紀の米国の文献、ルネッサンス期の文献のテキストデータを有している。このプロジェクトでは、Lewisが1890年に米国で出版した「An Elementary

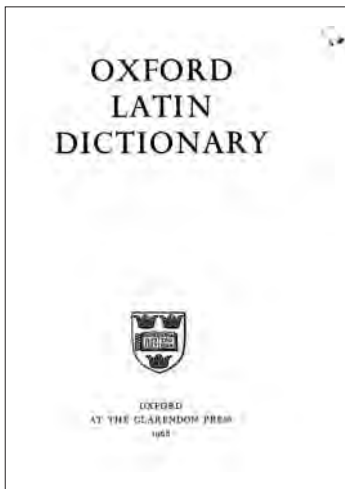


図3 Oxford Latin Dictionary, 1968

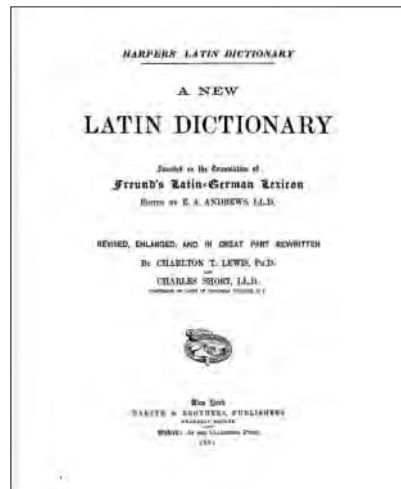


図4 Lewis and Short Latin Dictionary

(2) OLDと略される、First editor: Alexander Souter, Oxford Latin Dictionary, Oxford at the Clarendon Press, 1968.

(3) <http://www.perseus.tufts.edu/hopper/>

(4) Peter Heslin氏のDiogenes, <http://www.dur.ac.uk/p.j.heslin/Software/Diogenes/index.php>

(5) <http://www.archive.org/details/LatinDictionaryA>

Latin Dictionary」⁽⁶⁾ も収録しており、このテキストデータも一括して XML の形式でダウンロードすることが可能になっている。

「William Whitaker's Words」については、著者の Web ページ⁽⁷⁾があり、そこから辞書データのテキストや閲覧ソフトウェアである WORDS のダウンロードなどができるようになっている。テキストの解説についても、ノートルダム大学⁽⁸⁾や Classics Technology Center⁽⁹⁾を始めとする方々のラテン語についてのサイトで行なわれ、この Web ページへのリンクも張られている。ただし、この英羅辞典は、長音の表記はされていない。また、この辞典の特徴としては、現代まで用いられているラテン語の専門用語も列挙されている点がある。たとえば、放射能に関する用語も掲載されている。

英羅辞典については、archive.org 上に「A complete English-Latin dictionary⁽¹⁰⁾」などを始めとして、かなりのテキストデータがある。PDF 版でもオンライン版でも archive.org から読むことができる⁽¹¹⁾。archive.org 上の羅英・英羅辞典のテキストについては、OCR からテキストに起こしたために、純粋なデータとするには無理があり、今回の実装では見送ることにした。その他、archive.org には様々なラテン語の辞書のデータが置かれている⁽¹²⁾ので、今後の参考としたい。

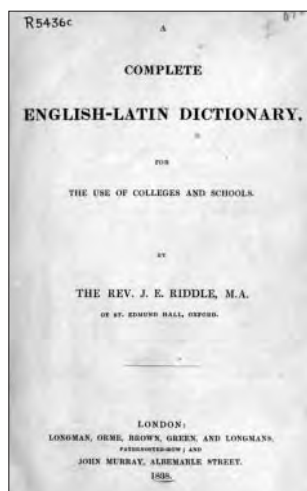


図5 A complete English-Latin dictionary, 1838

(6) Lewis, T. Charlton, An Elementary Latin Dictionary. New York, Cincinnati, and Chicago. American Book Company. 1890.

(7) Words と略される, William A. Whitaker, Latin-English Dictionary Program WORDS-Version 1.97FC, <http://users.erols.com/whitaker/words.htm>

(8) <http://archives.nd.edu/words.html>

(9) <http://ablemedia.com/ctcweb/showcase/index.html>

(10) Joseph Esmond Riddle, A complete English-Latin dictionary, for the use of colleges and schools, London: Longman, Orme, Brown, Green and Longmans, Paternoster-row; and John Murray, Albemarle street, 1838.

(11) A complete English-Latin dictionary の場合, <http://www.archive.org/details/completeenglishl00ridduoft>

(12) <http://www.archive.org/search.php?query=Latin%20Dictionary>

表1 テキストデータのあるラテン語の羅英辞典

辞典名	収録語数	内容	ファイル形式	サイズ
William Whitaker's Words	39225語	意味のみ	1行1単語	6.2MByte
Lewis and Short Latin Dictionary	51593語	長音・用例も	XML	84.5MByte
An Elementary Latin Dictionary	17581語	長音・用例も	XML	14.5MByte

テキストデータのあるラテン語の羅英辞典は、表1のようになる。表中の「XML」について若干の説明をしておく、XML形式⁽¹³⁾とは、マークアップ言語を通常のテキストデータの中に埋め込み、文章を構造化したテキスト形式のことであるが、この中で自身の構造化の定義も行なうことができ、その定義に従って、テキストを配列することができる。しかしながら、XMLを処理しながら84.5MBものデータ量のあるデータを読み込むのはスマートフォンでは重い。これを利用する場合は、もう少し簡易な形のフォーマットにしなければならない。

Mac OS X標準の辞書形式⁽¹⁴⁾は、まったく別個に定義されたXML形式になっているが、Perseusプロジェクトの辞書のXML形式を、Mac OS X標準の辞書形式に変換してしまえば、Mac OS X標準の辞書閲覧アプリケーション⁽¹⁵⁾で、閲覧することが可能となるだろう。さて、「Lewis and Short Latin Dictionary」を用いている幾つかのスマートフォン用のラテン語辞書では、アプリケーションのサイズが、20MBぐらいなので、ある程度XML形式を簡易な形にして読み込んでいと推測される。具体的には、長音や発音記号、あるいは引用されているギリシャ語などについての部分を残して、後はかなりXMLを簡略化してデータサイズを落としているものと思われる。一方、Wordsは、見出しの部分揃えるために、空白の文字コードが挿入されているので、これを除けば3.5MBぐらいになり、スマートフォン用の辞書データとしては手頃なサイズになる。発音記号や用例の部分は目をつぶれば、まずまず利用することができる辞書を構成することが可能となる。そのため、今回の実装では、まずは「Words」を対象にすることにした。

また、ラテン語は屈折語なので、格変化や活用形の種類は多い。名詞や形容詞については、男性・女性・中性の分類、および単数と複数があり、格としては主格・属格・与格・対格・従格・呼格の6格からなり、格変化の種類が多い。動詞については、未完了形と完了形の2つの相が現在・過去・未来の3時称に含まれており、直説法・接続法・命令法・不定法の4つの法および分詞・動名詞がある、加えて能動態と受動態の2つの態の区別が有り、単数と複数、および1人称・2人称・3人称の3つの人称がある。語尾変化としては、一つの動詞に140種類ぐらいある。ただし、不規則変化をする動詞は少ない。このような語形・語尾変化が著しい言語であるため、学習用の辞書としては、活用形からも引けるようにしたい。活用形としての語尾変化については、「Words」の付属ファイルの中に、語形・語尾が変化する品詞に関して、変化形ごとの規則的な変化の仕方の一覧や不規則動

(13) http://ja.wikipedia.org/wiki/Extensible_Markup_Language, 仕様は <http://www.w3.org/TR/REC-xml/>

(14) Apple Computer, Dictionary Source Schema for Dictionary Services XML. /Developer/Extras/Dictionary Development Kit/documents/DictionaryFormat.rtf, 2008.

(15) Dictionary.app

詞の変化に記述されたファイルがある。また、学習者の文法知識の習得の観点からすれば、ラテン語の文法書などに列挙されているように、名詞や形容詞の格変化の種類や動詞の活用形の変化の種類によって、グループ化するようなことを行なわれるのが良いだろう。「Words」の辞書のテキストデータの中には、品詞の項目に、活用形の分類などについても記述されている。加えて、それぞれの単語には分類コードが添えられている。これは、使用時代の区分・何を表すのか・使用された地域・使用頻度・語の原典となる辞典の5つの局面から語の分類をするものになっている。これらの分類コードは、単語を様々な範疇にしたがって分類して表示するときに役に立つだろう。

また、学習辞書には不可欠な引用文のデータベースも入れたいと考えている。逆に引用文で使われている各言葉から辞書の項目が引けるようにする必要もあるだろう。TUFTS大学のPerseusプロジェクトでは、膨大なラテン語のテキストデータが用意されている⁽¹⁶⁾。ただし、Perseusプロジェクト本体が所管しているデータは、128MByteほどであるから、パーソナルコンピュータでは全文をディスクに入れれば、オフラインでの検索の対象となりうる。ただし、スマートフォンに落とすためには、先に述べたことと同じように、詳細なXML形式で記述されたテキストデータのある程度XML形式を簡易な形にする必要があるだろう。これは、今後の実装の課題項目としたい。また、先に記したように、archive.orgにも様々なラテン語のテキストデータが用意されている。同じように、ラテン語のコーパスを蒐集したWebサイト⁽¹⁷⁾などもある。更に、ラテン語の入門サイトの中にはラテン語で書かれた原典から有名な文を抜き出し、文法的な解説を行なっているものがある⁽¹⁸⁾。今後は、このようなところからのデータも参照したいように思う。

3. データベースとしての専門辞書

スマートフォン用の学習辞書を編纂するのがこの研究の目的であるため、単なる検索だけができる辞書というのは最終的には想定していない。そこでは、学習者が文法的な知識を補完できる内容や例文などを通して、学習者が本として読めるものを求めている。そのためには、単なる辞書のデータに留まらず、文法的な知識あるいは例文データについても、ある程度まとまった形で格納されており、学習者の検索・閲覧の対象になっていることが望ましい。それを踏まえて、以下に学習辞書としてのデータモデルとインターフェースを提案する。

3.1. 学習辞書としてのデータモデル

学習辞書として、辞書の項目のデータモデルと文例の項目のデータモデルについて、表2に記述する。これらのデータモデルは、実際にはXMLの形で記述されている。Mac OS X 10.4以降の標準の辞書のXMLで定義されているもの⁽¹⁹⁾は、このような何を表すかについての分類はなされておらず、XHTML⁽²⁰⁾の拡張に終わっている。ただし、辞書上の

(16) <http://www.perseus.tufts.edu/hopper/collections>

(17) <http://www.forumromanum.org/literature/index.html>

(18) <http://www.kitashirakawa.jp/taro/latin5.html>

(19) 前掲、/Developer/Extras/Dictionary Development Kit/documents/Dictionary Format.rtf

各言語の標記については、Common Locale Data Repository (CLDR⁽²¹⁾) に準拠することが求められている。Perseus プロジェクトの方は、Mac OS X の辞書構造よりもだいたい項目の内容から構造化されているが、文章としての構造化の意味合いが強い。様々な側面から検索するためには、ここで提案したような辞書項目として構造化する必要がある。

表 2 辞書の項目のデータモデル

辞書項目	
ID=原形 (タイトル)	
発音表記	長音記号を含む
品詞の種類	
変化の種類	名詞・形容詞・動詞の場合
代表形	動詞の場合, 男性一人称主格, 原形, 男性一人称過去形
使用地域	
使用年代	
分類細分	概念語・対象語
英語の意味	
日本語の意味	
言葉の出典	
使用例へのリンク	
関連語へのリンク	反対語・上位概念・下位概念
文例項目	
文	
日本語の意味	
文中の単語を要素として羅列したものとその単語についての文法的な知識	
文の出典	

「Words」に含まれている分類コードから、使用地域・仕様年代・分類細分化なども自動的に得ることができる。しかしながら、言葉の出典や使用例のリンクなどについては、「Lewis and Short」の方から持ってくる必要がある。このデータモデルには入れていないが、語形・語尾変化の規則、および少数の不規則な変化をする語については、その変化の一覧についてのデータベースも必要になってくる。これも「Words」で提供されているが、「Words」では更に1つの語について、すべての語形・語尾変化を列挙したものをデータベースにしておいてある。すべての語形・語尾変化から検索できるようにするためには、これらのデータを併せてもっておかなければならない。

文例項目の中の文中の単語の要素の羅列は、1つの集合になっている。関係データベースなどで表現する場合は、正規化する必要があるだろう。文法的な知識については、単純に辞書の該当する語にリンクを張るだけでなく、その格・活用形の種類なども上記の語

(20) <http://www.w3.org/TR/2010/REC-xhtml-basic-20101123/>

(21) <http://www.unicode.org/cldr/>

形・語尾変化に関するデータベースと連動する形で、解説を加えたものを追記する必要がある。また、出典の一覧もデータベースにする必要があるだろう。

3.2. 学習辞書としてのインターフェース

学習辞書のインターフェースとしては、単純な検索ができるものだけでなく、様々な分類から辞書を閲覧できる必要があるだろう。表3に考えられる閲覧方法を示すビューの幾つかを提示し、その検索の仕様を、上記のデータモデルを関係データベースのスキーマと仮定し、データベース言語 SQL による問い合わせの記述の形で併記しておく。

表3 学習辞書としてのビュー

・ 羅英辞典としてのビュー	<pre>select 辞書項目.* from 辞書項目 where 代表形 like "%指定された綴り%"</pre>
・ 英羅辞典としてのビュー	<pre>select 辞書項目.* from 辞書項目 where 英語の意味 like "%指定された綴り%"</pre>
・ 和羅辞書としてのビュー	<pre>select 辞書項目.* from 辞書項目 where 日本語の意味 like "%指定された綴り%"</pre>
・ 格変化の種類ごとのビュー (名詞・形容詞)	<pre>select 辞書項目.* from 辞書項目 where 品詞 = "指定された品詞 (名詞または形容詞)" and 変化の種類 = "指定された変化の種類"</pre>
・ 活用形の種類ごとのビュー (動詞)	<pre>select 辞書項目.* from 辞書項目 where 品詞 = "動詞" and 変化の種類 = "指定された変化の種類"</pre>
・ 分類ごとのビュー (名詞・形容詞・動詞)	<pre>select 辞書項目.* from 辞書項目 where 分類細分 = "指定された分類"</pre>
・ 出典ごとのビュー	<pre>select 辞書項目.* from 辞書項目 where 言葉の出典 = "指定された出典"</pre>
・ 出典ごとの文例	<pre>select 文例項目.* from 文例項目 where 文の出典 = "指定された出典"</pre>

厳密には、「ごとの」という場合は、SQL では GROUP BY 句を用いてグループ化した検索をするのであるが、ここでは記述が込み入ってしまうので、単純に「指定された」値を用いての条件検索の形で記述した。また、出典などについては、関係データベースで考える場合は、本来は出典の一覧を表す表との JOIN 操作を行なう必要がある。分類ごとのビューでは、概念語（抽象語）と具体的な対象を示す語を分類されており、しかも具体的な対象の場合は、動植物や日常のどのような対象に対しての言葉なのかについての細分類がされているので、それを明示する。語学の入門の Web サイトでは、そのような分類を示すものもある。

なお、通常の辞書の場合は、スペースの分量を小さくするために、かなりの省略語が用いられている。たとえば、男性名詞については、「M」または「m」の一文字で示したり、引用についても、キケロなどは「Cic.」になっている。参考にされている他の辞書につい

でも、「Lewis and Short Latin Dictionary」は、「L-S」もしくは「L&S」が用いられ、「Oxford Latin Dictionary」は「OLD」のような形で略されていることが多い。専門家にはイメージできるこのような省略語は、学習者にとっては難解な印象を与えるだけなので、なるべく省略しない形での表記に直すことようにしたいと考えている。

4. 実装

4.1. 和文の意味の自動生成

日本人の学習者のために辞書を編纂するためには、どうしても意味のところで、和文で記述する必要がある。最初に行なったのは、英和・和英辞書を用いて、翻訳することであった。このために、Wordsの意味解説の英文の字句解析および簡単な構文解析を行なうことにした。そして、スクリプト言語 Python のスクリプトを用いて、英文の文を自動翻訳することにした。このために、フリーライセンスで用いることが可能な次の英和・和英辞典を用いた。

Gene 辞書 ⁽²²⁾	Kurumi 氏による	収録語数約63850語	英和辞典
Edict/JMdict 辞書 ⁽²³⁾	モナッシュ大学日本語アーカイブ	収録語数約57350語	和英辞典

実際に翻訳したところ、まず Gene 辞書はコンピュータ関係主体のもので、あまり結果は芳しくなかった。そこで、より一般的な Edict 辞書の方を用いて、次のような公式を用いて、和英辞典の見出し語を使って翻訳をすることにした。

和英辞書の英語訳 = 羅英辞書の英語訳

しかしながら、これも満足のゆく結果を得られなかった。ここまでの作業で、単純に英和・和英辞書を用いる翻訳はかなり良くないことがわかった。一般の翻訳を行なうサイトも基本的にはこのような手法を用いているので、自動翻訳でおかしな日本語を生成される所以が理解できた。ただし、羅英辞書の英語の表記は、それほど長い文章ではないため、翻訳サイトを利用の方が辞書を使ってそのまま翻訳するよりも、よっぽど良い結果が得られたので、翻訳サイトを利用することにした。幾つかの翻訳サイトを比較して、翻訳結果もまずまずで、一番利用しやすい Google の翻訳サイトを利用することにした。これも、途中までは Google API (Application Program Interface = プログラムから利用するためのライブラリ) を利用して、スクリプトで翻訳サイトに接続し、自動翻訳をするためのプログラミングをしていたが、ユーザ認証の部分がうまく動かないことと、この API で課金されないで利用できるデータの上限は、1 MByte までだったので、断念することにした。羅英辞典の意味の英文だけを抽出した部分だけでも、3 MByte ぐらいはあるからである。

⁽²²⁾ <http://www.namazu.org/~tsuchiya/sdic/data/gene.html>

⁽²³⁾ <http://ftp.monash.edu.au/pub/nihongo/00INDEX.html>

最終的には、Google の翻訳サイトを利用して、コピー&ペーストで、1000行ずつ手作業で翻訳して、和文の意味を作る作業を行なった。羅英辞典の英文だけの抽出したものは、同じ単語の意味でも、内容が違う場合は改行させたため、だいたい80000行程度あった。これを80回、手作業で翻訳した。人間は一日がかりでやれば、80回ぐらいの繰返し作業はなんとかこなすことができるものである。ただし、腱鞘炎には注意したい。1000行ずつに分割したのは、Google の翻訳サイト側で1000行を超えると表示結果がおかしくなるためである。これで和文による意味を生成し、英羅辞典に組み込んだ。元の英文だけのものが3.5MByte ぐらいであったのに対して、和文を組み込んだものは約5.8MByte のデータサイズになった。これから、時間を掛けて1行ごとに手を入れなければならない。単純に1単語で意味が書かれている場合も、和文でどの意味を拾ってくるかの取捨選択は人間が行なわなければならない。これからわかることは、自動翻訳を行なう場合は、どのような分野の何を目的として記述された文書、文というのがわからないと自動翻訳の質は下がるといことである。自動翻訳も手を染めてみたい技術ではあるが、今回の利用によって、自動翻訳では必ず状況を限定する形で行なうべきであることが理解でき、参考になった。

4.2. 辞書閲覧のアプリケーションの作成

まずは iPhone 用に検索だけができる辞書アプリケーションを開発した。iPhone の場合は、その基幹ソフトウェアである iOS の API を利用しなければならないが、この開発には Xcode という開発ソフトウェアを使う。常時、Xcode を利用していた経験もあり、加えて、iOS の API とほぼ共通の Cocoa ライブラリも以前研究会でも教えて慣れているつもりだったが、iOS のソフトウェア開発は難航した。これは、ユーザインタフェースの部分をインターフェース・ビルダという別の開発ソフトウェアと連携するのだが、その連携のために、専用のプログラミング言語 Objective-C で記述されたプログラム上に特別な指定を要するからである。この指定や、不可解な Objective-C の記述方法も加わって、通常のプログラミング言語で開発するよりも、かなり翻弄される部分があった。Web 上の日本人のプログラマが書いてくれた親切な解説を頼りに、なんとか開発することができた。



図6 英語の意味だけを表示するアプリケーション

最初は、羅英辞典の純粋なデータを用いて、英語で意味だけを表示するアプリケーションを開発した。

次に、ラテン語・英語・日本語から検索可能な辞書アプリケーションを開発した。検索バーで入力された文字列を使って中間一致での検索を行なうものである。



図7 和文の意味も表示するアプリケーション

様々なビューを持ったアプリケーションや Android 用のものは今後の課題に残して、ひとまず動くものが開発できたことを成果としたい。

5. 今後の課題

課題は山積している。まず、今回のアプリケーションについては、配布ライセンスを「Words」の制作者の William Whitaker に連絡を取る必要がある。大学の関係者は、このような研究開発で金銭的な利益を得てはならないと考えているので、無料での配布を行なう予定にしている。また、和文の意味の生成は、Google を用いたが、基本的には1つの文ごとに修正していかなければならないだろう。その結果は、たぶん、今回の自動生成で行なったものとは、だいぶ異なるものになると予想される。意味をすべて訂正した後に和文が入ったものを配布する必要がある。無料での配布を行なえば、かなり確率でユーザからの評価が得られるので、それを参考に今後の開発に繋げていきたいと考えている。

Lewis and Short の辞書を入れる場合については、Perseus プロジェクトに配布ライセンスを打診する必要がある。その後、同じく無償での配布をしたいと考えているが、Lewis and Short のデータについては、XML 形式を変換して、Mac OS X の辞書用にして配布することも考えたい。これで、標準の辞書閲覧ソフトウェアを用いて、辞書を閲覧することが可能になる。同じように、Words のデータを Mac OS X 用の辞書に変換する作業も行ないたい。

Words のアプリケーションが行なっているように、すべての活用形を機械的に生成し、元の単語を類推するような機能も実現する必要があるだろう。名詞・形容詞の語形変化や

動詞の活用形にしても、変化した後も同じ形になる場合が多い。Words は、その場合の選択肢を列挙している。

ビューの問題については、実装の部分の制約やまだ iOS の API に不慣れなこともあって、実現していないが、これも実現して、様々な方向から検索できるアプリケーションにしていきたいと考えている。また、iPad などのタブレット機用には、幾つかのビューを一画面に表示することが可能になる。このような実装も考えたい。iOS のアプリケーションでは、iPod Touch/iPhone 用と iPad 用と別に用意されているものも少なくない。大きな画面では、異なる提示の仕方を模索しなければならない。特に、例文から辞書を引くこと、逆に辞書から例文を提示することのリンクをスマートフォン用でもタブレット用でも実現したいと考えている。

archive.org サイトにある単なる OCR で読み取っただけのラテン語のテキストデータから XML 形式の辞書を編纂することも行ないたい。これによって、多くの辞典が、テキストデータとして有効に活用することが可能になるだろう。既に XML 形式になっている Perseus プロジェクトのテキストデータについては、これをスマートフォン・タブレット用の古典のラテン語のビューを作ることも計画に入れたい。

Android 用の実装については、準備ができ次第合わせて作業を開始したい。これもスマートフォン用とタブレット用を開発する予定である。これまで授業で教えてきた Java のフレームワークがほとんどそのまま利用することができるので、iOS の API よりは取りつきやすい。そのため、iOS 版の配布が一段落したら、同じように無償のアプリケーションとして配布することを予定している。

なお、日本語の意味を編纂する作業に合わせて、翻訳用の英単語・和単語の対を作成する予定である。このデータを元に、状況と目的に応じた翻訳を可能とするような自動翻訳用の辞書を編纂することも視野に入れている。最後に、今回はラテン語を対象にしたが、今後、古代文明の幾つかの言語、シュメール・エジプト・タミルなどの言語についても同じような作業を行なっていくことを予定していることを述べておきたい。

最後に、この辞書編纂から始まることであるが、言葉の上位概念・下位概念についてのリンクを張り、それをまとめることを予定している。これによって、用語の木構造あるいはシソーラスのような構造化を行なってゆけるのではないかと考えている。そのような言葉の概念的な構造は、各言語によって異なっているのか、それともほとんど同じになってしまうのかについては、比較言語学の範疇にも踏み込む話題になるかと思われるが、これから継続的に研究して行きたい。

[抄 録]

ラテン語は、フランス語、ロマンス諸語といったその系譜に繋がる現代の言語だけではなく、ドイツ語や英語のようなゲルマン系の言語も含めて、西洋の言語の基礎をなしている。また、その発音方式は日本語に極めて近いものであり、多くの日本人は意識していないが、日本語のローマ字の綴りの方式はラテン語の綴りになっている。加えて、新しい学術的な専門用語などは、ラテン語の造語規則に基づいて作られている。現代まで活きているラテン語について、日本ではあまり多くの教育がなされていない状況にある。英語やフランス語と同様に、そこに多大な影響を及ぼしているラテン語について知識や教養は、言語学的に見ても広く普及させる必要があると思われる。この研究は、潜在的に必要とされるそのようなラテン語教育を、狭い言語学の教室の中で行なうのではなく、もっと一般の人々に対して、開かれた形で、かつ自律的に学べるような形で提供できる言語学習のための辞書システムを編纂することを目的としている。そのために、このシステムの実装の対象をスマートフォンあるいはタブレット式のコンピュータに絞って開発を行なった。フリーの羅英辞書のテキストを用いて、辞書の原型を作成し、それをスマートフォンで提示する形を行なった。また、言葉の意味の英文を手作業で翻訳サイトを用いて和文にし、その和文を一つずつ修正していった。併せて、長音の表記や例文へのリンクなども作成している。この編纂を行なった後の羅日・日羅辞書についても、羅英・英羅辞書の中にも含める形で、かつ本論文で提案した学習用の様々なビューを付随させ、スマートフォンのアプリケーションとして公開する予定にしている。