

# An Improvement of the Hyper Dictionary through Indo-European Roots and Voices

Wataru Ishikawa s1080019

Supervised by Prof. Nobuyoshi Asai

## Abstract

The Hyper English Word Study Dictionary was invented to make the learning of English words more efficient and deep. The Hyper Dictionary possesses contents using Indo-European Roots, a virtual language called the ancestor of many present languages including English. Several English words can be learned via Indo-European roots related to them with functions of the Hyper Dictionary because one Indo-European root word branches to several English words. However, the Hyper Dictionary did not possess multimedia contents. Indo-European is a virtual language, so it is thought that pronunciation inherited from Indo-European roots to present English is significant. In this research, adding voices to the Hyper Dictionary was conducted to put more stress on learning English through pronunciation as a first step of multimedia contents. By development of the Hyper Dictionary, users can listen to voices when they access Word and Indo-European root web pages in the Hyper Dictionary, and pronunciation of words which have the same Indo-European root can be compared. Accomplishment of this research showed that the Hyper dictionary became more massive as an online dictionary and that relationships between English words which have the same Indo-European root became clearer.

## 1 Introduction

### 1.1 Background

Generally, it is said that English is a difficult language to learn for Japanese people. This phenomenon results from Japanese being different from English in the origin and structure of words. Many dictionaries to consult and learn English words exist, but they seem to be inadequate in helping people study English words well. Thus, Indo-European roots were aimed at to solve this problem. Can the learning of English words be conducted deeply using them? The project of the Hyper English Word Study Dictionary with Indo-European roots started based on this thought. [?] [?] [?]

### 1.2 Indo-European Roots

Conventional dictionaries can only provide users translations of English words, and users just memorize them. That is not effective, and it is difficult to strengthen learn-

ing English words. The Hyper Dictionary carries out usage the same as ordinary dictionaries, but it possesses an original feature. The Indo-European roots are inserted into the Hyper Dictionary. First, Proto-Indo-European is a virtual language obtained by theoretical reconstruction in the comparative linguistics field, and a gathering of its stems is grouped in various patterns as the Indo-European roots. It is thought that many languages including English were derived from Indo-European roots. Therefore, these roots can be called the original source language of English, or in other words, its ancestor. These Indo-European roots can be referred from the American Heritage Dictionary [?] [?]. American Heritage Dictionaries are compiled by authorities of various fields such as cultural anthropology, construction, law, military affairs, religion, movie, music, and sports. It is published by Houghton Mifflin Company. Many American Heritage Dictionaries exist, and the American Heritage Dictionary of Indo-European roots is the most important for us and is the basis of this research. A special function is provided in the Hyper Dictionary. The Indo-European roots and English words are linked, and can be found from each other. To learn English words with Indo-European roots is more efficient than to conduct it with nothing, since the words are learned thinking about relations between Indo-European roots and present English. It follows the fact that English words were derived from Indo-European roots. One Indo-European root word branches to several English words, so several English words can be learned via Indo-European roots related to them. This function makes learning English words more efficient. Also, English words can be learned deeply because we can learn them with tracing the origin.

### 1.3 An Improvement of the Hyper Dictionary

By last year, the following functions and information had been established in the Hyper Dictionary [?]: Indo-European roots, prefixes, stems, synonyms, example sentences, applications for study, and interim transitional words. The purposes of this research were to add voices to the Hyper Dictionary and to improve it. There is already a variety of functions in the Hyper Dictionary, especially Indo-European roots, but it has few multimedia

functions, and it seemed that its contents were weak as an online dictionary. It is said that the best way to learn English is to speak English in general, and many books and websites about the great effectiveness of pronunciation are published [?] [?]. Indo-European is a virtual language, so it is thought that pronunciation inherited from Indo-European roots to present English is significant. Therefore, voices were added to the Hyper Dictionary, putting a stress on learning English. [?]

## 2 Discussion

### 2.1 An Environment of the Hyper Dictionary

The Hyper Dictionary runs using several kinds of technology.

Operating System	RedHat 9.0
JAVA Version	J2SDK1.4.2
Web Server	Apache2.0.47
Servlet Container	Tomcat4.1.27
RDB Server	MySQL4.0.16
JDBC Driver	mm.mysql

- The Base of the Hyper Dictionary

The base of the Hyper Dictionary consists of RedHat Linux 9.0 and JAVA language.

- Web Server: Apache

Apache<sup>1</sup> is Hyper Text Transfer Protocol (HTTP) web server software based on The National Center for Supercomputing Applications httpd Ver.1.3, and it is a free and open source. Although Apache is free software, it has high reliability and many functions, and it is said that the reliability and functions are not inferior to those of business web servers, such as the Microsoft Internet Information Server (IIS) and the iPlanet Web Server. Also, Apache has the best share of web servers all over the world.

- Java Source Pages (JSP)

JSP is text documentation which produces responses to user's requests from set protocols such as HTTP. It is mainly used to build web applications. The construction of JSP is that particular tags for JSP were embedded in Hyper Text Markup Language (HTML). The JSP container interprets these tags, and dynamic HTML pages were generated.

- Servlet

Servlet is a technology which builds dynamic contents. It is a class of JAVA, and the class possesses a function which manages requests from HTTP and sends responses.

- Servlet Container: Tomcat

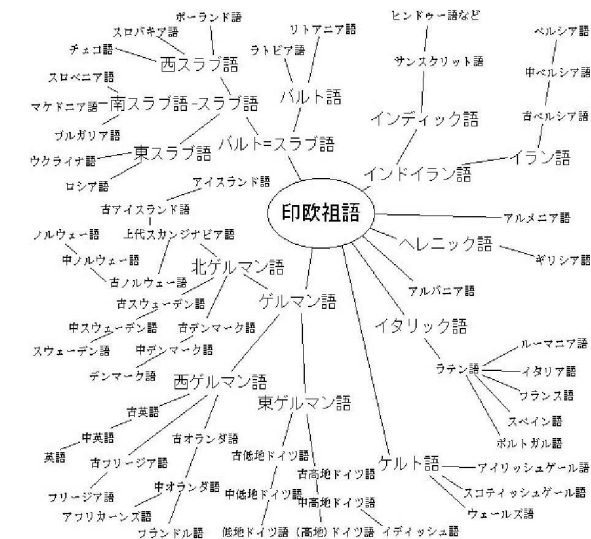


Figure 1: Indo-European Roots

### abdicate [v.]

(王位などを)捨てる, 退く

[用例]

She was forced to abdicate the throne of Spain.

印欧語根	
apo-	「...から離れて」「向こうの方へ」、分離などを表す。その他に「...に関して」の意。重要な派生語は、接頭辞 <b>ab-</b> を持つ語 ( <b>abjure</b> , <b>abrupt</b> , <b>absorb</b> など)、語幹 <b>pos-</b> を持つ語 ( <b>impose</b> など)、 <b>apostrophe</b> , <b>of</b> , <b>off</b> , <b>since</b> など。
deik-	見せることや、(厳重に) 言い渡すことを表す印欧語根。他人に指導する言葉などに関係する意味の語が多い。 1. <b>teach</b> などの由来として、見せることや指導すること。 2. <b>sign</b> の由来として、(指示するためのものとしての) 指。 3. 爪先 ( <b>toe</b> )。4. 語幹 <b>dic-</b> を持つ語 ( <b>dictate</b> , <b>predict</b> など) の由来として、話して教えること。5. 語幹 <b>dic-</b> を持つ語 ( <b>dedicate</b> など) の由来として、宣言すること。6. <b>indeed</b> , <b>prejudice</b> などの由来として、判断すること。

Figure 2: An Example Page of the Hyper Dictionary 2003

<sup>1</sup><http://www.apache.jp/misc/download.html>

Tomcat<sup>2</sup> is an openly developed web application server which runs Java servlet and JSP on a server machine. Sun Microsystems and Apache Software Foundation developed it together, and it is a reference implementation for Java 2 Platform Enterprise Edition, which is a platform in server side Java technology of Sun Microsystems.

- RDB Server: MySQL

MySQL<sup>3</sup> is the most popular open source Relational Database Management System (RDBMS) which is faster at searching for data than other RDBMSs in business. Also, that MySQL can accommodate to several platforms such as Unix, Linux, Microsoft Windows, Sun Solaris, Mac OS, and FreeBSD is an advantage.

- JDBC Driver: MySQL Connector

Java Database Connectivity (JDBC) is an Application Program Interface to access RDBMS from Java programs. Programs which are independent of the kind of database can be developed, but a driver adapted to each database must be introduced to that system in order to actually operate the JDBC. The driver is the JDBC Driver<sup>4</sup>. [?] [?] [?] [?]

## 2.2 About Adding Voices to the Hyper Dictionary

### 2.2.1 Background Information

By last year, there were 10700 English words which could be looked up in the Hyper Dictionary, and more words should be added to it. A new list of 12813 words was distributed to the members of our laboratory, but it had mistakes in spelling, and some words in it overlapped with words which had already been exhibited in the Hyper Dictionary. Since voices must be obtained from a native English speaker to obtain genuine pronunciation of English words, a list of words which would be inserted into the latest Hyper Dictionary needed to be made as soon as possible. Hence, the first step in conducting this research was to make the list. Then the list which was constructed by checking overlaps and mistakes between the new 12813-word list and the old one, and it was sent to an English teacher at the University of Aizu to record his native voice on computer. The voice was made as one WAV file, a music file format, and submitted to me by my supervisor. However, the file had to be split into several files so that each file had only

<sup>2</sup><http://jakarta.apache.org/>

<sup>3</sup><http://www.mysql.com/>

<sup>4</sup><http://www.mysql.com/products/connector-j/index.html>

one pronunciation since the WAV file included pronunciation of many English words. It was conducted using free software, spwave for Windows<sup>5</sup>, My Wave Cutter<sup>6</sup>, and a shareware ALLREC SOUND cutter<sup>7</sup>. In addition, they were converted to MP3 files with CDex<sup>8</sup>, another music file format, and it is generally said that MP3 files have smaller volumes than WAV files, so they can be reproduced more efficiently.

### 2.2.2 Implementation

To reproduce voices, a technique of the Real Player<sup>9</sup> which is also free software was selected because it adapts to Windows and Linux. Old JSP files were improved to link to Real Player with new HTML, JAVA, and Javascript sources. Words requested by clients are read in the Hyper Dictionary, and new JSP files interpret strings of these words as names of voice files. Then, these voice files are reproduced. It is not only a method in which voice reproductions will start automatically when clients click voice icons on web pages of the Hyper Dictionary, but another method in which voices are reproduced by clients with software on client's computers after downloading voice MP3 files. However, in either event, clients need to obtain the Real Player software to listen to voices. Therefore a link to Real.com and a leading to download it were located on the top page of the Hyper Dictionary. [?] [?] [?] [?]

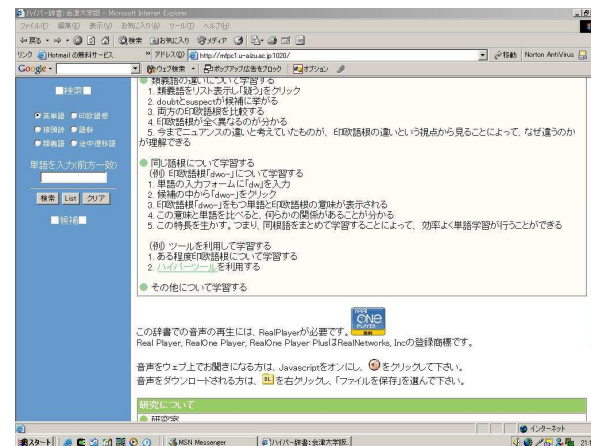


Figure 3: A Leading to Download the Real Player

<sup>5</sup><http://www.itakura.nuee.nagoya-u.ac.jp/people/banno/banno-j.html/>

<sup>6</sup><http://hp.vector.co.jp/authors/VA007724/>

<sup>7</sup><http://pine.zero.ad.jp/meronsoft/>

<sup>8</sup><http://www.cdex.n3.net/lang/langcheck.php/>

<sup>9</sup><http://www.jp.real.com/>

### 3 Conclusion

Clients can listen to voices when they access Words and Indo-European root web pages in the Hyper Dictionary. The Word pages have voice icons near the word searched by clients, and pronunciation can be listened to by clicking the icons. The Indo-European root web page shows a list of English words which have the same Indo-European root. Thus, voices were inserted into each word listed on the web page.

**abdicate** [v.]

〈王位などを捨てる、退く〉

〔用例〕  
She was forced to abdicate the throne of Spain.

A voice MP3 file will be downloaded.  
A voice reduction will start automatically.

印欧語根	
apo-	「...から離れて」「向こうの方へ」、分離などを表す。その他に「...に(関係して)」の意。重要な派生語は、接頭辞 <i>ab-</i> を持つ語 ( <i>abjure</i> , <i>abrupt</i> , <i>absorb</i> など)、語幹 <i>pose</i> を持つ語 ( <i>impose</i> など)、 <i>apostrophe</i> , <i>of, off, since</i> など。
deik-	見せることや、(厳粛に) 言い渡すことを表す印欧語根。他人に指導する言葉などに関する意味の語が多い。 1. <i>teach</i> などの由来として、見せることや指導すること。 2. <i>digit</i> の由来として、(指示するためのものとしての) 指。3. 爪先 ( <i>toe</i> )。4. 語幹 <i>dic-</i> を持つ語 ( <i>dictate</i> , <i>predict</i> など) の由来として、話して教えること。5. 語幹 <i>dicate</i> を持つ語 ( <i>dedicate</i> など) の由来として、宣言すること。6. <i>judge</i> , <i>prejudice</i> などの由来として、判断すること。

Figure 4: A word page featuring voice of the Hyper Dictionary 2004

**aus-**

輝くことを表す (*aurora* など)。 *east* などの由来として、(日が昇る方向としての) 東。

Pronunciations of words which have the same Indo-European root can be compared.

この語根を持つ語	
<i>aurora</i>	オーロラ、極光、曙光
<i>auroral</i>	あけぼのの、バラ色に輝く、極光の(ような)、曙光の(ような)
<i>east</i>	東
<i>Easter</i>	《キリスト教》復活祭、イースター
<i>eastern</i>	東の、東部にある、東からの

Figure 5: An Indo-European Root page featuring Voice of the Hyper Dictionary 2004

### 3.1 Evaluation

The new Hyper Dictionary obtained a valuable function by the addition of voices to the Hyper Dictionary for users to enhance learning efficiency, and it became more massive as an online dictionary. We were able to recognize relationships and connections between Indo-European roots and present English words since pronunciations of words which have the same Indo-European root could be compared. In this research, relationships between English words which have the same Indo-European root became clearer. Hence, it was concluded that addition of voices was very profitable in the development of the Hyper Dictionary.

### 3.2 Future Work

The Hyper Dictionary does not carry phonetic symbols. Therefore, the addition of them and their pronunciation to the Hyper Dictionary will be profitable certainly. Also, with gaining a foothold in addition of voices, dynamic improvements should be brought to the Hyper Dictionary by multimedia contents in the future.

## Acknowledgement

First, I would like to express my gratitude to my supervisor Prof. Nobuyoshi Asai for his great assistance in weekly meetings and personal sessions throughout my research. Also, I am grateful to Prof. Kenneth Cranker and prof. Takafumi Hayashi for providing his voice and making voice files. I received great cooperation from members of my laboratory, Mr. Owada, Mr. Otake, Mr. Matsumoto, and Mr. Takahashi. Finally, I am very thankful to my parents for support and encouragement throughout college.

## References

- [1] “Nihonno eigo kyouiku”,  
<http://www.geocities.co.jp/CollegeLife-Club/5117/page012.html/>
- [2] “WR0160”,  
<http://world-reader.ne.jp/renasci/another/uzaku-990728.html/>
- [3] “SPACE ALC”,  
<http://www.alc.co.jp/>
- [4] C. Watkins, *The American Heritage Dictionary of Indo-European Roots Second Edition*, Houghton Mifflin Company, 2000.
- [5] M. S. Berube, *The American Heritage Dictionary of the English Language Fourth Edition*, Houghton Mifflin Company, 2000.

- [6] “N. Asai laboratory at the University of Aizu, Hyper English Dictionary 2003 Edition”,  
<http://hidic.u-aizu.ac.jp/>
- [7] M.Kunihiro, *Eikaiwa zettai ondoku chousenhen*, Koudansha, 2001.
- [8] Y.Noguchi, *Chou Benkyouhou*, Koudansha, 1995.
- [9] Y.Horikawa, “Hyper English Word Study Dictionary with Indo-European Roots and Interim Transitional Words”,  
Graduation thesis in the department of Computer Science and Engineering at University of Aizu, 2003
- [10] “IBM Japan”,  
<http://www-6.ibm.com/jp/software/data/developer/library/techdoc/tomcat.html>
- [11] “@IT: Linux Square”,  
<http://www.atmarkit.co.jp/flinux/index.html>
- [12] “e-Words”,  
<http://e-words.jp/>
- [13] “Wings”,  
<http://www.wings.msn.to/>
- [14] “Tohoho no WWW Nyumon”,  
<http://tohoho.wakusei.ne.jp/www.htm>
- [15] “Redworks Products”,  
<http://www.redworks.org/tips/real/real2.html>
- [16] M. Hurukawa, “Tomcat de manabu J2EE”,  
*JAVA Developer*, SOFT BANK Publishing, No.11, JUNE, 2003, pp.88-95.
- [17] “Web kibako”,  
<http://www5c.biglobe.ne.jp/horoau/index.html>