# Improvement of The Hyper Dictionary with Indo-European Roots in Classification

Shingo Ootake        s1080047                    Supervised by Prof. Nobuyoshi Asai

## Abstract

This paper describes the development and improvement of the English-Japanese Hyper Dictionary characterized by Indo-European Roots. Indo-European Roots is a virtual language that made by comparison of many languages which have similarities. The previous English-Japanese Hyper Dictionary had a problem in that users were not able to reference the special meanings of several words which are commonly used in particular fields in the Hyper Dictionary. I considered that classifying English words by according to specialized fields was necessary in the Hyper Dictionary, and suggested adding new classifications and a new data table to the Hyper Dictionary. These improvements should help Japanese learners to understand English words more deeply and by solving the above problem.

## 1 Introduction

### 1.1 Back Ground

#### 1.1.1 The Needs of Study English

Presently, internationalization is spreading into all fields (business, entertainment, education, and any other fields), so people must be able to communicate with each other more effectively. Since English has become the global language, it is a very important language for Japanese to learn. Presently, many Japanese desperately seek for some effective learning methods. Most learners generally study, by listenning to native English pronunciation, memorizing the structures of English sentences, and talking with native English speaker. In most cases, an English-Japanese and Japanese-English dictionary is needed.

#### 1.1.2 Study of English Words

A Japanese-English or English-Japanese dictionary is often necessary for understanding the definitions of English words. Vocabulary is the foundation of English. Vocabulary is thus one of the most important elements when a person learns a language because a large vocabulary is necessary to select appropriate expressions in a language. Therefore, the study of English words is very important when studying English.

#### 1.1.3 Online Dictionary

A paper dictionary can be used anywhere, but it can easily deteriorate. However, an online dictionary does not deteriorate, and its database can be updated easily. The Internet is diffused in almost areas of the world, and is free except for equipment and connection charges. When a person wants to know about an English word, accessing information is easier in an online dictionary with a mouse click than by turning the pages of a paper dictionary. A web page can also include sound, video and graphics data. Both audio and visual information are more helpful for human recall and understanding than sight only. Language is living and perpetually changing. Width of usable area and immidiate responce of online dictionary may be adapt for international people because they often move all around the world with baggage as small as possible. Online dictionary is not big and heavy baggage. This is different from paper dictionary.

#### 1.1.4 Development of The *Hyper Dictionary*

Learning and understanding English words is difficult for Japanese because few Japanese know the origins and structures of the words they study. Developers have tried to revolutionize in the field of language education. [15] As one solution, the Hyper Dictionary was developed. The Hyper Dictionary is an English-Japanese web dictionary which is useful for the study of English words by Japanese learners. The Hyper Dictionary utilizes Indo-European roots so learners can understand the meanings of English words more deeply. The American Heritage Dictionary features the roots of words [2]. Indo-European roots were used between 3000 B.C. and 5000 B.C. and are the original source language of English. Therefore, these roots are considered as a good tool for understanding the meanings of the origins of English words. Now, the Hyper Dictionary has several searching methods for English words. A user can search and know not only English words but also Indo-European roots.

#### 1.1.5 About American Heritage Dictionary

The American Heritage Dictionary (AHD) is special dictionary. The editors of AHD are experts of several especial fields (Anthropology, Architecture, Cinema, Dance, Law, Mathematics, Military science and Weapons, Music, Nautical, Philosophy, Religion, Sports etc.). This dictionary enables native English users to know the def-

initions of English words.

### 1.1.6   Indo-European Roots

Indo-European roots are specific languages that were made by comparing many languages (Sanskrit, Greek, Latin, German). In the 19th century, comparative linguist were able to group the languages used in Iceland and Ireland in the west to India in the east, and from Scandinavia in the north to Italy and Greece in the south, into a family called *Indo-European* or *Proto-Indo-European*. The dialects or branches of Indo-European are still represented today by one or more of the following languages: Indo-Iranian, Greek, Armenian, Balto-Slavic, Albanian, Celtic, Italic, and Germanic etc. (see Figure 1). Some patterns of word stems of Indo-European are called *Indo-European roots* as hypothetical languages.
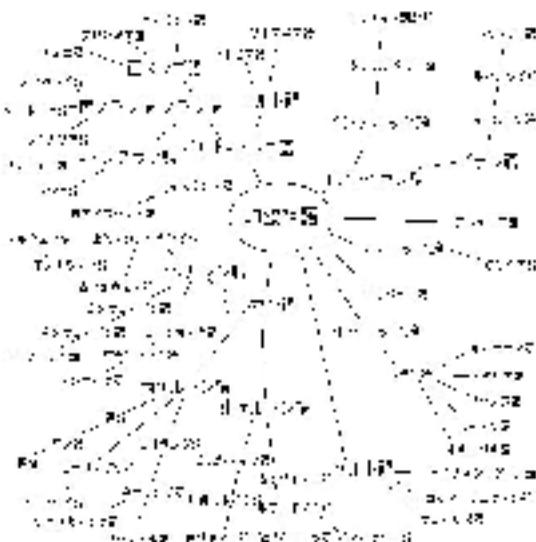


Figure 1: derivation

## 1.2   Functions of Former Hyper Dictionary

Previous work created several functions in the Hyper Dictionary. The Hyper Dictionary has searching functions with English words, Indo-European roots, English prefixes, English synonyms, stems of English words, and interim transitional words (see Table 1). In 2003, researchers changed the operating system from Windows to Linux, and added an application for learning English words and Indo-European roots called a *Hyper Tool* (see Table 1).

These searching methods are front matching.

Table 1: All functions of the Hyper Dictionary added in 2003 [1]

| Functions | Result information |
|---|---|
| English word search | spelling, part of speech, meaning, using example, related Indo-European roots, stems of the word, prefix, interim transitional words. |
| Indo-European root search | spelling, meaning, English words that contained the Indo-European root. |
| English prefix search | spelling, meaning, related English word list containing spelling and meanings. |
| English synonyms | select English words with Japanese meanings, so a list and link for jumping to a selected English word's page was created. |
| Interim transitional words search | related English word list containing the spelling and meanings. |
| Stem of English words search | spelling, meaning, related English word list containing spelling and meanings. |
| Hyper Tool | An application to study English words and Indo-European roots. |

## 1.3   Main Purpose

The main purpose of this research is to improve the Hyper Dictionary and add a function which would be more useful for users to learn English words. The goal is to improve the classification of English words, Indo-European roots, and interim transitional words by adding definite meanings and addition meanings of English words in specialized fields, e.g., Mathematics, Medicine, and Computer science, etc. The task of other researches in our laboratory is to add examples of English sentences that included links to jump to an English word's information page, a function to allow learners to hear native English pronunciation, and a database of suffixes of English words to the Hyper Dictionary. In addition, we added new English word and Indo-European root data, as well as required prefixes, interim transitional words, and stems of English words.

## 2   Classification

This research focused on classification of English words, Indo-European roots and interim transitional words.

First, I will describe the method of classification used. Presently, English words are classified as an editor's original pattern, part of speech, and origin of words, etc. Classification of parts of speech is suitable for exams or short time remembrance, but is unsuitable for a deep understanding of English words or long time remembrance; therefore, I did not select this method. Classification with origin of words is not suitable for the Hyper Dictionary became English words included in the Hyper dictionary have already been classified by Indo-European roots. I didn't have right or license to classify English words originally, so I classified English words with definite meanings [2] [5] [6] [7] [8]. Definite categories pertain to special fields. I added meanings of English words included in various categories by expanding the searching function so users can know the special word meanings and understand the words more deeply. Equally, I classified Indo-European roots and interim transitional words in case the meanings of these words can be classified into categories used in English words. (Please refer to *Table 2* for a list of classified fields.) Next, 74 percent of 17000 English words were classified.

Table 2: Used Categories

| Category | English words | Indo-European roots | interim transitional words |
|---|---|---|---|
| Mathematics and physics | * | | |
| electricity and election | * | | |
| computer | * | | |
| medicine | * | | |
| law | * | | * |
| plant | * | * | * |
| animal | * | * | |
| part of (human) body | * | * | |
| color | * | * | |

# 3  Implementation

## 3.1  Explanation of Technology used the Hyper Dictionary

Since much importance was given to a free development environment, the development of the Hyper Dictionary is used as freeware or free development system. The development environment is the Linux operating system, and the Hyper Dictionary used MySQL for database development, as well as Tomcat and Apache for JSP and Servlet.

### 3.1.1  About Linux

Linux is a free Unix-type operating system originally created by Linus Torvalds with the assistance of developers from around the world. Developed under the GNU General Public License, the source code for Linux is freely available to anyone. [9]

### 3.1.2  About MySQL

MySQL is the world's most popular open source database, recognized for its speed and reliability. MySQL AB, the company founded by the creators of the MySQL database, provides MySQL software development and related support and services. [11]

### 3.1.3  About Tomcat

Tomcat is servlet container that is used in the official Reference Implementation for the Java Servlet and JavaServer Pages technologies. The Java Servlet and JavaServer Pages specifications were developed by Sun under the Java Community Process. [10]

### 3.1.4  About Apache

Apache is a free web HTTP server software based on NCSA httpd Ver.1.3 and developed by the National Center for Super-computing Applications (NCSA). It can operate in most environments (Linux, UNIX, Windows, and Macintosh). The Apache Project is a collaborative software development effort aimed at creating a robust, commercial-grade, featureful, and freely-available source code implementation of an HTTP (Web) server. The project is jointly managed by a group of volunteers located around the world, using the Internet and the Web to communicate, plan, and develop the server and its related documentation. These volunteers are known as the Apache Group. [10]

### 3.1.5  About JSP/Servlet

Java Server Pages (JSP) is a technology that lets users mix regular, static HTML with dynamically-generated HTML. Many Web pages that are built by CGI programs are mostly static, with the dynamic part limited to a few small locations; but most CGI variations, including servlets, make users generate the entire page via their program, even though most of it is always the same. JSP allows users to create the two parts separately. [13]

Servlets are Java technology's answer to CGI programming. They are programs that run on a Web server and build Web pages. [13]

## 3.2 New addition

I added two thousand English words to the Hyper Dictionary. I checked and added parts of speech, and general meanings. Indo-European Roots were checked and updated as well. I made a new data table for classification with MySQL in the database of the Hyper Dictionary. The name of new data table is called *"Ebunrui"* which contains data of the spellings of English words, special meanings with Japanese translation, ID numbers, and names of classified categories. I expanded the data table of Indo-European roots and interim transitional words, and changed visual expressions on the web pages, as shown in the following figure.
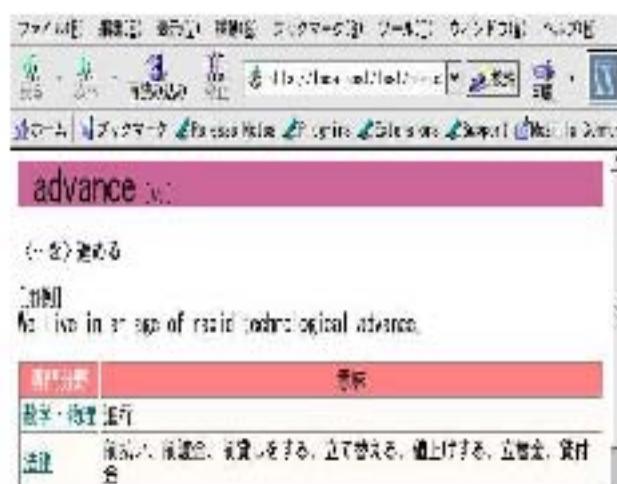


Figure 2: Implemented Sample

## 3.3 Implementation Environment

We implemented an environment of functions as shown in then Table 3.

Table 3: Implementation Environment

| OS | RedHat9.0 |
|---|---|
| Web Server | Apache2.0.47 |
| Java | J2SDK1.4.2-01 |
| Servlet Container | Tomcat4.1.27 |
| RDB Server | MySQL4.0.16 |
| JDBC Driver | mm.mysql |

Changed Point in the Web display

- Addition of a new table, which includes classified names and special meanings to English word's page.

- Addition of a new column, which includes classified names to Indo-European roots and interim transitional words.

## 4 Result and Evaluation

As a result, I succeeded in making a database of classified English words, Indo-European roots, interim transitional words, and added these to the Hyper Dictionary.

I asked ten participant to use the improved Hyper Dictionary. A few people thought there were a few visual problems, but others considered the improvements to be very good and helpful. They thought the new data which classified meanings as very useful in word searching and understanding of words.

## 5 Conclusion

In this research, a new data table called "bunrui" was added to the Hyper Dictionary which contains classified English words, and their meanings in specialized fields (e.g., Mathematics, Medicines and Computers, etc.). The problem of users not being able to reference the specific meanings of several words which are used in special fields with Hyper Dictionary was solved by this research. The usability of the Hyper Dictionary was improved.

## Acknowledgement

## References

[1] N. Asai laboratory at University of Aizu, Hyper English Dictionary,
http://tnfntsv1.u-aizu.ac.jp/hidic/

[2] Margery S. Berube, *The American Heritage Dictionary of The English Language 4th Edition*, Houghton Mifflin Company, 2000.

[3] C. Watkins, *The American Heritage Dictionary of Indo-European Roots Second Edition*, Houghton Mifflin Company, 2000.

[4] A. S. Hornby and Sally Wehmeier, *Oxford Advanced Learner's Dictionary*, Oxford University Press, 2002.

[5] K. Kihara, *GRAND CONCISE ENGLISH-JAPANESE DICTIONARY* SANSEIDO, 2001.

[6] T. Konishi and K. Minamide, *TAISHUKAN'S Unbridged Genius English-Japanese Dictionary* Taishukan Publishing Company, Tokyo, 2001.

[7] T. Konishi and M. Yasui and T. Kunihiro and K. Horiuchi, *Random House English-Japanese Dictionary* Shogakukan, 2002.

[8] H. Michihata, 英辞郎 , ALC Inc., 2002.

[9] Linux Operating System, "Linux Online Web Site" http://www.linux.org/

[10] Apache Jakarta Project, "The Jakarta Site — Apache Tomcat" http://jakarta.apache.org/tomcat/index.html

[11] MySQL AB, http://www.mysql.com/

[12] Sun microsystems, "Site of Sun microsytems – JAVA" http://java.sun.com/

[13] Java Boutique, "WEB technology with java Site" http://javaboutique.internet.com/tutorials/JSP/part01/page03.html

[14] S. Yoshimura, "A study about the hyper-dictionary used Indo-European roots information", Master's thesis in the department of Science and Engineering at University of Tsukuba, 1997.

[15] Yasuhiko Ikebe, 印欧語根を用いたハイパー英単語辞書の公開, 研究報告 「人文科学とコンピュータ」 アブストラクト No.045 - 009, Tue Jun 19 14:53:57 2001

[16] 専 門 語 辞 書 参 照 サ ー ビ ス, http://wwwd.nova.co.jp/webdic/webdic.html