

Upgrading the English-Japanese Hyper-Dictionary

Hideo Sugiura s1061091

Supervised by Nobuyoshi Asai

Abstract

In this study, we verify a function insufficient in an *English-Japanese Hyper-Dictionary* with using *Indo-European roots*, and try to adding new function. We added the function which is able to search *middle change word* to clear the flow of change word, and bug fix. This research continues the preceding research of Ikebe at University of Tsukuba. This study is based on the idea that of English education in Japan cannot be reformed without reform of a dictionary. This research is done by the student of Shizuoka Sangyo University, and Tsukuba university, and the University of Aizu.

1 Introduction

1.1 Background

The plan which educates English from an elementary school or a kindergarten is examined recently in Japan. In order to understand English, it is necessary to understand the grammar of English culture, but If they are memorized once on the body, we can apply it, and what must be studied permanently is English words. And when learning English words, the learning method using the Indo-European roots are more efficient than other learning methods.

1.1.1 What is Indo-European roots?

A *Indo-European roots* is the ancestor of Indo-European languages containing English, and it is examined that is the single language group from B.C 5000 years to 3000 years. Many languages progressed and derived from *Indo-European roots* such as Iceland in Irish, Ireland, Russian, Indian, Chinese, etc., and they are divided into about 12 word groups they form the most important language family in cultural history. Research of the *Indo-European roots* began by William Jones who is British and India scholar and lawyer and he has discovered the similarity of the remarkable language feature between Sanskrit, and Greek and Latin which are a classic word of Europe from 1786. Recently, the genealogy and the feature of Indo-European languages family are becoming clear by theoretical reconstruction of following the regular phonological transition in Indo-European languages or comparative linguistics etc. [1]

1.1.2 English study using the Indo-European roots

Professor Ikebe devised the English teaching materials for study using the *Indo-European roots* paying attention to this *Indo-European roots*. [2] The English teaching materials for study are called the *English-Japanese Hyper-Dictionary*, and research has been conducted of the Tsukuba University until now. We are doing the joint research with Shizuoka Sangyo University students now. [3] The *English-Japanese Hyper-Dictionary* is an English-Japanese dictionary on the web for "searching the *Indo-European roots* of a given English word, searching the whole of the derivative, and studying these collectively." Now, the *English-Japanese Hyper-Dictionary* is on the web (<http://mfntsv1/hidic/hdhome.htm>) [4], and can search a Indo-European root from the present English word, or a present English word from an Indo-European root. The big merit of using this English word study hyper-dictionary is being able to learn the English word as it originates from the same root collectively. It is easy to use the difficult study method which can learn the difference in a synonymous meaning and the difference in usage from the Indo-European root. Weaknesses of the current English word study Hyper dictionary are as follows.

- It cannot be known along what route the present English word has passed to the present *English-Japanese Hyper-Dictionary*.
- There are a few English words registered too often.
- A bug of the program is that some *Indo-European roots* which are registered cannot be searched.
- A Rarity in the explanations about the difference in synonyms.

These Four points which we have noticed and should be improved especially. The target of our research is to make an *English-Japanese Hyper-Dictionary* that is easier to use and more complete. I focused on fixing bug and searching for the words between *Indo-European roots* and present English words.

1.2 Purpose

This *English-Japanese Hyper-Dictionary* cannot be called completed one yet, has and it points which should

still be improved partly. An English word did not become its present form from a Indo-European root suddenly. It became the present form through various routes, such as Germanic language, old English, and Greek. We cannot know along what route the present English word has passed by the present *English-Japanese Hyper-Dictionary*. One bug in the program is that there are some *Indo-European roots* which cannot be searched from among the words registered, and there are a few words registered too often, and The explanation about the difference in synonyms is not detailed. These four points should be improved in particular. The target of our research is to make an English word study hyper-dictionary easier to use, and to be more completed. In these points which should be improved, I researched that correct the bug on the program, and possible to search the middle words between *Indo-European roots* and present English words. About the increase in a registration word, and the addition of synonyms and detailed explanation of that, please refer to a paper by Hoshino and Yamanaka who studied the English word study hyper-dictionary with me. [5], [6]

2 Discussion

2.1 Necessarity of this research

The purpose of this research is to make an English word study hyper-dictionary easier and more practical to use. The greatest selling point of an English word study hyper-dictionary is it can give a story to a word by classifying it according to its root. We will be interested in English words by these stories. However, if a English word study hyper-dictionary is used, it may be difficult to imagine the path by which a present word is produced from this root. In the case of a synonym, in spite of seeing the flow from a Indo-European root and judging the difference in delicate use, the flow can seldom be seen. A concrete example is shown below [7].

2.1.1 The result when searching a English word with the present English word study hyper-dictionary

The words *abandon*, *forsake*, *desert*, etc. all have the meaning abandon. First, when referring to *abandon* in a present hyper-dictionary, the Indo-European root is *bh#a-2*, and the meaning is to talk, declare. Next, when *forsake* is searched (Figure: 1), Indo-European root is *s#ag-*, and the meaning is discovered. The last *desert's* Indo-European root is *ser-3*, and the meaning is 1 sequence and securing. Probably, it will be very difficult to visualize the difference in meaning between each word and the difference in use from only this information. The story between each word and their *Indo-European roots*

are all difficult to visualize. We think the reason for this problem is information in the *English-Japanese Hyper-Dictionary* are only present English and *Indo-European roots*.



Figure 1: Case of serching forsake

2.1.2 The result when adding the middle change word and searching an English word

In order to solve this problem, we investigate the language in the middle changing from an Indo-European root to the present word. The flow of *abandon* is from *bh#a-2* through the old French *ban*; then it became the present form *abandon*. We can be reminded that the meaning of *ban* is from *bh#a-2* which meaning is talking or meeting, and set to a meaning called the summons for military business. We can imagine that a military command does not depend on individual convenience, having been set to *abandon* from the meaning of abandoning individual convenience unavoidably. Next, in the word *forsake*, the Indo-European root is *s#ag-* meaning discovered (table1: 1), and it becomes old English *for-sacan* the meaning is stop spontaneously(discovering). This becomes its present English word *forsake*. Finally, *desert* became the *ser-3* the meaning prepared and 1 sequence; it becomes the Latin *serere*, meaning securing of arguing, and became the present word *desert*. If the three above words are viewed, the change of a word is clearer than with only a Indo-European root and the present English word. Next, we call the *middle change word* which is the language between Indo-European root and present words generically.

2.2 Input to database

We say that a *middle change word* makes more clear the flow of word change in 2.1.2. Then, the middle change word in stored the database which flows the word having synonym. Synonym is important to the flow of word change to understand the delicate difference of meaning.

Table 1: Information of forsake in AHD

<i>s#ag-</i> To seek out. FORSAKE, from Old English forsacan, to renounce, refuse

For this purpose a relational database *Microsoft Access* is used.

2.2.1 What middle changes word of the present English word did we investigated?

A list of the words which have the *middle change word* is the *Learner's Progressive English-Japanese Dictionary Second Edition* (Shogakukan). [8] In this dictionary, if a word has a synonym explained, other synonyms are shown in another window.(table: 2)

Table 2: Information of synonym in English-Japanese dictionary

類語 ability: 頭腦的、肉体的に何かを 成し遂げることができる能力、技能全般 をいう。capability: タイプ、運転など 特定の技能に優れた能力。talent: 人が 生来備えている、特に芸術的な天分のこ と。genius: 限られた小数の人間にしか 与えられていない生まれつきの独創的な才能。
--

In this research, we added to a database to the *middle change word* with synonyms indicated by this dictionary.

2.2.2 Explanation of database relation

We explain the relation of the *middle change word* in the database. The relation of the database *word master* which has stored the English word was as follows.

- Word master [spelling, headline distinction, comment, word root1, word root2, word root3, word root4, prefix1, prefix2, stem]

The attribute of root1-4 or prefix1-2 takes the integer value which marks the Indo-European root and prefix. If an English word has some Indo-European root or prefix, it takes these data from each attribute 1-4, and if not, it takes -1 not to take the null value. We added the relation of middle change word to the above relation, which became the following:

- Word master [spelling, headline distinction, comment, word root1, word root2, word root3, word

root4, middle word1, middle word2, prefix1, prefix2, stem]

The attribute of the middle word1-2 takes the integer value to indicate the middle change words such as that of root1-4 or prefix1-2. If an English word does not have a *middle change word*, it takes -1 not to take null value.

- middle word [distinction, spelling, meaning]

Distinction is made of the consecutive numbers of the *middle change word*, and they used to attach in order to increase the efficiency of the work which treats the *middle change word*. Also, this distinction has the same contents as the attribute of middle word 1-2 in relation of the database *word master* which has the present English word stored. Since the character which cannot be treated in the usual European-languages font is contained in the notation of the *middle change word*, it is replaced as follows:

$\hat{a} \hat{i} \hat{u} \hat{e} \hat{o} \longrightarrow \bar{a} \bar{i} \bar{u} \bar{e} \bar{o}$
 $a@ \longrightarrow \bar{ae}$

Figure 2: Replaced font table

2.2.3 The realization method of the English word study hyper-dictionary on WWW

In order to realize an *English-Japanese Hyper-Dictionary* on WWW, an environment which enables use of a relational database system from a WWW server is required. The present *English-Japanese Hyper-Dictionary* has planned the convenience of research using the original server (IIS) installed in the computer which our laboratory owns. However, by this research, we thought that it was better to use another server which operates lightly by the local host, because we need to test the English word study hyper-dictionary program, and it is easy to setup server by GUI. For the above reason, an environment which operates the WWW server was used, and the system of English word study hyper-dictionary used at the time of this research is as follows:

- OS *Microsoft WindowsNT (Microsoft)*
- WWW server *AN HTTP Daemon version1.36b*

AN HTTP is a WWW server for Windows95/98/Me/NT/2000. This server can offer home pages from a Windows machine connected to LAN, and SSI/CGI can be local tested, and it is free-ware. Also,

the test of CGI operation can run PC which is not connected to a network. Detailed information is indicated by the creator's home-page at (<http://www.st.rim.or.jp/~nakata/>). This server program is not suitable for many people's use, but only for test of research and development, because this program can treat CGI by a stand-alone computer, and it is easy to carry out a server setup.

- Relational database system *Microsoft Access2000* (*Microsoft*)
- Program language *Microsoft Visual Basic ver.6.0* (*VB*)(*Microsoft*)

The BASIC compiler can create the execution file using the external runtime library. As a big feature, it is able to use a relational database system easily out of a program. A database can be controlled from a program without writing a complicated code. Past researchers developed the program in this programming language by these advantages. When a program is improved, since the advantage of this programming language is very attractive, we can utilize the past property. This research also uses this programming language. The structure of the English word study hyper-dictionary is shown below.

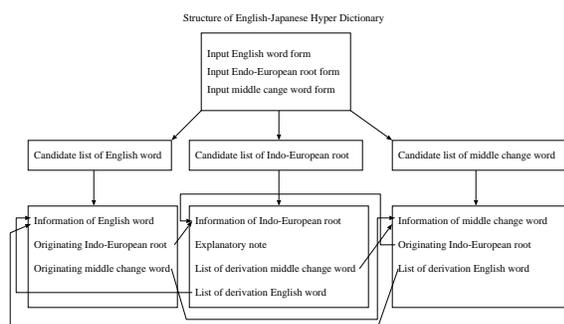


Figure 3: Structure of Hyper-Dictionary

2.3 Bug fix

Some fatal bugs exist in the present English word study hyper-dictionary. One bug is that there is a Indo-European root which cannot be searched. Some reasons for bugs like this include the problem in a program, the problem of a server, etc. Two examined as a reason which cannot be searched, as shown in the example below.

2.3.1 Searching a word which has

First, if there is Indo-European root which has \bar{a} , or \bar{b} , etc, they are replaced by $\#a$, $\#b$, etc. When an Indo-

European root like a $\bar{s}ag$ - are searched, this string is inputted as $s\#ag$ -. Then, in spite of registering the Indo-European root into the database, the result in which the Indo-European root is not registered occurs. It becomes clear that # is recognized as a special character which has the arbitrary integers of 0-9 in the programming language VB. In order to # is recognized not as a special character but normal character by VB, just to attach [] to the both ends. Then, in the program, we solved the problem of when the character sequence containing # was inputted, inserted [] before and after it to create a new character sequence with [].

2.3.2 Overflow

The second bug is when Indo-European root are searched, if there are 100 or more English words derived from it. Since there are more English words than the number of the arrangements for storing them, there is overflow, and a CGI error is occurs. For this problem, I made the number of derivation English words correspond to 400 words, and solved it. The number of English words derived from one Indo-European root is at most 200. But when a researcher tries to increase *Indo-European roots* or English words in the database, there may be *Indo-European roots* whose number of derivation English words is 200 or more. Even when this occurs, we thought that it was unnecessary to rewrite the program, so it is referred to as 400 in the margin.

3 Conclusion

3.1 Consider about the focus to the middle change words

When learning an English word, the concept of investigating a word from *middle change words* or Indo-European roots is seldom known. However, as described in sections 1.1 and 2.1, the merit of learning Indo-European roots and *middle change words* cannot be disregarded. By having included the information about middle change words in the English word hyper-dictionary, it became easier to memorize an English word than before. A classification or explanation of chronological order was not able to describe a word which has two or more *middle change words*, since this was the first time that *middle change words* were classified. Moreover, the *middle change words* which were investigated were the parts of English words now registered into the present database, we feel it necessary to investigate *middle change words* of other English words not yet registered. However, these are so important that they should be examined as one research theme in the future.

3.2 Future work

By having included the *middle change words* into the English word study hyper-dictionary, we thought that information required to learn English words using *Indo-European roots*. However, the number of registered *middle change words*, or *Indo-European roots*, are still too few than in a regular dictionary. In the future, it will be necessary to increase the number of registered words, or correct the bugs which still exist, so this hyper-dictionary will be used by more people. It should be evaluated.

Acknowledgment

I wish to thank my supervisor, Prof. Yasuhiko Ikebe, for not only providing this interesting topic and helpful suggestions, but for help in learning English. I also want to thank Prof. Nobuyoshi Asai, and Prof. Yasusi Kikuchi for technical advice, Prof. Stephen Lambacher for checking English drafts, and my collaborator Yoshihiro Hoshino and Chie Yamanaka for providing other valuable information.

References

- [1] Y.Terasawa, "The kenkyusya Dictionary of English etymology", IDG Kenkyusya, pp.1627-1685, 1997(in Japanese)
- [2] Y.Ikebe, "Eigogijutu purezentesyon no skill up number1", Human Interfae Society, pp.27-30, 2001(in Japanese)
- [3] Y.Ikebe, "hyper-text dictionary: sogenteki oyobi goshitekigogenkenkyu ni motodoku eigoi yousei-hou", University of Tukuba, pp.801-808, 2000(in Japanese)
- [4] English-Japanese Hyper-Dictionary at Aizu University
(<http://mfntsv1.u-aizu.ac.jp:8080/hidic/hdhome.htm>)
- [5] Y.Hoshino, "Upgrading the Indo-European-Root-Based Hyper English Dictionary", University of Aizu,2001
- [6] C.Yamanaka, "Synonym Usage viewed from Pre-Historic Roots", University of Aizu,2001
- [7] Calvert Watkins, "The American Heritage Dictionary of INDO-EUROPEAN ROOTS, second edition", Houghton Mifflin Company, 1995
- [8] Lerner's Progressive English-Japanese Dictionary Second Edition, Shogakukan, 1997(in Japanese)