Detection of inconsistencies in concept classifications in a large dictionary
― Toward an improvement of the EDR electronic dictionary ―

Eiko Yamamoto, Kyoko Kanzaki, and Hitoshi Isahara

Computational Linguistics Group

National Institute of Information and Communications Technology

3-5 Hikari-dai, Seika-cho, Soraku-gun, Kyoto 619-0289, Japan

E-mail: {eiko, kanzaki, isahara}@nict.go.jp

Abstract

The EDR electronic dictionary is a machine-tractable dictionary developed for advanced computer-based processing of natural language. This dictionary comprises eleven sub-dictionaries, including a concept dictionary, word dictionaries, bilingual dictionaries, co-occurrence dictionaries, and a technical terminology dictionary. In this study, we focus on the concept dictionary and aim to revise the arrangement of concepts for improving the EDR electronic dictionary. We believe that unsuitable concepts in a class differ from other concepts in the same class from an abstract perspective. From this notion, we first try to automatically extract those concepts unsuited to the class. We then try semi-automatically to amend the concept explications used to explain the meanings to human users and rearrange them in suitable classes. In the experiment, we try to revise those concepts that are the lower-concepts of the concept "human" in the concept hierarchy and that are directly arranged under concepts with concept explications such as "person as defined by –" and "person viewed from –." We analyze the result and evaluate our approach.

1. Introduction

Although some corpus-based dictionaries exist, such as the EDR electronic dictionary compiled with the EDR corpus (EDR, 1995) and the Collins COBUILD English language dictionary (Sinclair, 1987) compiled with a large corpus, that is, the Bank of English (Renouf, 1987), no dictionary currently exists that has a conceptual structure extracted automatically from scratch with a computer. That is, the information used to construct the dictionary structure is still extracted from the corpus by hand. Because several lexicographers compile it manually, several inconsistent structures appear in the dictionary. In this paper, we present a semi-automatic method for revising such errors using only the information contained in the dictionary itself.

National Institute of Information and Communications Technology (NICT) holds all copyrights for the EDR electronic dictionary. The dictionary comprises eleven sub-dictionaries, which include a concept dictionary, word dictionaries, bilingual dictionaries, co-occurrence dictionaries, a technical terminology dictionary, and the EDR corpus. Recently, we have been revising the EDR electronic dictionary and will be publishing the next edition. We first try to revise the concept dictionary of the EDR electronic dictionary by using information already stored in the EDR electronic dictionary. In other word, we take another second look at the EDR electronic dictionary in order to revise it. Because several lexicographers performed the concept classification individually and their results were assembled in the concept dictionary without a comparison of the results of all lexicographers, we can find concepts that are not suitably classified in the hierarchy.

For example, there are 43 prefectures in Japan and all of them should be classified into the same class, i.e., they must be linked to the concept "name of prefectures," which is linked to the upper-concept "region of land indicated by place name (place name of administrative jurisdictions, districts, etc.)" in the EDR electronic dictionary. However, the concept "a prefecture in Japan called Miyagi" is classified under the concept "region of land indicated by place name (name of Japanese localities)." This is clearly an error, and the concept "a prefecture in Japan called Miyagi" should be rearranged under the concept "name of prefectures" to maintain uniformity with the classification of the names of the other prefectures. The EDR electronic dictionary contains numerous similar concepts requiring rearrangement.

In this study, we aim to revise the arrangement of such concepts in order to improve the EDR electronic dictionary. However, it is a vast dictionary with records of more than 400,000 concepts. This suggests that checking all links between concepts would be all but impossible with manpower alone. Therefore, we need to automatically extract the candidates in the EDR concept dictionary that should be revised.

We automatically extract the concepts that might be in unsuitable classes and try to rearrange them semi-automatically in suitable classes. Each concept in the EDR concept dictionary has a phrase explaining its meaning to the user (concept explication). We use this concept explication to extract the candidate concepts mentioned above, correct the concept explications, and place the concepts in the appropriate positions.

In this paper, we try to revise concepts that are lower-concepts in the hierarchy of the concept ningen ("human") and that are directly arranged under concepts that tend to have concept explications such as – *de toraeta ningen* ("person as defined by –", "person viewed from –", etc.). We then analyze the result and evaluate our approach.

2. EDR Electronic Dictionary

2.1 Overview

The EDR electronic dictionary (EDR, 1995) was developed for advanced processing of natural language by computers. This dictionary is the result of a nine-year project (from fiscal 1986 to fiscal 1994) aimed at establishing an infrastructure for knowledge information processing. The project was funded by the Japan Key Technology Center and eight computer manufacturers[1].

The EDR electronic dictionary catalogues the lexical knowledge of Japanese and English (the word dictionaries, the bilingual dictionaries, and the co-occurrence dictionaries), and has unified thesaurus-like concept classifications (the concept dictionary) with corpus databases (the EDR corpus). The concept classification dictionary, a sub-dictionary of the concept dictionary, describes the similarity relation among concepts listed in the word dictionary. The EDR corpus is the source for the information described in each of the sub-dictionaries. The basic approach taken during the development of the dictionary was to avoid a particular linguistic theory and to allow for adoptability to various applications.

The EDR electronic dictionary, thus developed, is believed to be useful in R&D of natural language processing and the next generation of knowledge processing systems. In addition, it will become part of an infrastructure that provides new types of activities in information services.

A total of 1,742 EDR electronic dictionary packages have been sold as of this writing. On October 1, 2003, NICT released updated versions of the dictionaries. We are currently revising them for their next release.

2.2 Concept dictionary

The concepts of the EDR electronic dictionary are uniquely identified and are the same in each of the dictionaries that compose the EDR electronic dictionary. The record of a concept in the dictionary comprises the record number, the concept identifier, the headconcept, the concept explication, and management information. The concept identifier is indicated by a hexadecimal number and is the method of identifying the concept. When a concept is described by assigning a word to represent the concept, the concept assumes a physical shape or becomes more concrete. The headconcept is the word assigned to the concept. Because a headconcept may be assigned to more than one concept identifier, a one-to-one correspondence between the concept identifier and the headconcept is not necessarily assured. Moreover, to facilitate human understanding of the concept, the concept is explained with a phrase. This phrase is the *concept explication*.

The concept classification of the EDR electronic dictionary can be viewed with the dictionary browsers, ED-BroW and EDRCPCview. EDBrow is the browser attached to the EDR electronic dictionary. We are currently developing the EDRCPCview browser with a more user-friendly interface. When a word or a concept identifier is input, both browsers output bottom-up hierarchical structures expressing the upper-concept classification for

the input. Fig. 1 shows the hierarchical structure for the word "dog" in EDRCPCview.

On the browser, concepts are indicated such as "3bdc67: an animal called dog [dog, doggy, pooch and …]" in the eleventh line of Fig. 1. There are the concept identifier "3bdc67," the concept explication "animal called dog," and words such as "dog" and "doggy." Hereafter, we call the list of these words the *wordlist of the concept*. Thus, from this output hierarchical structure, for example, we can find the above concept that can express with the word "dog" has the upper-concept "444d86: thing and matter (thing)" around the top layers.

3. Approach

What are concepts unsuitable to the class? We believe that an unsuitable concept differs from other concepts classified into the same class from an abstract perspective. If an upper-lower relation exists between concepts, the lower-concept succeeds to the attributes of the upper-concept. In the other words, the upper-concept has all the attributes of the lower-concept; that is, no exclusive attributes exist between the upper-concept and lower-concept. Therefore, for example, no upper-lower relation exists between the lower-concept of "thing, matter" and the lower-concept of "event" because the concept hierarchy breaks up into the concept "thing, matter" and the concept "event" at the top layer. In Japanese, the concept explained by – *suru mono* ("a thing which does –") cannot be an upper-concept of the concept explained by – *suru koto* ("an act of –ing"), and vice versa. They are the lower-concept of "thing, matter" and the lower-concept of "event," respectively.

Suppose "*XX*" is a noun in Japanese and there exists a concept with a concept explication "*XX*." If a concept exists with a concept explication "*modifier + XX*" — that is, *XX* restricted semantically by the modifier — they probably have the upper-lower relation that the former (e.g., box) is an upper-concept of the latter (e.g., shoe box). For example, the Japanese word *eiyu* ("hero") is a word in the wordlist of the concept *toku ni sugureta jinbutsu* ("an especially outstanding person"). In the concept dictionary, there are concepts wherein the concept explication has the noun *eiyu* at the end, such as *chikara no hakuchu shita futari no eiyu* ("two heroes who have similarly great abilities"), and *warudie ni taketa eiyu* ("a hero who is a master at cunningness"). They should be linked to the concept with *eiyu* in its wordlist as its lower-concepts; in other words, they should be classified into the class defined by the concept that can express with *eiyu*.

From these considerations, we try to extract the concepts not suited to their current class and rearrange them in a suitable class.

However, the qualities of concept explications in the EDR electronic dictionary are not guaranteed because the consistency of their descriptions has not been verified, though they are phrases created by several lexicographers. Such concept explications should be amended to suitable phrases for the class. For the concept explications of concepts extracted as unsuitable concepts for their current classes, we investigate amending them to suitable phrases for their classes.

---

[1] Fujitsu, Ltd., NEC Corporation, Hitachi, Ltd., Sharp Corporation, Toshiba Corporation, Oki Electric Industry Co., Ltd., Mitsubishi Electric Corporation, and Matsushita Electric Industrial Co., Ltd.
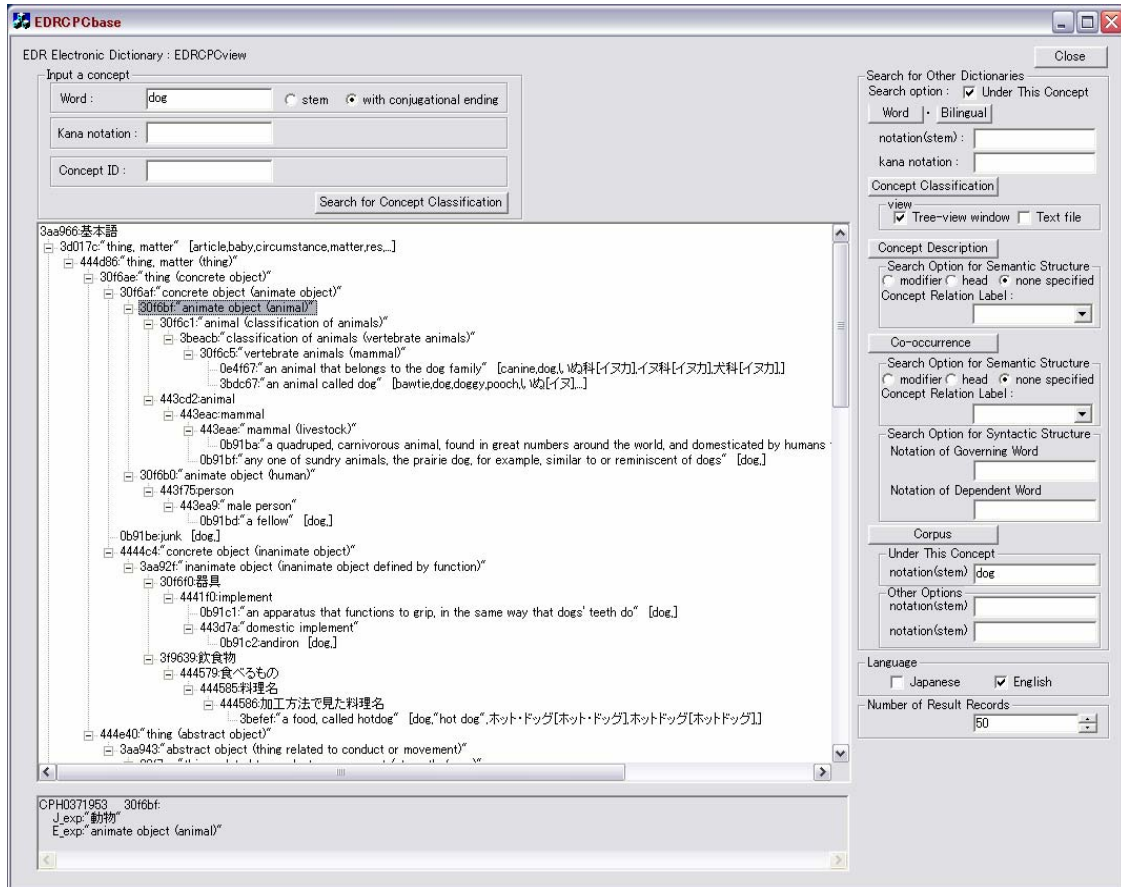
Figure 1: The layer structures for the word "dog"

4. Experimental data

To verify the applicability of our approach, we focused on the hierarchies for the concept *ningen* ("human") having the upper-concept *monogoto* ("thing and matter") at the top layer, as an example. We then gathered as our experimental data the lower-concepts directly linked to the concept – *de toraeta ningen* ("person as defined by –"). The number of concepts coming under this pattern – *de toraeta ningen* is 92. The total number of the gathered

lower-concepts is 5502. Some of the 67 lower-concepts directly linked to the concept *shintai-kino no jotai de toraeta ningen* ("person viewed from their state of physical health") are shown as examples in Fig. 2.

5. Revision method

In this section, we show our method for revising the arrangement of concepts in the EDR concept dictionary.

5.1 Process

We revise the arrangement of concepts from the top of the hierarchies. We compare a concept with its directly lower concept. As an example, we focused on the concept *ningen* ("human") and gathered the lower concepts directly linked to the concept – *de toraeta ningen* ("person as defined by –"), which are the lower-concepts of the concept *ningen.*

Step 1. Among the directly lower concepts of the focused concepts, collect concepts which do not inherit attributes of the focused concepts.

We analyze the concept explications with the Japanese morphological analysis system ChaSen (Matsumoto et al., 2001) and extract the noun at the end of each concept explication described by a noun phrase. In our experimental data, we collected many nouns such as *hito* ("person") at the end of the concept explication *kyoshi to_iu shokugyo no hito* ("a person who is a teacher"). In the rest of this paper, we will describe such a noun as the *last noun*.

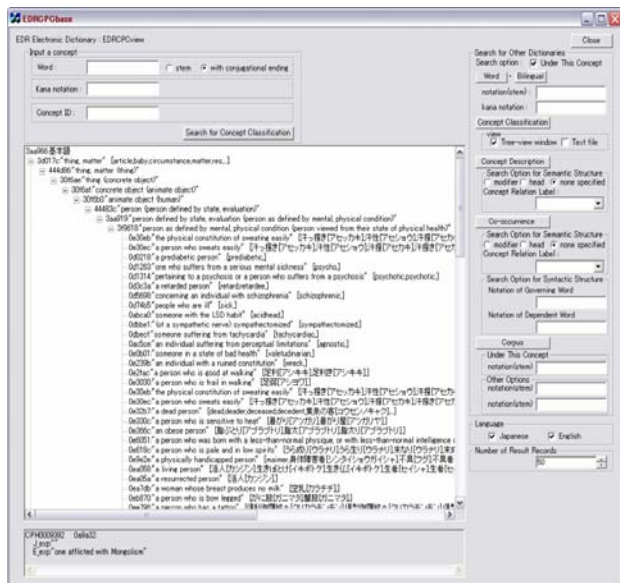We collect the concepts whose last noun is not in-

Figure 2: Example of the gathered lower-concepts

volved in the wordlist of the lower concepts of the concept *ningen* ("human"). For example, the lower-concept *ase_kkaki to_iu taishitsu* ("the physical constitution of sweating easily"), which is linked to the concept *shintai_kino no jotai de toraeta ningen* ("person viewed from their state of physical health"), has its last noun *taishitsu* ("physical constitution"). We collect this concept because the concept hierarchies of the concepts having the last noun *taishitsu* in their wordlists do not agree with the hierarchy of our focused concept *ningen*.

Step 2. To revise the handmade concept hierarchies in the EDR concept dictionary, for all the concepts collected in Step 1**,** amend the concept explications and rearrange the concepts to the suitable positions.

Step 2-1. Amendment of the concept explications and rearrangement of the concepts

Because some of the concept explications do not suit the concept hierarchy, we try to amend the explications by simple substitutions.

a) We gather the concept explication from the concepts collected in Step 1 in which the last noun has a high degree of abstraction, e.g., concept explications having abstract nouns such as *koto* ("matter") and *ichi* ("position") at the end, and amend the concept explications by adding a typical word that expresses the focused concept to the tail of the concept explications, replacing the last noun with the typical word, or deleting the last noun.

This is the process for amending the concept explications to a suitable phrase to the class. For example, there is a concept whose concept identifier is "0f3824" and whose concept explication is *yakyu ni_oite dageki san_bumon de ichi_i wo shimeru <u>koto</u>* ("in baseball, the act of winning a Triple Crown award"). This explication can be amended to *yakyu ni_oite dageki san_bumon de ichi_i wo shimeru <u>hito</u>* ("in baseball, <u>a winner of</u> a Triple Crown award") by replacing the noun *koto* ("matter") with the typical word *hito* ("person") that expresses the concept *ningen* ("human"). In this case, the amended phrase becomes a suitable concept explication to be classified into the class. In the case of the concept explication *oganemochi* ("the state of being very rich") of the concept "ofb272," it can also be amended to *oganemochi <u>no hito</u>* ("a person who is the state of being very rich") by adding *no hito* ("a person of –," "a person who –"). On the other hand, the concept explication *kijin no wakai danshi tachi* ("young noblemen") can be amended to *kijin no wakai danshi* ("a young nobleman") by deleting *tachi*, which is a Japanese suffix representing a plurality of the previous word because the singular noun *danshi* ("man") has the same concept as the plural noun *danshi tachi* ("men").

b) If amended concept explication does not exist in the EDR concept dictionary, create a new concept, and therefore its concept identifier, with the amended concept explication and substitute the new concept for the original concept in the class. Create links from all lower concepts of the original concept to the new concept. If amended concept explication exists in a concept in the EDR

concept dictionary, create links from all lower concepts of the original concept to the concept.

Because the above amended phrase *yakyu ni_oite dageki san_bumon de ichi_i wo shimeru <u>hito</u>* does not exist in the amended explication, arrange it instead of the original concept that the last noun is *<u>koto</u>* in the class. The concept explication *onaji gakko de mananda <u>koto</u>* ("the act of having studied at the same school") can be also amended to *onaji gakko de mananda <u>hito</u>* ("a person who has studied at the same school"). However, a different concept having the amended phrase as the concept explication already exists in the dictionary. Then, we simply add new links from all the lower concepts of the original concept to the concept with the amended phrase *onaji gakko de mananda <u>hito</u>*.

c) If the original concept does not have an upper concept except the concept – *de toraeta ningen*, we rearrange the original concept under the concept expressed by the last noun of the original concept, retaining all links from the lower concepts. If the original concept has at least one upper concept in addition to the concept – *de toraeta ningen*, we simply remove the link to the concept – *de toraeta ningen*.

Step 2-2. Rearrangement of the concepts without amendment of the concept explications

For the concepts whose concept explications are not amended, if the concept does not have an upper-concept except the concept – *de toraeta ningen*, we rearrange the concept under the concept having the last noun in its wordlist, retaining all links from the lower concepts. If the concept has at least one upper-concept in addition to the concept – *de toraeta ningen*, we simply remove the link to the concept – *de toraeta ningen*.

Because the concept *ase_kkaki to_iu taishitsu* does not have an upper-concept except the concept *shintai_kino no jotai de toraeta ningen*, we rearrange this concept under the concept having the last noun *taishitsu* in its wordlist, retaining all links from the lower concepts.

Step 3. Have lexicographers check the revised arrangement of concepts.

Our method extracts the candidates of concepts which should be amended and/or rearranged. When we try to rearrange the concept under some concept, exact place of rearrangement can not be decided automatically. Therefore, the check by lexicographers is indispensable. Because above steps are repeated from the top of the hierarchies, some of the errors created in the earlier stage of the revision will be corrected.

5.2 Amending concept explications

First, for concepts extracted as not suited to their current class, we consider ways of amending the concept explication to a suitable phrase to the class in a simple way as shown in Step 2-1. This is intended to maintain the concept hierarchy in the EDR electronic dictionary.

For example, our experimental data contain concept explications having the abstract noun *koto* ("matter") or *mono* ("thing") at the end. They can be viewed as lower-concepts of the upper-concepts such as *jisho*

("event") and *jotai* ("condition of event"). In Japanese, when the last nouns *koto* and *mono* in such concept explications are simply replaced with *hito* ("person"), they can become suitable concept explications for concepts in which the upper-concept is the concept *ningen* ("human"). We investigate giving a concept with such an amended concept explication a new concept identifier and arranging it instead of the concept with the original concept explication.

We can automatically amend concept explications, once the last nouns used for the amendment are selected. Also, it is easy to automatically check whether the EDR concept dictionary contains phrases we amended or others similar to them. For example, the concept explication *namae ga onaji de betsu_jin de aru koto* ("the situation in which different people have the same name") can be amended to *namae ga onaji de betsu_jin de aru hito* ("a person who is a different person has the same name") and the similar phrase *namae ga onaji de betsu no hito* ("a different person with the same name") already exists as a concept explication in the dictionary. We recognize these two phrases as same phrase. Then, we remove from the class the original concept whose last noun is *koto*, and add new links from all lower-concepts of this original concept to the existing concept with similar phrase to the amended phrase. .

From the above considerations, we try to amend concept explications to revise the arrangement of the concepts.

5.3 Rearranging concepts

We try to rearrange under a suitable upper-concept the collected concepts that are not amended. It is easy to remove the link from a concept to an upper-concept that corresponds to an unsuitable class. However, we try not only to remove such links but also to rearrange them under a suitable upper-concept based on the concept explication.

We determine that a concept is not suited to its class if the hierarchy of the concept with the last noun in its wordlist does not agree with the hierarchy of the focused concept as shown in Step 2-2. This is based on the hypothesis that the head noun in a concept explication determines the attribute of the concept. The attribute of a new word tends to succeed to the attribute of its head noun. This succession between its head noun and the new word indicates a hypernym-hyponym relation between them. In Japanese, the new word is either a noun phrase or a compound noun and the head nouns tend to be at the end of the noun phrase or of the compound noun. The concept explication in the EDR concept dictionary we use is a noun phrase or a noun. According to these heuristics, for each concept that needs to be rearranged, we supposed the noun at the end of the concept explication (*last noun*) is the head noun and determined that the concept with the head noun in its wordlist is an upper-concept of the concept with this concept explication. We then rearrange the concept not suited to the current class under the upper-concept thus determined.

6. Experimental result

The number of concepts in our experimental data is 5502, where the total number of concepts in the EDR electronic dictionary exceeds 400,000. As shown in Table 1, among the 5502 concepts, there exist 492 kinds of last nouns.

Among the 492 last nouns, there are 87 last nouns where the hierarchy of the concept with such a noun in its wordlist does not agree with the hierarchy of the focused concept *ningen* ("human"). There are 244 concepts in which one of the 87 nouns is the last noun in the concept explication. Such concepts correspond to about 4.4% of the experimental data. We revised such concepts by amending their concept explications or rearranging them under another concept.

Concept	Number	No. of last nouns
Under – *de traeta ningen*	5502	492
Not suited to the class	244	87
Amended	59	10
Rearranged	185	77

Table 1: Number of revised concepts.

As shown in Table 1, we amended concept explications for 59 of the 244 concepts. We selected 10 last nouns by hand and defined three amendment rules as follows.

Rule 1. Replace *koto* (matter), *mono* (thing), and *ichiin* (member) with *hito* (person).

Rule 2. Add *no_hito* (person of) to *kyoto* (believer), *ichi* (position, place), *saikai* (booby), *okanemochi* (millionaire), *chii* (position, rank), and *ryosaido* (both sides).

Rule 3. Delete *no_koto* (matter of) and *tachi* (a Japanese suffix representing a plurality of the previous word; i.e., the last letter 's'of "persons").

As a result, Rule 1 was applied to 27 of the 59 concepts. Rules 2 and 3 were each applied to 16 concepts. We show examples of the amended concept explication in Table 2. We did not create a concept with the amended phrase in the first example because this concept exists in the EDR concept dictionary.

meiyo_kyoju de aru koto
    ("the condition which is a professor emeritus")
→ *meiyo_kyoju de aru hito*
    ("a person who is a professor emeritus")

ta ni masatte sugurete_iru mono
    ("a thing which is superior to others")
→ *ta ni masatte sugurete_iru hito*
    ("a person who is superior to others")

gekidan no ichiin ("the member of the dramatic company") → *gekidan no hito* ("the person of the dramatic company")
hoka no hito ni eikyo_ryoku no-aru shohi_kodo wo suru hito tachi ("consumers who influence others") → *hoka no hito ni eikyo_ryoku no-aru shohi_kodo wo suru hito*   ("a consumer who influences others")

jinja ni tsukaeru shinjin no osa to_iu chii
    ("the position of chief priest of a Shinto shrine")
→ *jinja ni tsukaeru shinjin no osa to_iu chii no hito*
    ("a person who is a chief priest of a Shinto shrine")

Table 2: Examples of amended concept explications.

We show examples of the rearranged concept in Table 3, where the pattern of the concept explications is shown for each last noun. We determined the concepts that have to be rearranged by their last nouns and supposed that a hierarchical relation — such as a hypernym-hyponym relation — exists between the concept for its last noun and the concept. That is, when a last noun is *XX*, we defined the concepts with the concept explications *modifier + XX*, such as – *to_iu XX*, – *no XX*, and – *XX*, as the lower-concepts of the concept having *XX* in its wordlist.

From this relation, they can be rearranged under the concept of *XX*. In our experiment, except the 10 last nouns used for the amendment, the rest 77 last nouns can be used for rearranging the 185 concepts with them. However, it is necessary to be checked by lexicographers before rearranging them.

Upper-concept (last noun)	Rearranged concept as lower-concept
jinshu bunrui jo_no tani to_shite_no shuzoku ("a race as a unit of classification of human beings") (*minzoku* ("affinal"))	– *to_iu minzoku* – *yuboku minzoku* – *senjyu minzoku*
uke_motteiru shigoto ("assigned work that must be carried out") (*shokumu* ("duty"))	– *to_iu shokumu* – *to_shite_no shokumu* – *no shokumu*
sorezore no hito ni wariaterareta yakume ("duties that are assigned to each respective person") (*yakuwari* ("cue"))	– *to_iu yakuwari* – *suru yakuwari*
hito ya monogoto no kakawariai ("a relationship between people or things") (*kankei* ("relation"))	– *to_iu kankei*

Table 3: Examples of rearranged concepts.

7. Discussion

In our experiment, we apply our method of revising concepts that are the lower-concepts of the concept "*ningen* (human)" and are directly linked to the concept – *de toraeta ningen* ("a person as defined by –", "a person viewed from –", etc.).

First, in determining the concepts whose concept explications can be amended, we manually selected the last nouns and amended them automatically. We discovered one phrase that failed to be amended. It is "*noke hito*," a phrase whose concept explication *noke mono* ("outcast") was amended by Rule 1; it is a mere string that is neither phrase nor compound noun. We found that, in order to avoid creating a concept with an incorrect phrase like this as concept explication, we need to manually check whether the amended phrases are correct Japanese after the amendment.

Next, we considered the positions to which the concepts were rearranged. In this paper, we defined the position under the concept having the last noun in its wordlist as a suitable position based on the heuristics in Japanese; that is, noun phrases and compound nouns succeed the attribute of the head noun and the head noun tends to be at the end of them. Actually, we found that some concepts are linked to the upper-concept, based on the same notion as ours. For example, the noun *yakuwari* ("cue") is a word in the wordlist of the concept *sorezore no hito ni wariaterareta yakume* ("duties that are assigned to each respective person"), and the hierarchy of its concept does not agree with the hierarchy of the focused concept *ningen* ("human"). Under this concept, the concepts with the concept explication *modifier + yakuwari* such as *sono hito ni yoku tekishita shigoto no_ue nado de_no yakuwari* ("a job that is well suited to a person") and *shuyona yakuwari* ("the leading role") are arranged as the lower-concepts. We often found such hierarchical structures in the EDR electronic dictionary. However, we also found upper-concepts that do not have any such lower-concepts. For the concepts we thus defined as upper-concepts, we will investigate whether there are any concepts in the EDR electronic dictionary that can be their lower-concepts ac-

cording to our notion; or, if such concepts do exist, whether they have been arranged under a concept we define as one of their upper-concepts. However, we found that our method can at least maintain consistency with the concept hierarchical structure in the EDR electronic dictionary.

Previously, we tried to extract from corpora hierarchies of abstract nouns, which are defined as hypernyms of adjectives (Kanzaki et al., 2004; Yamamoto et al., 2004; Yamamoto et al., 2005). As a result, we found differences between the extracted hierarchies and those of adjectives from around the middle layers of the EDR electronic dictionary. We might be able to use this difference to revise the EDR electronic dictionary. This approach is considered one method of revision with information from actual data such as a corpus; that is, other information that is not built into the dictionary. Our future work will include the application of this approach to the revision of the EDR electronic dictionary.

8. Conclusion

In an effort to improve the EDR electronic dictionary, we tried to revise the arrangement of the concepts in the concept dictionary. To do this, we automatically extracted the concepts not suited to the class. We then amended some of the concept explications that explain their meanings and rearranged the rest of them in a suitable class, semi-automatically. We revised the arrangement of the concepts that are the lower-concepts in the hierarchy of the concept *ningen* ("human") and that are linked to the concepts with the concept explications – *de toraeta ningen* ("person as defined by –", "person viewed from –", etc.). As a result, we found our method is capable of revising the concepts while at least maintaining consistency with the concept hierarchical structure in the EDR electronic dictionary.

References

EDR electronic dictionary (1995). http://www2.nict.go.jp/kk/e416/EDR/index.html

Kanzaki, K., Yamamoto, E., Ma, Q. and Isahara, H. (2004). Construction of an objective hierarchy of abstract concepts via directional similarity. In Proceedings of the 20th International Conference on Computational Linguistics (Coling2004), Vol.2, 1147-1153.

Matsumoto, Y, Kitauchi, A, Yamashita, T, Hirano, Y, Matsuda, H, Takaoka, K, and Asahara, M. (2000). Japanese morphological analysis system ChaSen version 2.2.1.

Renouf, A. (1987). Corpus development. In Sinclair (Ed.), Looking Up (pp.1-40).

Sinclair, J. (1987). Looking Up. London: Collins ELT.

Yamamoto, E., Kanzaki, K. and Isahara, H. (2004) Hierarchy extraction based on inclusion of appearance. In ACL04 Companion Volume to the Proceedings of the Conference, 149--152.

Yamamoto, E., Kanzaki, K. and Isahara, H. (2005). Extraction of hierarchies based on inclusion of co-occurring words with frequency information. In Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI05), 1166-1172.