Objective Hierarchy of Abstract Concepts
- Organization of Abstract Nouns via Distribution of Adjectives -

Kyoko Kanzaki, Qing Ma, Eiko Yamamoto, Masaki Murata and Hitoshi Isahara

Communications Research Laboratory
Keihanna human Info-Communications
Research Center,
3-5 Hikaridai, Seika-cho, Soraku-gun, Kyoto,
619-0289, Japan
{kanzaki, eiko, murata, isahara} @crl.go.jp

Ryukoku University
Faculty of Science and Technology
Seta, Otsu, 520-2194,Japan
qma@math.ryukoku.ac.jp

Abstract

Our purpose in this research is to find an objective way to organize word meanings by using large corpora. At first we are treating adjectives which have complicated meanings. We focus on the semantic relations between abstract nouns and adjectives. And then we construct a Semantic Map of abstract nouns based on the classification of adjectives by using a Self Organizing Semantic Map (SOM). We use the CSM(Complementary Similarity Measure) as their input data of SOM. We can see the SOM from two viewpoints. One is a position of an abstract noun of all abstract nouns and another is a superordinate/subordinate relation. Finally we explain the aspect of the superordinate abstract concepts that all adjectives have commonly.

1 Introduction

Our purpose in this research is to find hypernimic concepts of words experimentally by using a large corpora and a neural network model. At first we treat adjectives. We focused on semantic relations between abstract nouns and adjectives. We made linguistic data by extracting semantic relations between abstract nouns and adjectives from large corpora, and we use them as an input data for the Self-Organizing Semantic Map (SOM), which is a neural network model (Kohonen 1995). On the Semantic Map, words are located near to or far from each other depending on their similarities. In some previous research, word meanings were classified from linguistic data based on syntactic information using a statistical method. Hindle (1990) used syntactic relations between nouns and verbs, and Hatzivassiloglou and McKeown (1993) used semantic relations between adjective-adjective pairs, both of which are modifiers for a head noun. Their methods are useful for an organization of words in a subordinate layer, however, if we consider it in a superordinate layer, their methods seem not to be enough to solve the problem. Previous researches on the classification of words by a SOM treat a small amount of data (Kohonen 1995, et.al), however, we made an experiment with a large amount of data, such as 42 years worth of newspaper articles.

In our previous experience of creating a SOM, we calculated the similarities of words using simple Euclid distance. We could get meaningful distribution of words based on their meanings, however, the hierarchy (super/subordinate relation) between them was not clear. In this paper, we utilize the Complementary Similarity Measure (CSM) as a similarity measure for the creation of a SOM, which estimates an one-to-many relation, such as a super/subordinate relation (Yamamoto

and Umemura 2002). By using the CSM, we could find the hierarchical relations between words on the map, which is the first step to make an objective thesaurus based on a real data.

2 The Linguistic Clue to Find an Abstract Concept of Adjectives

Takahashi (1975) illustrated a function of an abstract noun in a sentence with the following examples.

A. *Yagi wa seishitsu ga otonashii*.
(goat) topic (nature) subject (gentle)
The nature of goats is gentle
B. *Zou wa hana ga nagai*.
(elephant) topic (nose) subject (long)
The nose of an elephant is long

He found out the difference of a semantic function between "*seishitsu* (nature)" in Example A and "*hana* (nose)" in Example B, and explained that an abstract noun in Example A, i.e., "*seishitsu* (nature)", indicates a side of something, i.e., a goat, and "*hana* (nose)" in Example B indicates a part of something, i.e., an elephant. He recognized an abstract noun in Example A as a superordinate concept of an attribute that an adjective in predicate position expresses. Isahara and Kanzaki (1999) classified adnominal relationships between adjectives and their head nouns, and found two kinds of semantic relations among various adnominal relations between adjectives and their head nouns, which are similar to the above. One is a type like "*yuruyakana keisha* (gentle slope)", which we can paraphrase into a phrase whose head noun is an abstract noun, i.e., "*keisha ga yuruyakada* (the slope is gentle)", and another is a type like "*kanashii kimochi* (sad feeling)", which cannot be paraphrased into "*kimochi ga kanashii* (the feeling is sad)". An adjective, "sad", is a description of the head noun "feeling", a cognate noun. In these types of adnominal relations the head noun is identified as an abstract concept like a superordinate concept for an adjective. We consider that this pattern is useful for finding abstract concepts of adjectives in real data, i.e., a corpus.

3 Data

We extracted abstract nouns from two years of a newspaper, the Mainichi Shinbun, and extracted adjectives co-occurring with abstract nouns from 42 years of newspapers, including 11 years of the Mainichi Shinbun, 10 years of the Nihon Keizai Shinbun (Japanese economic newspaper), 7 years of the Sangyou-kinyu-ryutusu shinbun (a kind of economic newspaper), and 14 years of the Yomiuri Shinbun, 100 novels and 100 essays. The number of different abstract nouns is 365, the number of different adjectives is 10,525, and the total number of adjectives is 35,173. The maximum number of co-occurring adjectives for one abstract noun is 1594. We made the list of collocations like the following from the corpus.

OMOI (feeling):
ureshii (glad), *kanashii* (sad)
shiawasena (happy), …
KIMOCHI (though) :
ureshii (glad), *tanoshii* (pleased)
hokorashii (proud), …
KANTEN (viewpoint):
igakutekina (medical),
rekishitekina (historical), ...

A group of co-occurring adjectives is regarded as a definition of an abstract noun in our input data.

4 Encoding

It is necessary to encode the above linguistic data in order to put it into a SOM as the input data (Ma 2000). We use the two methods of similarity measurement for encoding the input data of a SOM. One is that we calculate and normalize the similarity in the degree of the

intersection of adjectives commonly appearing with two abstract nouns. Another is that we used the Complementary Similarity Measure (CSM) as a similarity measure for a SOM, which estimates an one-to-many relation, such as a superordinate–subordinate relation (Yamamoto and Umemura 2002). The reason for introducing the CSM is that in the case of the SOM using the first similarity measure for the input data, that is, the normalized value of the degree of intersection of adjectives between two head nouns, the clue for analyzing the distribution of words on the Semantic Map is just the location itself. It is possible to find a rough sketch of the similarities between nouns based on the visual image of the distribution of nouns on the Semantic Map. However, we had to classify words on a Semantic Map into groups of similar words manually, that is, by our intuition. CSM is a good similarity measure for finding the relation between two words such as a super/subordinate relation and a similar relation. The similarity calculated via the CSM is a numerical value, and therefore, if the superordinate-subordinate relation can be reflected on the Semantic Map, it is possible to understand the relation between words located on the map objectively.

4.1 Encoding by Ma (2000)

For example, we can define "*kimochi* (feeling)" as the set of its adnominal constituents, i.e. "*kimochi*" = {"*shiawasena* (happy)", "*hokorashii* (proud)", "*kanashii* (sad)", "*kinodokuna* (unfortunate)"…}. Suppose there is a set of nouns w_i ($i = 1, \ldots, \omega$) that we are planning to use for self-organizing. Any noun w_i can be defined by a set of its adnominal constituents as

$$w_i = \{ a_1^{(i)}, a_2^{(i)}, \ldots, a_{\alpha i}^{(i)} \} \quad \text{-------(1)}$$

where $a_j^{(i)}$ is the jth adnominal constituent of w_i and α_i is the number of adnominal constituents of w_i. One method of encoding nouns so that they can be treated by a SOM is to use random coding, which is a common method used for constructing SOMs (see details in Kohonen (1995)). By several preceding computer experiments, however, we found that this method is not suitable for our task. We therefore used a new method as described below.

Suppose we have a correlative matrix where d_{ij} is some metric of correlation (or distance) between nouns w_i and w_j. We can encode noun w_i from the correlative matrix as

$$V(w_i) = [d_{i1}, d_{i2}, \ldots, d_{i\omega}]^T. \quad \text{---------(2)}$$

The $V(w_i) \in \Re^{\omega}$ is the input to the SOM. In this paper, d_{ij} is measured by

$$d_{ij} = \begin{cases} \dfrac{(\alpha_i - c_{ij}) + (\alpha_j - c_{ij})}{\alpha_i + \alpha_j - c_{ij}} & \text{If } i \neq j \\ 0, & \end{cases} \quad \text{-------(3)}$$

where α_i and α_j are respectively the numbers of the adnominal constituents of w_i and w_j, and c_{ij} is the total number of common adnominal constituents of both w_i and w_j. The term d_{ij} is therefore a normalized distance between w_i and w_j in the context of the number of adnominal constituents they have in common; the smaller d_{ij} is, the closer w_i and w_j are in terms of their adnominal constituents.

4.2 Encoding by using CSM

The CSM (Yamamoto and Umemura 2002) is a similarity measure for finding one-to-many relations. According to Yamamoto and Umemura (2002), the similarity value of the CSM is calculated as follows.

If
$\vec{F} = (f1, f2, \ldots, fi, \ldots, fn)(fi = 0 or 1)$,
$\vec{T} = (t1, t2, \ldots, ti, \ldots tn)(ti = 0 or 1)$,

$$Sc(\vec{F}, \vec{T}) = \frac{ad - bc}{\sqrt{(a+c)(b+d)}}$$

Here "a" indicates the number of the data in which two labels appear together, "b" the number of the data in which "label 1" appears

but “label 2” does not appear, “c” the number of the data in which “label 1” does not appear but “label 2” appears, and “d” the number of the data in which neither labels appears. In the case of our research, a “label” refers to an abstract noun and “a” indicates the number of adjectives co-occurring with both abstract nouns, “b” and “c” indicate the number of adjectives co-occurring with either abstract noun, “label 1” or “label 2”, and “d” indicates the number of adjectives co-occurring with neither abstract noun. After the similarity value between two abstract nouns that was calculated by CSM is normalized, we made the matrix and then encoded all nouns from the correlative matrix as we mentioned above.

5 Semantic Map

The learning step of a SOM consists of an ordering phase and a final phase. The number of learning steps is 30,000 in the ordering phase and 100,000 in the final phase. Here is the SOM of a 45*45 array in which a hexagonal topology type of neighborhood is used.

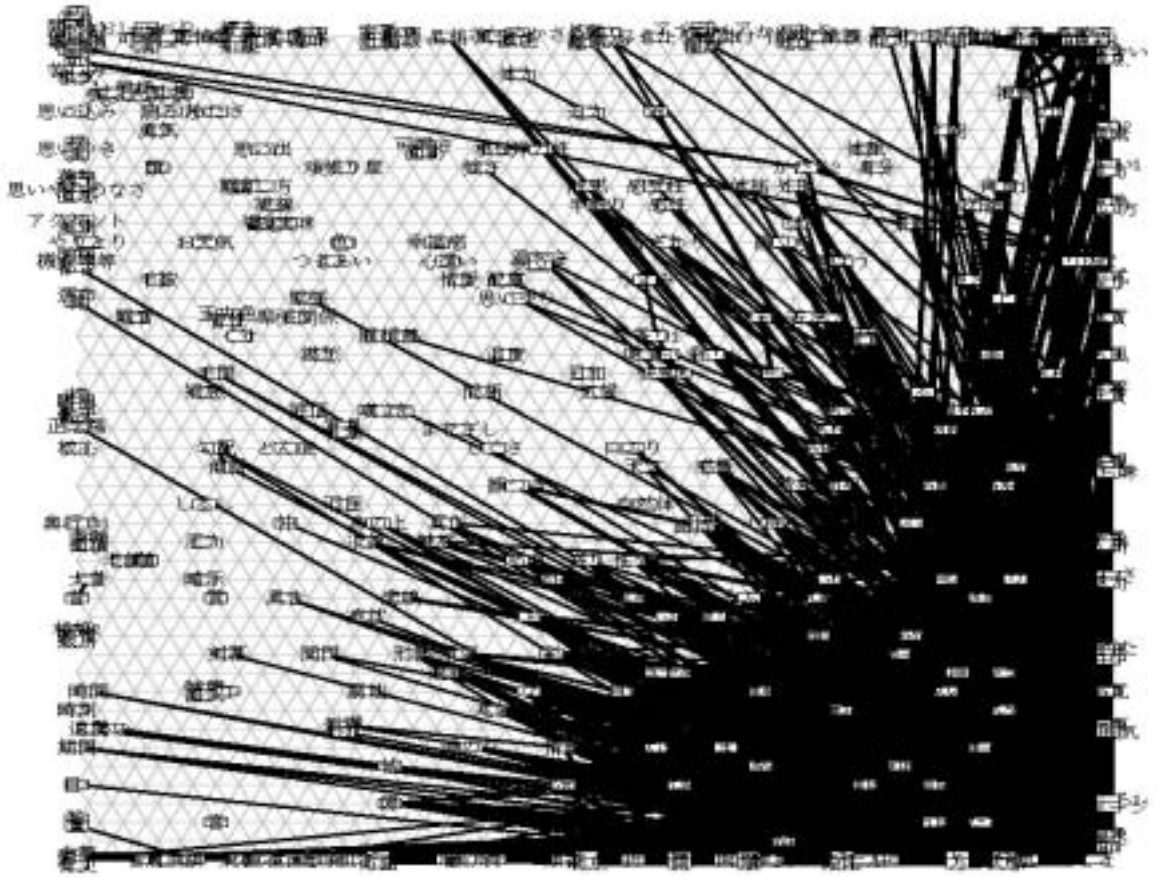

Fig 1. Map of abstract nouns whose CSM values are 1.0-0.3.

The following figure is the Semantic Map of abstract nouns. In this figure, abstract nouns with a CSM value between 1.0-0.3 are linked by lines. On the Semantic Map, the abstract nouns are distributed from the high value of CSM to the low value. The abstract nouns spread out from the bottom right-hand corner radially.

6. A Classification of abstract nouns until Normalized Value, 0.5

In this section, we explain the distribution of the abstract nouns on the map according to the normalized value of the CSM. The lines between words are linked automatically. The similarity value that we show step by step is from 1.0 until 0.5 because the superordinate abstract nouns related to most adjectives almost appear between the normalized value 1.0 and 0.5.

First, we explain what the normalized value indicates. A value of word A for word B (we call it the value of AB) and a value of word B for word A (we call it the value of BA) are different, i.e., asymmetric. If word A and word B are in an entirely superordinate/suborinate relation, the difference in the value of AB and BA is big. If word A is a superordinate word and word B is a subordinate word, the relation between two is "*a value of AB* > *a value of BA*", i.e., in respect to the common co-occurring adjectives, word A is similar to word B, but word B is not so similar to word A. If word B is interchangeable with word A, "*a value of AB = a value of BA*", i.e., the number of the adjectives co-occurring with word B and the number of adjectives co-occurring with word A are the same. If the values of both AB and BA are high, the number of common co-occurring adjectives is large. On the other hand, if the values of both AB and BA are low, the common co-occurring adjectives are few.

6.1. The super/subordinate between abstract nouns according to the CSM value

We describe the relations between abstract nouns basically according to a normalized value; however, we sometimes explain them by referring to another normalized value.
We regard a value of more than 0.1 as a super/subordinate relation.

The Normalized Value: 1.0- 0.9

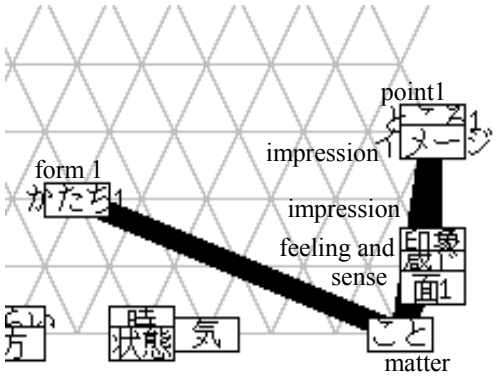

Fig 2. Abstract nouns of 1.0-0.90

Word A	Word B	CSM(A:B)	CSM(B:A)	Difference
imeji (impression)	inshou (impression)	1.0	0.951	0.049
koto (matter)	katachi1 (form1)	0.955	0.605	0.35
inshou (impression)	kanji (feeling and /or sense)	0.936	0.778	0.158
koto (matter)	kanji (feeling and/ or sense)	0.925	0.694	0.231
koto (matter)	tokoro 1 (point1)	0.910	0.795	0.115

At first, "*koto* (matter)", "*kanji* (feeling and/or sense)", "*inshou* (impression)", "*ime-ji* (impression)", "*tokoro*1 (point)", and "*katachi*1 (form1)" are located at the bottom right-hand corner on the map.

The meaning of "*imeji* (impression) " which is a loanword, is almost the same as "*inshou* (impression)". The values of both directions of CSM for these words are almost same. The difference is 0.049.

"*Kanji* (feeling and/or sense)" is a subordinate noun of "*inshou* (impression)" and "*koto* (things)" because the difference of CSM("*inshou*" : "*kanji*") and CSM("*kanji*" : "*inshou*") is 0.158 and the difference of CSM("*koto*" : "*kanji*") and CSM("*kanji*" : "*koto*") is 0.231.

"*Tokoro*1 (point)" and "*katachi*1 (form1)" are subordinate nouns of "*koto* (matter)" because the differences are 0.115 and 0.35. The example of "*tokoro*1(point)" is "*Kare wa*

yasashii tokoro ga aru. (He has a kind side (point)) " and the example of "*katachi*1 (form1)" is "*Koushou wa tyutohanpana*(halves) *katachi* (form) *de*(by) *owatta* (The negotiation was finished *by halves*)".

After all, at the normalized value of 0.7 we can find that "*koto* (matter)" and "*inshou* (impression)" are similar from the difference of their CSM values. "*Koto* (matter)" and "*inshou* (impression)", like a seed of adjectives, can co-occur with most adjectives. From these two words all abstract nouns branch off and their co-occurring adjectives become distinct gradually like a sprouting seed.

The Normalized Value: 0.8

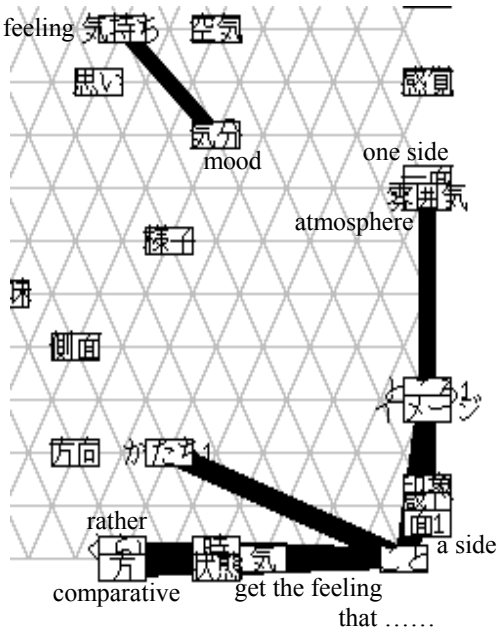

Fig3. Abstract nouns of 1.0-0.80

In this level, a mental state like "*kimochi* (feeling)", "*kibun* (mood)" and "*ki* (get the feeling that)", a state like "*toki* (when)" and "*joutai* (state)", a certain point that someone chooses in the comparison like "*hou* (comparatively)" and "*kurai* (rather)", an atmosphere like "*fun'iki* (atmosphere)" and a side of something like "*men*1 (a side)" and "*ichimen* (one side)" appear.

The values of "*kibun* (mood)" and "*kimochi* (feeling)" are like the following.

WordA	Word B	CSM(A:B)	CSM(B:A)	Difference
kibun (mood)	kimochi (feeling)	0.819	0.771	0.048

"*Kibun* (mood)" and "*kimochi* (feeling)" have similar co-occuring adjectives because the difference between them, 0.048. You can find a line located a little far from the bottom right hand corner on the map. "*Kimochi* (feeling)" and "*kibun* (mood)" in the neighborhood are combined with each other by the line.

"*Hou* (comparatively)", "*kurai* (rather)", "*toki* (when)" and "*joutai* (state)" are subordinates of "*koto* (matter)" at this level.

WordA	Word B	CSM(A:B)	CSM(B:A)	Difference
koto (matter)	hou (compara-tively)	0.891	0.533	0.358
koto (matter)	kurai (rather)	0.846	0.543	0.303
koto (matter)	toki (when)	0.878	0.627	0.251
koto (matter)	joutai (state)	0.813	0.616	0.197

In the lower level, we can find that the difference of CSM ("*hou*(comparative-ly)" : "*kurai* (rather)") and the difference of CSM("*toki* (when)" and "*joutai* (state)") are small as follows. "*Kurai* (rather)" and "*hou* (comparatively)" means that a degree or a state that someone choose in the comparison.

WordA	Word B	CSM(A:B)	CSM(B:A)	Difference
kurai (rather)	hou (compara tively)	0.596	0.558	0.038
joutai (state)	toki (when)	0.817	0.772	0.045

According to the CSM value, "*hou* (comparatively)" is similar to "*kurai* (rather)", and "*toki* (when)" is similar to "*joutai* (state)". In the normalized value, 0.6, we can find

"*kurai* (rather)" and "*hou* (comparatively)" are related to "*toki* (when)" and "*joutai* (state)".

At this level we can find that "*fun'iki* (atmosphere)" is a subordinate of "*ime-ji* (impression)" and "*inshou* (impression)". The value of the CSM is as follows:

WordA	Word B	CSM(A:B)	CSM(B:A)	Difference
imeji (impression)	fun'iki (atmosphere)	0.885	0.652	0.233
inshou (impression)	fun'iki (atmosphere)	0.867	0.669	0.198

As a similar word to "*inshou* (impression)" and "*ime-ji* (impression)", "*men1* (a side)" appears. "*Men*1(a side)" is an abstract noun representing one side/point of someone or something. But at first "*men*1(a side)" is combined with "*ichimen* (one side)" at this level.

WordA	Word B	CSM(A:B)	CSM(B:A)	Difference
men1 (a side)	ichimen (one side)	0.841	0.646	0.195

From the viewpoint of word meanings "*men* 1(a side)" and "*ichimen* (one side)" seem to indicate almost the same meaning, however, according to the data, "*men* 1 (a side)" is a superordinate of "*ichimen* (one side)"[1]. Then, "*men* 1 (a side)" combines with "*inshou* (impression)" and "*ime-ji* (impression)" as similar words.

WordA	Word B	CSM(A:B)	CSM(B:A)	Difference
men1 (a side)	inshou (impression)	0.801	0.760	0.041
men1 (a side)	ime-ji	0.719	0.715	0.004

"*Men* 1 (a side)" is similar to "*tokoro1* (point1)" as in the following:

WordA	Word B	CSM(A:B)	CSM(B:A)	Difference
men1 (a side)	tokoro1 (point)	0.816	0.755	0.006

[1] this is because "*men* (a side)" co-occurs with more adjectives than "*ichimen* (one side)" co-occurs; however, this may be caused by the sparseness of data.

We described that "*men* 1 (a side)" combined with "*inshou* (impression)" and "*ime-ji* (impression)" as the similar words. "*Tokoro1* (point1)" is not only a subordinate noun of "*koto* (matter)", but also is similar to "*inshou* (impression)" and "*ime-ji* (impression)" like "*men*1 (a side)". We can see the value which indicates the similarities between "*tokoro1* (point1)" and "*inshou* (impression)" in the normalized value of 0.7. In the Semantic Map, we can find that "*men* 1 (a side)" is located at the same coordinates as "*inshou* (impression)", and "*tokoro1* (point1)" is located at the same coordinates as "*ime-ji* (impression)".

As a subordinate noun of "*koto* (matter)", at this level, "*ki* (get the feeling that)" appears.

WordA	Word B	CSM(A:B)	CSM(B:A)	Difference
koto (matter)	ki (get the feeling that)	0.878	0.659	0.219

"*Ki* (get the feeling that)" is also related to "*kanji* (feeling and/or sense)" as a similar word in the normalized value of 0.7.

WordA	Word B	CSM(A:B)	CSM(B:A)	Difference
ki (get the feeling that)	kanji (feeling and/or sense)	0.732	0.730	0.002

In the normalized value of 0.9, we described that "*kanji* (feeling and/or sense)" was a subordinate noun of "*inshou* (impression)" and "*koto* (matter)". The subordinate noun of "*inshou* (impression)" and "*koto* (matter)" is not only "*kanji* (feeling and/or sense)" but also "*ki* (get the feeling that)". However, in the Semantic Map, "*ki* (get the feeling that)" is not located at the same coordinates as "*kanji* (feeling and/or sense)". This is because abstract nouns in the lower level related to "*ki* (get the feeling that)" and "*kanji* (feeling and/or sense)" are different. For example, at this level, 0.8, "*ki* (get the feeling that)" is related to "*kurai* (rather)" as a superordinate noun.

WordA	Word B	CSM(A:B)	CSM(B:A)	Difference
ki (get the feeling that)	kurai (rather)	0.801	0.670	0.231

The relation between "*kanji* (feeling and/or sense)" and "*kurai* (rather)" appears in the normalized value of 0.5. So, "*ki* (get the feeling that)" is away from "*kanji* (feeling and/or sense)" and moves near the "*kurai* (rather)".

Also in the normalized value of 0.6, we can find "*kanji* (feeling and/or sense)" and "*ki* (have got a feeling that)" are related to "*kimochi* (feeling)", "kibun (mood)" and "omoi (thought)".

The Normalized Value: 0.7

Here, new superordinate nouns do not appear, but subordinate abstract nouns appear. "*Seikaku* (character)", "*miryoku* (charm)", "*hitogara* (personality)" and "*utsukushisa* (beauty)" appear here as a subordinates of "*inshou* (impression)".
In this level "*omoi* (thought)" also appears.

WordA	Word B	CSM(A:B)	CSM(B:A)	Difference
kimochi (feeling)	omoi (thought)	0.778	0.763	0.015
kibun (mood)	omoi (thought)	0.724	0.670	0.054

The difference of CSM("*kimochi* (feeling)" : "*omoi* (thought)") is 0.015 and the difference of CSM(*kibun* (mood)" : "*omoi* (thought)") is 0.054. As the values are very low, "*omoi* (thought)", "*kimochi* (feeling)" and "*kibun* (mood)" are similar.

WordA	Word B	CSM(A:B)	CSM(B:A)	Difference
fun'iki (atmosphere)	kuuki (atmosphere)	0.755	0.664	0.091

As the difference of CSM("*fun'iki* (atmosphere)" : "*kuuki* (atmosphere)") is 0.091, they are almost super/subordinations. And "*kuuki* (atmosphere)" is located near "*kimochi* (feeling)".

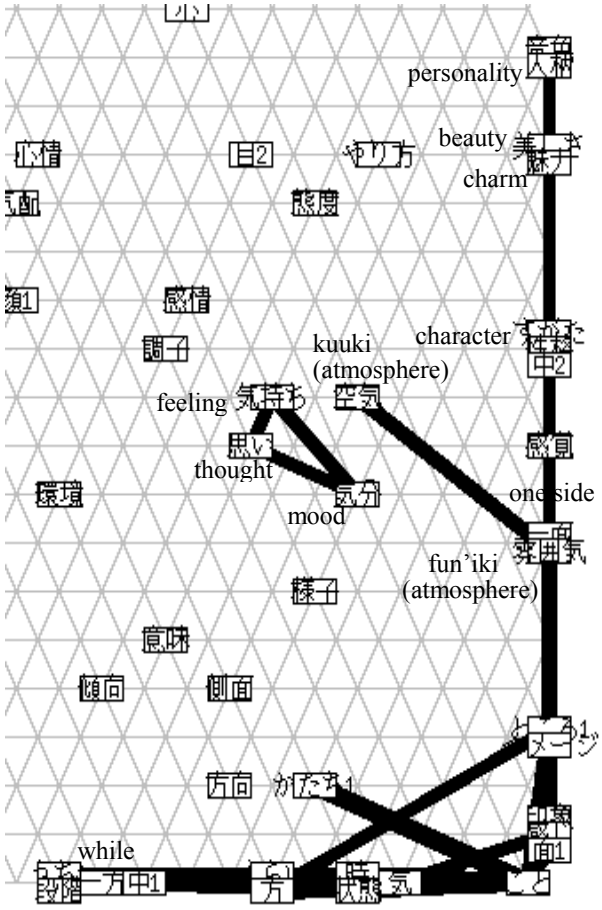

Fig4. Abstract nouns of 1.0-0.70

By this location of "*kuuki* (atmosphere)", the direction of "*fun'iki* (atmosphere) and "*kuuki* (atmosphere)" is appearing and we can find "*kuuki* (atmosphere)" is related to a mental state.

In this level, "*naka*1 (while)" and "*uchi*1 (while/before)" are located in the subordinate position of "*toki* (when)" and "*joutai* (state)".

WordA	Word B	CSM(A:B)	CSM(B:A)	Difference
toki (when)	naka1 (while)	0.781	0.651	0.13
joutai (state)	naka1 (while)	0.747	0.594	0.153
joutai (state)	uchi1 (while)	0.772	0.515	0.257
toki (when)	uchi1 (while)	0.702	0.500	0.202

An example of "*naka*1 (while-)" is "*isogashii naka (adj + naka) kitekurete arigatou*. (Thank you for coming while you are busy.)"

The Normalized Value: 0.6

At this level, words representing a mental state are related to each other.

WordA	Word B	CSM(A:B)	CSM(B:A)	Difference
ki (have got a feeling that)	omoi (thought)	0.645	0.576	0.069
ki (have got a feeling that)	kimochi (feeling)	0.632	0.576	0.056
ki (have got a feeling that)	kibun (mood)	0.552	0.518	0.034
kanji (feeling and/or sense)	omoi (thought)	0.626	0.546	0.080
kanji (feeling and/or sense)	kimochi (feeling)	0.642	0.570	0.072
kanji (feeling and/or sense)	kibun (mood)	0.652	0.611	0.041

"*Kanji* (feeling and/or sense)" and "*ki* (have got a feeling that)" are related to "*kimochi* (feeling)", "kibun (mood)" and "omoi (thought)".

Also at this level, "*kankaku* (sense)" appears.

WordA	Word B	CSM(A:B)	CSM(B:A)	Difference
imeji (impression)	kankaku (sense)	0.693	0.564	0.129
inshou (impression)	kankaku (sense)	0.627	0.536	0.091
kanji (feeling and/or sense)	kankaku (sense)	0.546	0.543	0.003

The difference of CSM("*kankaku* (sense)" : "*inshou* (impression)") is 0.091 and the difference of CSM("*kankaku* (sense)" : "*ime-ji* (impression)") is 0.129. both values are around 0.1. "*Kankaku* (sense)" is in a subordinate position of "*inshou* (impression)" and "*ime-ji* (impression)". Also in normalized value 0.5 we can find that the similarity of "*kankaku* (sense)" and "*kanji* (feeling and/or sense) is high.

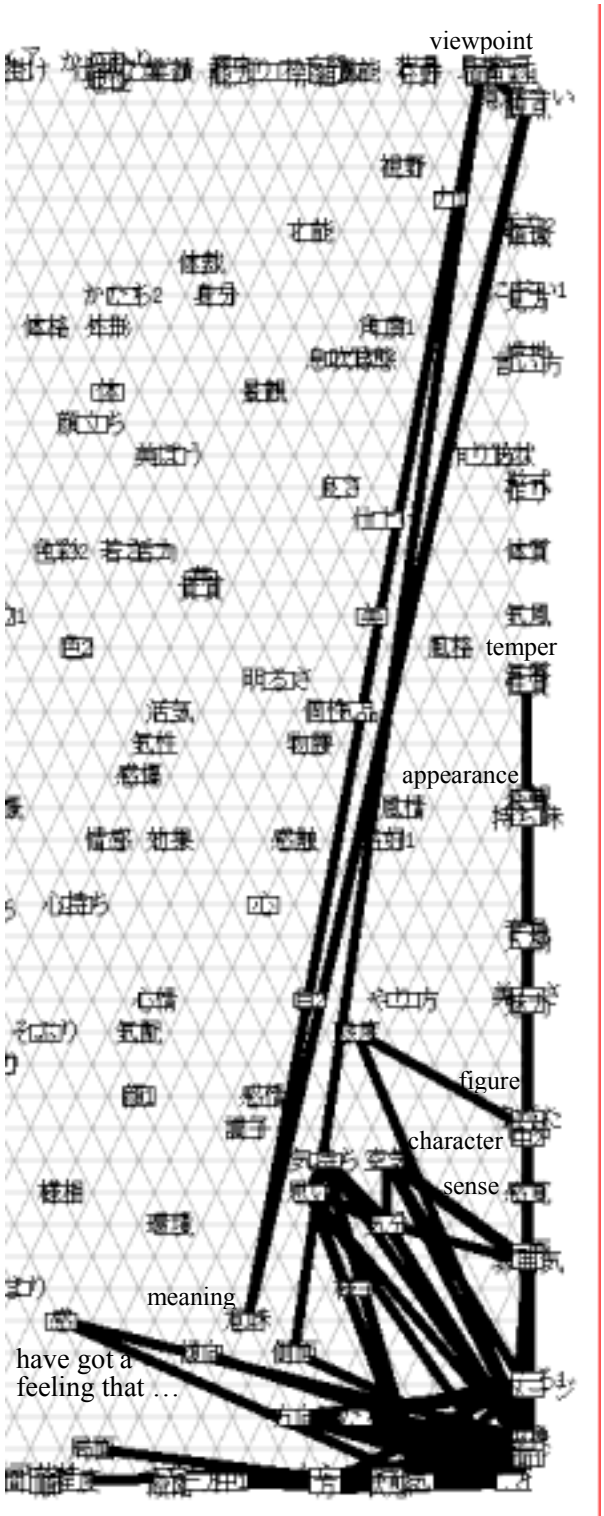

Fig5. Abstract nouns of 1.0-0.60

The difference of CSM("*kankaku* (sense)" : "*kanji* (feeling and/or sense)") is 0.003. "*Kankaku* (sense)" and "*kanji* (feeling and/or sense)" are very similar words in terms of the value.

"*Sugata* (figure)" and "*taido* (attitude)" are located in the subordinate position of "*inshou* (impression)" and also "*taido* (attitude)" is in a subordinate position of "*seikaku* (character)".

Then, "*kyokumen* (phase)" appears in a little higher position from the bottom. "*Kyokumen* (phase)" is related to "*joutai* (state)".

WordA	Word B	CSM(A:B)	CSM(B:A)	Difference
joutai (state)	kyokumen (phase)	0.618	0.395	0.223

From the value of the difference "*kyokumen* (phase)" is a subordinate word of "*joutai* (state)".

In the normalized value of 0.6, we can find "*kurai* (rather)" and "*hou* (comparatively)" are related to "*toki* (when)" and "*joutai* (state)". An example is the case of "*kurai* (rather)"

WordA	Word B	CSM(A:B)	CSM(B:A)	Difference
toki (when)	kurai (rather)	0.681	0.621	0.060
joutai (state)	kurai (rather)	0.619	0.528	0.091

As for "*hou* (comparatively)", the difference of CSM("*hou* (comparatively)" : "*toki* (when)") is 0.103 and the difference of CSM ("*hou* (comparatively)" and "*joutai* (state)") is 0.115.

"*Houkou* (direction)" is a subordinate noun of "*koto* (matter)" at this level, 0.6. In the normalized values 0.5 and 0.4, "*houkou* (direction)" is related to "*kurai* (rather)", "*hou* (comparatively)", "*toki* (when)" and "*joutai* (state)". The difference between them is less than 0.1. In the Semantic Map, "*houkou* (direction)" is located near "*kurai* (rather)", "*hou* (comparatively)", "*toki* (when)" and "*joutai* (state)".

Then, "*keikou* (tendency)" is more similar to "*houkou* (direction)" than those nouns.

WordA	Word B	CSM(A:B)	CSM(B:A)	Difference
houkou (direction)	keikou (tendency)	0.534	0.525	0.009

By this value, we can find that the line between "*houkou* (direction)" and "*keikou* (tendency)" appear in the Semantic Map.

"*Fun'iki* (atmosphere)" is related to "kibun (mood)" at this level.

WordA	Word B	CSM(A:B)	CSM(B:A)	Difference
inshou (impression)	sugata (figure)	0.633	0.485	0.148
inshou (impression)	taido (attitude)	0.613	0.463	0.15
seikaku (character)	taido (attitude)	0.612	0.452	0.16

WordA	Word B	CSM(A:B)	CSM(B:A)	Difference
fun'iki (atmosphere)	kibun (mood)	0.635	0.623	0.012

In normalized value 0.4 and 0.3, "*fun'iki* (atmosphere)" and "*kuuki* (atmosphere)" are related to "*omoi* (thought)", "*kimochi* (feeling)" and "*kibun* (mood)". So, in the Semantic Map they are located near each other.

"*Kan* (have got a feeling that)" appears in this level and is in a subordinate position of "*inshou* (impression)" and "*ki* (have got a feeling that)". Though in the Semantic Map, "*inshou* (impression)" and "*ki* (have got a feeling that)" are not located at the same coordinates, their values are similar. The CSM ("*inshou* (impression)" : "*ki* (have got a feeling that)") is 0.647; on the other hand, the CSM ("*ki* (have got a feeling that)" : "*inshou* (impression)") is 0.565, and the difference between these values is 0.082. So, "*inshou* (impression)", "*ki* (have got a feeling that)" and "*Kan* (have got a feeling that)" is very related to each other.

"*Gaikan* (appearance)" is a subordinate noun of "*inshou* (impression)". "*Kishitsu* (temper) " is a subordinate noun of "*seikaku* (character)".

"*Imi* (meaning)" and "*kanten* (viewpoint)" are in a super/subordinate relation. The examples are "*igakutekina kanten* (a medical

viewpoint)" and "*gakumonteki na imi* (scholarly meaning)". On the map, you can see the line between "*imi* (meaning)" in the lower position and "*kanten* (viewpoint)" in the upper right hand corner.

The Normalized Value: 0.5

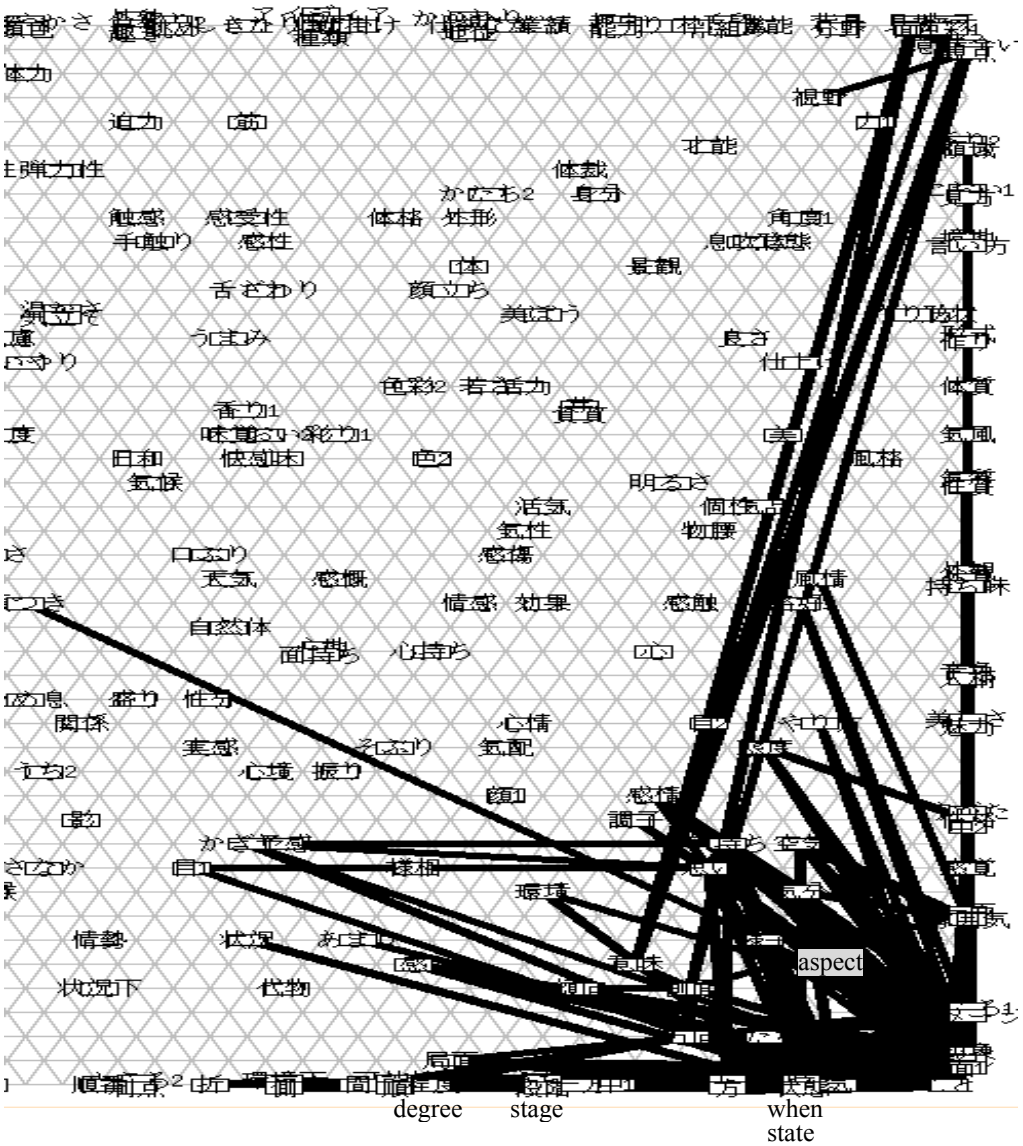

Fig6. Abstract nouns of 1.0-0.50

In this level, many subordinate nouns appear. The feature of this level is that nouns related to a state and a degree are located in the same line, at the bottom end.

For example, "*dankai* (stage)" and "*teido* (degree)" appear in a subordinate position of "*joutai* (state)" and so on.

WordA	Word B	CSM(A:B)	CSM(B:A)	Difference
koto (matter)	dankai (stage)	0.590	0.374	0.216
joutai (state)	dankai (stage)	0.560	0.442	0.118
men1 (a side)	dankai (stage)	0.523	0.353	0.17
toki (when)	dankai (stage)	0.519	0.432	0.087

"*Dankai* (stage)" is located in the direction of "*joutai* (state)" and "*toki* (when)" on the map, and from the viewpoint of the value, the difference of CSM("*toki* (when)" : "*dankai* (stage)") is 0.087. They are similar but they are somewhat super/subordinates.

"*Teido* (degree)" is also located in the direction of "*toki* (when)" and "*joutai* (state)" on the SOM.

WordA	Word B	CSM(A:B)	CSM(B:A)	Difference
joutai (state)	teido (degree)	0.568	0.378	0.19
koto (matter)	teido (degree)	0.539	0.305	0.174
kurai (rather)	teido (degree)	0.527	0.404	0.123
toki (when)	teido (degree)	0.512	0.362	0.15

From CSM value "*teido* (degree)" is a subordinate position of "*toki* (when)", "*joutai* (state)" and "kurai (rather)"

Also, "*yousu* (aspect)" appears from this level. It is a subordinate word of "*koto* (matter)", "*toki* (when)", "*inshou* (impression)" and "*kanji* (feeling and/or sense)" and so on in terms of the CSM value.

6.2 The Classifications of Abstract Nouns

The superordinate abstract nouns related to most adjectives appear between the normalized values 1.0 and 0.5.

According to the CSM value, the most abstract nouns to related adjectives are "*koto* (matter)", "*inshou* (impression)" and "*kanji* (feeling and/or sense)". Morita (1868) supposed that an adjective was an individual subjective representation and a representation for the relative judgment. The result in this research seems to prove his analysis.

In the Semantic Map we can find some directions of the distribution of nouns according to the above relations. We show the outline figure of the SOM constructed below (Fig 7).

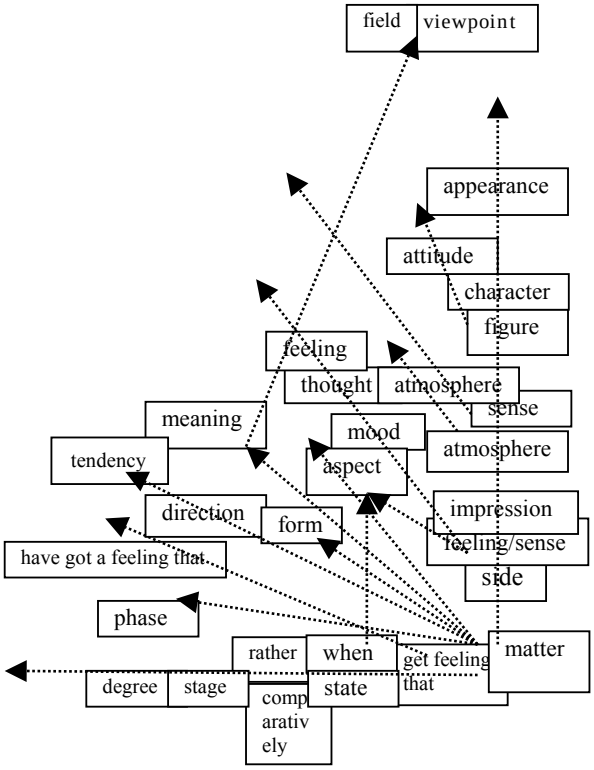

Fig.7. The outline of a distribution of abstract nouns

a. A starting point: "*koto* (matter)"
The direction of
[*joutai* (state)group]
toki (when), *joutai* (state),
naka1 (while), *uchi1* (while/before),
[*hou* (comparatively)/*teido* (degree) group]
hou (comparatively), *kurai* (rather), *dankai* (stage) and *teido* (degree).
In this direction, "*joutai* (state)" group and "*hou* (comparatively)/*teido* (degree)" group are mixed.)

The direction of *houkou* (direction)
keikou (tendency)
The direction of *katachi*1 (form)
The direction of *imi* (meaning),
kanten (viewpoint)
The direction of *joutai* (state),
"*kyokumen* (phase)"

b. A Starting points: *inshou* (impression),
imeji (impression)
The direction of *men*1 (one side),
ichimen (one side),
[*seikaku* (character) group]
seikaku (character), *miryoku* (charm),
hitogara (personality),
utsukushisa (beauty), *neiro* (tone),
[*sugata* (figure)]
sugata (figure), *kakkou* (appearance),
gaikan (outward appearance),
gaiken (look)
In this direction "*seikaku* (character)" group and "*sugata* (figure)" group are on the same line.
The direction of *seikaku* (character),
sugata (figure) and *taido* (attitude)

c. A starting point: *kanji* (feeling and/or sense)

At the first stage, value 0.9, *"kanji* (feeling and/or sense)*"* became a subordinate noun of "*koto* (matter)" and "*inshou* (impression)".
The direction of *kimochi* (feeling),
kibun (mood), *omoi* (thought)
The direction of *fun'iki* (atmosphere),
kuuki (atmosphere)
The direction of *ki* (get the feeling that),
kan (have got a feeling that)
The direction of *kankaku* (sense) (This direction is the same line as "b"group)
The direction of *ki* (get the feeling that),
kan (have got a feeling that)
The direction of *yousu* (aspect)

Japanese adjectives are often classified mainly into the "characteristics", the "mental state", the "state" and the "degree". In our result, firstly, "*koto* (matter)" and "*inshou* (impression)" are the most abstract nouns related to adjectives, and then "*kanji* (feeling and/or sense)" appears as the subordinate noun of "*koto* (matter)"and "*inshou* (impression)". "*Koto* (matter)" is a base of "state" and "degree", "*inshou* (impression)" is a base of "characteristics" and finally "*kanji* (feeling and/or sense)" is a base of "mental state".

However, "*koto* (matter)", "*inshou* (impression)", "*kanji* (feeling and/or sense)", "*men*1 (one side)", "*joutai* (state)", " *imi* (meaning)", "*hou* (comparatively)" and so on are related to each other at the top level because the difference of the CSM value between them is low. This means that abstract nouns at the top level co-occur with many common adjectives, so, most co-occurring adjectives of their abstract nouns can be combined with any abstract noun at the top level. So we suppose that at the top level of abstract concept all adjectives potentially have these abstract meanings. However, as the concept is more concrete, the adjective is more specified like the concept of the "distance" has adjectival meanings such as "far/ nearby/…"

7. Conclusion

We sketched roughly the directions of the distributions of abstract nouns and adjectives on the SOM. In our method, we extracted abstract nouns and co-occurring adjectives like a cognate relation, that is, we extracted the relation that adjectives represent concrete instances of abstract nouns, e.g., "*ureshii kimochi* (happy feeling)". So we consider that a SOM of abstract nouns classifies an abstract meaning that adjectives have. Though adjectives are not classified clearly in the superordinate layer, if the abstract noun is more subordinate, sets of co-occurring adjectives are more specific. For example, "*hiroi* (wide) / *semai* (narrow)", "*hayai* (early) / *osoi* (late)", "*kouteitekina* (affirmative) / *hiteitekina* (negative)", "*igakutekina* (medical)

/ *kagakutekina* (scientific)" and "*koukina* (noble) / *yuishotadashii* (with a long and distinguished history)" and so on co-occur with abstract nouns in the subordinate position.

As a next step, We need to compare this result of the Semantic Map with the Semantic Map using frequency. In the future, we aim to detect a system of intermediate concept linking adjectives with concrete nouns by using adnominal relationship between adjectives and abstract nouns.

References

Kohonen, T. (1995), Self-Organizing Maps, Springer.

Hindle, D. (1990), *Noun Classification From Predicate-Argument Structures*, In the Proceedings of the 28th Annual Meeting of the Association for Computational Linguistics, *pp.268-275*

Hatzivassiloglou,V. and McKeown.R.K. (1993), *Towards the Automatic Identification of Adjectival Scales: Clustering Adjectives According to Meaning*, In the Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics, *pp. 172-182.*

Yamamoto. E, and Umemura, K. (2002), *A Similarity Measure for estimation of One–to- Many Relationship in Corpus*, Journal of Natural Language Processing, pp.45-75.

Takahashi, T. (1975), *A various phase related to the part-whole relation investigated in the sentence*, Studies in the Japanese language 103, The society of Japanese Linguistics, *pp.1-16*

Isahara, H. and Kanzaki, K. (1999), Lexical Semantics to Disambiguate Polysemous Phenomena of Japanese Adnominal Constituents, In Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics (ACL'99), *pp.489-496*

Ma.Q, Kanzaki. K, Murata. M, Uchimoto. K and Isahara. H (2000), Self-Organization Semantic Maps of Japanese Noun in Terms of Adnominal Constituents, In Proceedings of IJCNN'2000, Como, Italy, vol.VI.

Morita.Y (1968) Grammer- *The way to represent a motion/state*, A course of Japanese, Institute of Language Teaching, WASEDA University.