An Experimental Classification of English Noun Phrases Used in Metaphorical Expressions

Yasutake ISHII
Graduate School of Area and Culture Studies,
Tokyo University of Foreign Studies

Abstract

Metaphors play an essential part in human communication, but it is difficult to deal with metaphors using a computer. One of the reasons for this difficulty is that a computer program can only interpret the explicit meaning of a sentence using a dictionary. Constructing a comprehensive machine-readable dictionary would seem to solve the problem, but it is virtually impossible to make a fully objective and comprehensive dictionary manually. What is necessary is a machine-readable dictionary based on knowledge inductively acquired through the real use of language. This paper will show how it was possible to collect from the BNC a large number of English noun phrases used in metaphorical expressions consisting of some selected prepositions. The resulting data show the possibility of constructing a database for computational processing of metaphors.

1 Introduction

Metaphors are one of the most difficult phenomena to deal with in natural language processing. Metaphors play a basic and essential part in human communication, but they communicate a meaning that is not derived directly from linguistic symbols. Without the consideration of metaphor, advancement in natural language processing would be quite limited. A database of vocabulary utilizable for the explanation of metaphor is called for.

In our previous research[1], it was proved that it is possible to extract the complement noun phrases (henceforth just "nouns") of particular English prepositions (*in*, *into*, and *out of*) from large corpora, which enables the categorization, to some extent, of words that can be interpreted as a container. This method enables us to construct a comprehensive dictionary which is free from editors' subjectivity and can be used to process the implicit meanings of metaphors.

In this paper, we solidify our linguistic basis, which is necessary for proper identification of linguistic variables and for a valid analysis of language[2]. Also, we reconfirm the validity of our approach using the BNC (described in detail below), and extend the scope of the research.

メタファに着目した英語の名詞分類の試み
石井 康毅
東京外国語大学大学院 地域文化研究科
3-11-1 Asahi-cho, Fuchu-shi, Tokyo, 183-8534 JAPAN
ishii@sano.tufs.ac.jp

2 Metaphor

2.1 Basic Concept

Metaphor is often misunderstood to be only a device of poetic imagination or rhetorical expressions, which is far from true. Our conceptual system is fundamentally metaphorical in nature. Whether we notice it or not, our everyday language is full of metaphorical expressions.

Lakoff and Johnson[3] analyzed various metaphorical expressions and the conceptual structures behind them, and pointed out that the essence of metaphor is understanding and experiencing one thing in terms of another.

Metaphor is also considered to be of much significance in relevance theory. According to Sperber and Wilson[4], metaphor is a natural outcome of some very general abilities and procedures used in verbal communication.

2.2 Container Metaphors

There are many metaphorical concepts, but here we concentrate on "container" metaphors. According to Lakoff, even if there is no natural physical boundary, we imagine boundaries and mark off territory where there is an inside and a bounding surface. He argues that this is one of the most basic human instincts. The metaphorical concept of "VISUAL FIELDS ARE CONTAINERS"[1] is an example of the container metaphors. These container metaphors are often expressed in the form of prepositional phrases, and the typical prepositions used in them are *into*, *out of*, *in*, etc. So the complement nouns of these prepositions can be seen as containers, and they constitute a category sharing some semantic features.

The following sentences[3][2] are examples of the metaphorical concept of "VISUAL FIELDS ARE CONTAINERS."

(1) a. The ship is coming *into* view.
b. He's *out of* sight now.
c. I can't see him – the tree is *in* the way.

[1] Metaphorical concepts are a metalanguage, and written in the upper case.

[2] Slight changes were made by the author.

For the sentence (1a), for example, the complement noun of the preposition *into* is "view."

2.3 Difficulty in Processing Metaphors

We human beings can understand the above metaphorical expressions (1a-c) easily. We can judge that these sentences are related to each other in their use of visual fields. When these sentences are understood, particular aspects of container-images are highlighted through the metaphorical concept of "VISUAL FIELDS ARE CONTAINERS," and at the same time the other aspects are hidden. Thus, when we hear or think of these sentences, we regard visual fields as containers. In order to understand the sentences, we have to recognize either consciously or unconsciously the metaphorical relationship between the prepositions and the complement nouns. A computer program can, however, only interpret the explicit meaning of a sentence depending on a dictionary. If it is not explicitly shown in a dictionary that these words have some semantic features in common, the program can not judge that these sentences are related to each other. What is necessary is a machine-readable dictionary based on knowledge inductively acquired through the real use of language. This would be much more comprehensive and far more effective for computational processing of metaphorical expressions than existing man-made dictionaries.

3 Approach

3.1 Prepositional Phrases

Prepositions have fundamental meanings of physical places or locations, but when used in the container-metaphorical expressions, they are also used for meaning abstract spaces or states. So in these examples, different from the common classification of vocabulary constructed in terms of content words, it is effective to make an analysis in terms of function words. Our hypothesis is that the complement nouns of these prepositions can be seen as containers and constitute a category sharing some semantic features. We selected *in, inside, into, out of, outside*, and *within* as typical prepositions used for the container metaphors, and collected the complement nouns of them.

3.2 The BNC

Although Chomsky is critical about depending on corpora in linguistic studies, many works in corpus linguistics show their effectiveness in that they can give us various insights and reveal various unknown facts[5].

In our research, we used the second edition of the British National Corpus (henceforth "BNC")[3]. It contains about 100 million words of present-day British English. About 10% of the data is spoken English and the rest is written. The text is grammatically tagged (the line below is a sample from the BNC).

[3]http://www.hcu.ox.ac.uk/BNC/.

Sample line from the BNC:
<s n="694"><w AT0>The <w AJ0>Roman <w AJ0>catholic <w NN1>church <w VBZ>is <w VVN>financed <w PRP>out of <w AJ0>voluntary <w NN2>contributions<c PUN>.

3.3 Automatic Extraction

In our research, we developed a Perl script that reads an entire text, finds selected prepositions, and identifies only the complement nouns while ignoring the adjuncts. Since the BNC is a POS(part of speech)-tagged corpus, we need not use a syntactic parser to analyse the part of speech of each word, which frees us from very time-consuming preparatory work.

3.4 Statistics

We can calculate the difference coefficients from the relative frequencies (not the absolute numbers of occurrences) of the complement nouns of each selected preposition and their relative frequencies in any possible positions for nouns in the whole corpus except for the complement positions.

$$\text{Difference Coefficient} = \frac{Freq._A - Freq._B}{Freq._A + Freq._B}$$

This value can be between -1 and +1. If the value is close to +1, it means that the noun appears exclusively in the complement positions of the particular preposition, and if it is close to -1, it means that the noun appears anywhere except for the complement positions. If the value is close to 0, it is equally proportionate and there is no relationship between the particular preposition and the noun.

The statistical test to confirm the validity of the data is the chi-square test. We take the null hypothesis that there is basically no difference between the complement positions and all the possible positions for nouns. If this null hypothesis is rejected, we can reasonably judge that the two conditions are different.

Another statistical index used here is the correlation coefficient. It is used to confirm the deflection of particular words to particular conditions.

4 A Survey of the Complements of Selected Prepositions

4.1 Some Problematic Procedures

When collecting the complement nouns of the selected prepositions, we faced some procedural problems. Some of the problems and how we coped with them are listed below.
It was difficult:

- to distinguish between some singular and plural forms. We made it a rule not to distinguish the forms when we did not have significant difference in the usage or meaning of singular/plural forms. One of the exceptions is "account" in the complement position of *into*, which is much more common than "accounts." However, the delineation is

necessarily arbitrary to some extent. Those words marked "-" are counted only when they appear as the very form shown there.

- to deal with proper nouns and compound nouns. We included them into our scope of research because in most cases their syntactic behaviors are the same.
- to deal with characters other than ASCII characters. Unicode would solve the problem, but other elements such as mathematical formulae are hopeless. Therefore, we concentrated only on ASCII characters.

4.2 Counting the Number of Occurrences

We have to count the total number of nouns which appear in the BNC. Although for the approximate figures, we can refer to Leech et al.[6], we developed a script to count the correct number of occurrences of all the nouns using the POS-tags. The same method was used for counting the exact numbers of each item under discussion.

4.3 Results

Figures 1-6 show the difference coefficients of the most frequent complement nouns of each preposition[4], with the correlation coefficients (abbreviated as "CC").

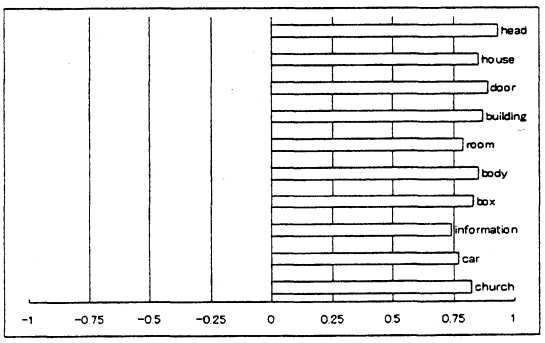

Figure 1: Complements of *inside*; CC=0.5626

All the data are statistically significant at the 1 percent level. Also, the correlation coefficients are low in general, which means that there is little relationship between the two environments.

5 Conclusion

5.1 Discussion of the Results

As for the complements of *inside*, they can be divided into three groups; abstract objects, physical constructs, and bounding surfaces of the constructs. All three groups relate to the container concept, and this

[4]As for *in*, since the frequency of *in* is too high, we counted the first 200,000 complements.

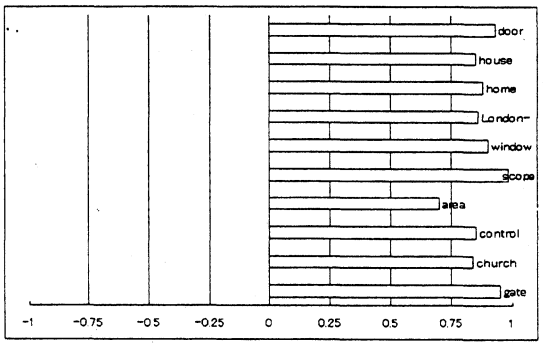

Figure 2: Complements of *outside*; CC=0.4969

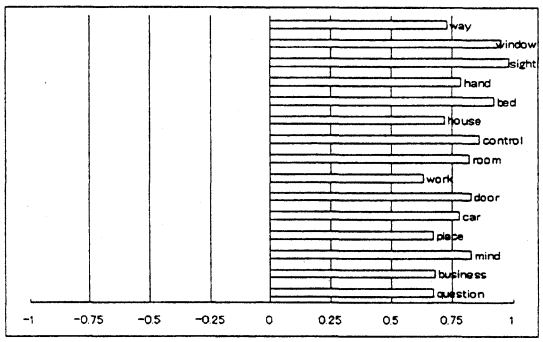

Figure 3: Complements of *out of*; CC=0.2038

model applies to all the results. The first group consists of "head," "body," and "information." The second group includes "house," "building," etc. The third has "door," and it is interesting in that its concern is not a container itself but its bounding surface. A door or a window attracts our attention as a path to the outer world when we are in a room. The meaning of *inside* is fairly limited, which must be the reason why the difference coefficients are relatively high. It follows, then, that the two environments, the complement positions of *inside* and all the other possible positions for nouns, are independent to a large extent.

The complements of *outside*, the antonym of *inside*, show the similar tendency. Typical nouns relating to the container concept appear very frequently here. It is also notable that nouns of typical boundary concepts are included such as "scope" and "area."

Most of the words listed in the two results above have an image of a bounding surface and in-out orientation.

The result of *out of* and *into* show that they are also typical (equivalent) prepositions of container metaphors. Some of the complements are often used as frequent set phrases; "*out of* the way," "*out of* sight, *out of* mind," and "take *into* account," for example.

Many of the complements of *in* are also used in frequent phrases such as "*in* fact." There are, however,

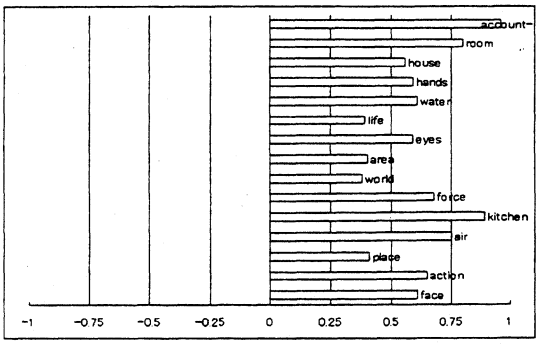

Figure 4: Complements of *into*; CC=-0.3137

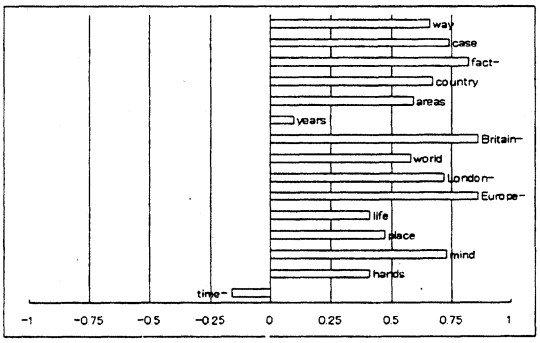

Figure 5: Complements of *in*; CC=0.0776

some words which can not be interpreted as a container object such as "years" and "time." This must be due to the fact that *in* has a wider range of meaning and is not used exclusively for container metaphors.

When we look at the results of *into* and *in*, the difference coefficients are relatively low compared with the above three prepositions. One reason may be that *into* and *in* are complementarily used for static/dynamic context, and some words are split into the two environments.

As for the result of *within*, some words ("framework" and "community," for example) are typical container-like objects and have very high difference coefficients, but at the same time, some words have meanings related to time. It is of course the same case as *in*.

Those words that appear in more than two complement positions of the six selected prepositions are "area," "door," "hands," "house," "place," and "room." These words can be seen as typical physical or abstract containers.

We referred to WordNet[5] to see whether these nouns are categorized in the way the container metaphor can be explained. We could not, however, gather very descriptive data or explanations of the nouns. This shows that our result is, to some extent, successful in classifying words in a way that has not been attempted.

[5] http://www.cogsci.princeton.edu/~wn/.

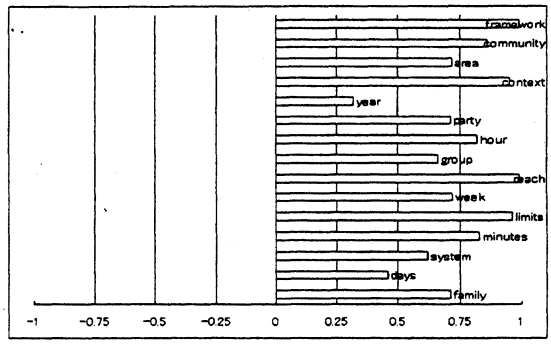

Figure 6: Complements of *within*; CC=-0.2600

5.2 Further Study

We have to refine the rules of extracting complement nouns of prepositions because at this point some omissions remain. For example, the following are ignored at this point.

- *inside* most of the old city's public houses
- *inside* NOUN or/and NOUN (...)
- *inside* which NOUN (...)

We have to make some adjustments and difficult decisions to deal with them.

We could apply our research to educational purposes if we concentrate on the differences between certain text categories or registers.

We plan to extend our research to various metaphorical concepts and adopt other statistical indices. We aim to cluster the results and construct databases of English vocabulary applicable to the processing of metaphors.

References

[1] Ishii, Y. and H. Sano. 2001. "An Experiment in Classifying Nouns Associated with "Metaphorical Prepositional Phrases."" (written in Japanese) In *Proceedings of The Seventh Annual Meeting of The Association for Natural Language Processing*, pp. 82-85.

[2] Wolfram, W. 1993. "Identifying and Interpreting Variables." In D. Preston, ed., *American Dialect Research*, pp. 193-221. John Benjamins.

[3] Lakoff, G. and M. Johnson. 1980. *Metaphors We Live By*. The University of Chicago Press.

[4] Sperber, D. and D. Wilson. 1995. *Relevance : Communication and Cognition*, Second Edition. Blackwell.

[5] McEnery, T. and Wilson, A. 1996. *Corpus Linguistics*. Edinburgh University Press.

[6] Leech, G., P. Rayson and A. Wilson. 2001. *Word Frequencies in Written and Spoken English*. Longman.