数理解析研究所講究録
495巻 1983年 71-86

A Mathematical Structure of Dictionaries

Shuji Shiraishi

Department of Information Systems
Graduate School of Engineering Science
Kyushu University

ABSTRACT

In this paper we first present a formal definition of dictionaries and introduce a semantic space of dictionaries which is used for giving formal meanings of entry words. Then we show that the semantic space is uniquely determined up to isomorphism. We also construct the semantic space which consists of infinite trees with no leaf and some other trees where only leaves are labeled. So it is enough to take this constructed semantic space when considering semantics of dictionaries.

1. Introduction

Dictionaries are indispensable not only for our daily lives but also for computerized systems such as database systems and knowledge systems [1,2].

According to the dictionary [3], "dictionary" is "a book containing a selection of the words of a language, usually arranged alphabetically giving information about their meanings, pronounciations, etymologies, inflected forms, etc., expressed in either the same or another language." However in the present paper we simply take a dictionary as a book containing words of a language and their meanings expressed in the same language.

Although there may be several standpoints to treat meanings of words in a dictionary, we take them in the following manner. Consider our consulting a dictionary for the meaning of a word. The explanation of the word is expressed with a finite sequence of words in the dictionary. But if we find unknown words in it, we may again consult the dictionary for them. In this way we will get the meaning of a word by consulting a dictionary finitely many times.

As far as the author's knowledge is concerned, no formal definition of a dictionary has been given and semantics of dictionaries has not yet been studied in a mathematical way. So in this paper we give a formal definition of a dictionary and a semantic space of dictionaries. A dictionary is made up of a set of entry words, their explanations which are expressed with finite sequences of entry words in the dictionary and possibly undefined words. A semantic space is made up of a set Y and a bijection # from Y^+ (the union of n-products of Y) to Y satisfying some commutative diagram.

In Section 2 we present a formal definition of a dictionary, which gives the framework of a dictionary and enables a mathematical discussion. In Section 3 we introduce a semantic space to treat formal meanings of words in a dictionary. Then we show that the semantic space is uniquely determined up to isomorphism. In Section 4 we construct the semantic space which made up of infinite trees with no leaf and some other trees where only leaves are labeled.

2. Dictionary

We give in this section a formal definition of a dictionary. In general explanations of entry words in a dictionary are expressed with finite sequences of entry words and possibly undefined words.

Notation. Let X be a set. Then X is defined by:

$$X^+ = X + X^2 + \ldots + X^n + \ldots,$$

where X^n is n-fold product of X and + is disjoint union.

Definition 1. A **dictionary** is a triple DIC = (X,A,D), where

(1) X is a nonempty set of **entry words,**

(2) A is a set of **undefined words,**

(3) D is a mapping from X to $(X+A)^+$.

We call words other than **entry words undefined words** and also call D(x) the **explanation** of x. The mapping D corresponds to the action of consulting a dictionary. If A is empty, we call a dictionary **complete.**

Example. We give examples.

D(concept) = (a, general, notion, or, idea),

D(notion) = (a, general, or, vague, idea), D(idea) = (any, conception, existing, in, the, mind),

D(general) = (of, or, pertaining, to, all, persons, or, things, belonging, to, a, group, or, category),

D(mind) = (the, part, in, a, human, being, that, reasons, understands, perceives).

For clearness, the expression $D(x) = (x_1, x_2, \ldots, x_n)$ may be illustrated by using the following tree (Fig 2.1):

D(x):

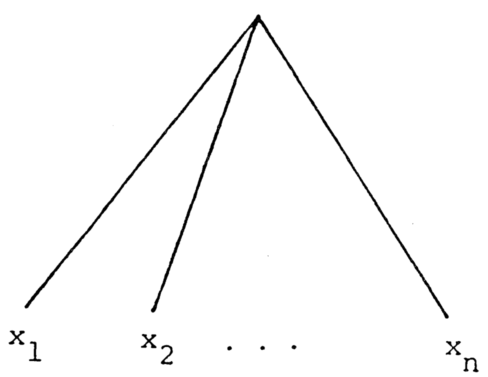

Fig 2.1 $D(x) = (x_1, x_2, \ldots, x_n)$.

Next we consider semantics of dictionaries. Several approaches to the semantics may be considered from the various standpoints, such as linguistic, philosophical and mathematical standpoints. But we here take the semantics formally (mathematically) in the following way, that is, a formal meaning of any word in a dictionary is obtained by combining those of the words which appear in the explanation.

For exmaple let us consider the semantics of very tiny dictionary (X, A, D), where

$$X = \{x_1, x_2, x_3\},$$

$$A = \{a\},$$

and D is defined by:

$$D(x_1) = (x_1, x_2),$$

$$D(x_2) = (a),$$

$$D(x_3) = (x_3, a).$$

As $D(x_1) = (x_1, x_2)$, the formal meaning of x_1 is the combination of those of x_1 and x_2. Now let us see the process of getting the meaning of x_1 (Fig 2.2).

1st step.

2nd step.

3rd step.

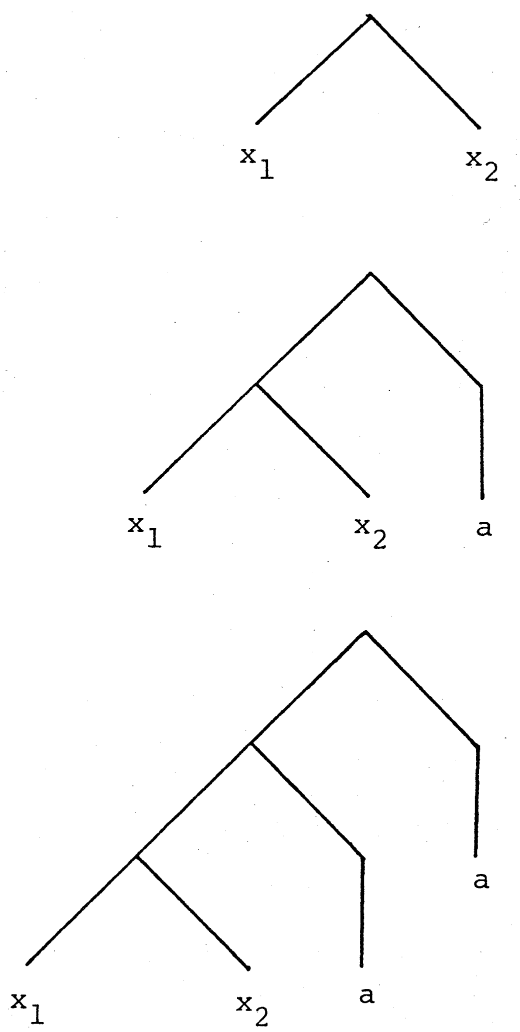

Fig 2.2 The first three partial meanings of x_1.

The first step shows the (partial) formal meaning of the word x_1 consulting the dictionary one time, the second step two times and the third step three times respectively. That is, the (total) meaning of x_1 is gained by consulting the dictionary repeatedly. This may be the process by which we acquire meanings of words. If we encounter undefined words (here "a"), then this process partially stop there. Since we cannot consult them for the dictionary any more. This is the same for x_2, x_3. So we may take the meanings of x_1,x_2,x_3 as the following tree diagrams respectively (Fig 2.3):

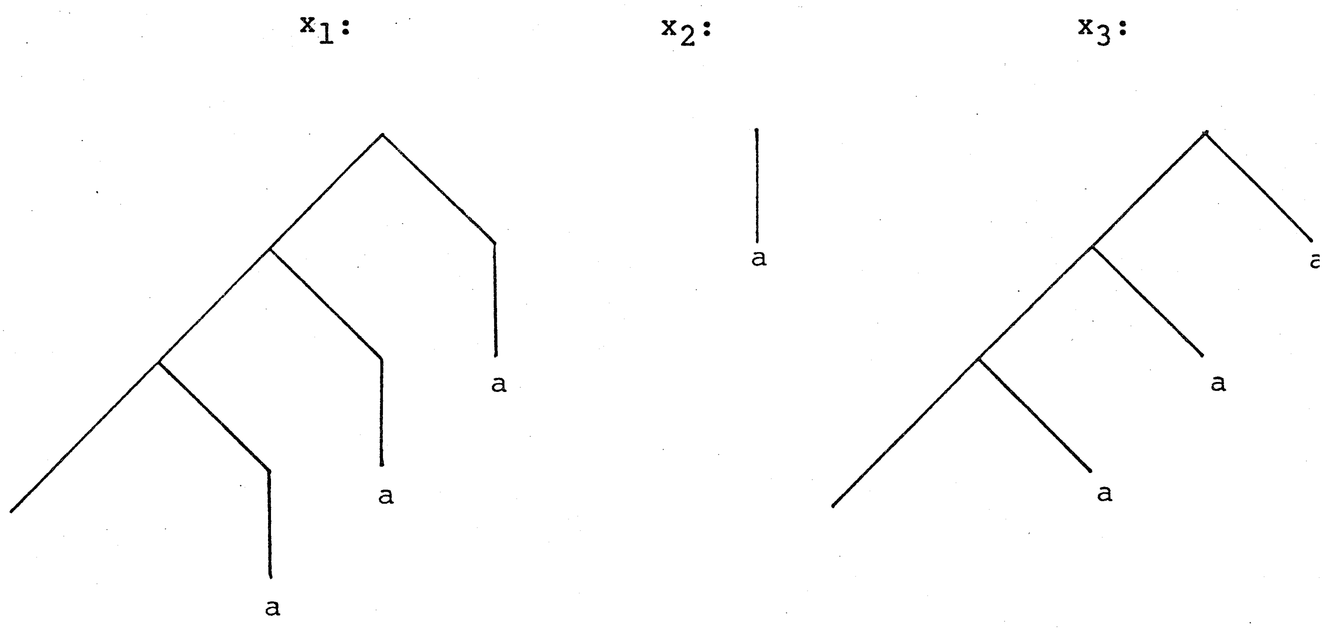

Fig 2.3 The formal meanings of x_1,x_2 and x_3.

As looked at above, in order to consider semantics of dictionaries, we are necessary some space which is used for giving formal meanings of words such as trees illustrated above. So in the next section we introduce a semantic space of dictionaries.

3. Semantic space

We first define a semantic space of dictionaries and state the reason why this definition is sound. Then we show that semantic spaces are isomorphic.

Now we define a semantic space formally.

Definition 2. A **semantic space** for dictionaries is a pair $(Y,\#)$ with the following conditions:

(1) Y is a set.

(2) $\# : Y^+ \to Y$ is a bijection.

(3) For any dictionary $DIC = (X,A,D)$, there exists only one mapping, called a semantic mapping, $s : X \to Y$ such that for any mapping $\tilde{s} : X+A \to Y$ with $\tilde{s}|X = s$ the diagram below commutes:

$$\begin{tikzcd} X \arrow[r, "s"] \arrow[d, "D"'] & Y \\ (X+A)^+ \arrow[r, "\tilde{s}^+"'] & Y^+ \arrow[u, "\#"'] \end{tikzcd}$$

where $\tilde{s}^+ = s + s^1 + \ldots + s^n + \ldots$.

We explain briefly the reason why this definition is sound. It should be natural that we require the semantics $s(x)$ of a entry word x in X is uniquely determined. This semantics $s(x)$ is successively obtained along the commutative diagram above. Firstly, the word x in X is expressed with the finite sequence $x_1,\dots,x_n$ ($n \geq 1$) of entry words and possibly undefined words by using the mapping D, that is,

$$D(x) = (x_1,\dots,x_n).$$

Secondly the semantics of the explanation of x is defined by:

$$\tilde{s}(D(x)) = \tilde{s}^n(x_1,\dots,x_n) = (\tilde{s}(x_1),\dots,\tilde{s}(x_n)).$$

Here if x_i is in X, then $\tilde{s}(x_i) = s(x_i)$. Otherwise, that is, if x_i is undefined words (x_i in A), the semantics may be any element of Y as long as our definition is satisfied.

Lastly the semantics s(x) of x in X is obtained by combining these semantics $s(x_1),\dots,s(x_n)$ using the mapping #, that is,

$$s(x) = \#(\tilde{s}(x_1),\dots,\tilde{s}(x_n)).$$

The mapping # should be a bijection, since if the semantics of the explanations of words are mutually distinct, so should be the semantics of the words, and since extra semantics is not in Y, that is, any semantics of Y is always the combination of some finite number of semantics of Y.

By this definition, we get the following proposition, which asserts that any semantic spaces are isomorphic.

Proposition 3. Let $(Y_i,\#_i)$ ($i = 1,2$) be semantic spaces. Then

$$(Y_1,\#_1) \cong (Y_2,\#_2)$$

holds.

proof. $\#_i^{-1} : Y_i \to Y_i^{+}$ ($i = 1,2$) are considered to be complete dictionaries. Let $f : Y_1 \to Y_2$ and $g : Y_2 \to Y_1$ be the semantic mappings. Then the following commutative diagram holds:

$$\begin{tikzcd} Y_1 \arrow[r, "f"] \arrow[d, "\#_1^{-1}"'] & Y_2 \arrow[r, "g"] \arrow[d, shift left, "\#_2^{-1}"] & Y_1 \\ Y_1^+ \arrow[r, "\tilde{f}^+"'] & Y_2^+ \arrow[u, shift left, "\#_2"] \arrow[r, "\tilde{g}^+"'] & Y_1^+ \arrow[u, "\#_1"'] \end{tikzcd}$$

and the diagram

$$\begin{tikzcd} Y_1 \arrow[r, "i_{Y_1}"] \arrow[d, "\#_1^{-1}"'] & Y_1 \\ Y_1^+ \arrow[r, "\tilde{i}^+_{Y_1}"'] & Y_1^+ \arrow[u, "\#_1"'] \end{tikzcd}$$

also holds, where i_Y the identity mapping. By the uniqueness of semantic mapping we have $gf = i_Y$. Similarly we have $fg = i_Y$. Therefore the proposition is obtained. □

4. A semantic space construction

We now construct the semantic space which consists of infinite trees with no leaf and some other trees where only leaves are labeled. For this end, we first give some notions on trees.

semantic mappings. Then the following commutative diagram holds:

$$\begin{tikzcd}
Y_1 \arrow[r, "f"] \arrow[d, "\#_1^{-1}"'] & Y_2 \arrow[r, "g"] \arrow[d, shift left, "\#_2^{-1}"] & Y_1 \\
Y_1^+ \arrow[r, "\tilde{f}^+"'] & Y_2^+ \arrow[r, "\tilde{g}^+"'] \arrow[u, shift left, "\#_2"] & Y_1^+ \arrow[u, "\#_1"']
\end{tikzcd}$$

and the diagram

$$\begin{tikzcd}
Y_1 \arrow[r, "i_{Y_1}"] \arrow[d, "\#_1^{-1}"'] & Y_1 \\
Y_1^+ \arrow[r, "\tilde{i}^+_{Y_1}"'] & Y_1^+ \arrow[u, "\#_1"']
\end{tikzcd}$$

also holds, where i_Y the identity mapping. By the uniqueness of semantic mapping we have $gf = i_Y$. Similarly we have $fg = i_Y$. Therefore the proposition is obtained. □

4. A semantic space construction

We now construct the semantic space which consists of infinite trees with no leaf and some other trees where only leaves are labeled. For this end, we first give some notions on trees.

Definition. Let $J_n = \{1,\ldots,n\}$ and let $J^* = J_1^* + \ldots + J_n^* + \ldots$. Then $J^* \supset \alpha$ is said to be a **tree** if the following conditions are satisfied:

(1) $st \in \alpha$ implies $s \in \alpha$.

(2) $sk \in \alpha$ and $k \in N$ imply $s\{1,\ldots,k\} \subset \alpha$,

where N denotes the set of all natural numbers.

Any tree is finitely branching. Δ and $L(\alpha)$ denote the set of all trees and the set of all leaves of a tree α, respectively.

Definition.

α is a **partial tree** iff $L(\alpha) \neq \emptyset$.

α is a **total tree** iff $L(\alpha) = \emptyset$.

Partial trees are trees with leaves. Total trees are (infinite) trees with no leaf. We denote by Δ_p Δ_t the set of all partial trees and the set of all total trees, respectively. Clearly $\Delta = \Delta_p + \Delta_t$. We also define labeled trees.

Definition. Let Z be a labeled set. Then the set of all Z-labeled trees is

$$Z^\Delta = \bigcup_{\alpha\in\Delta}\{ m \mid m : \alpha \rightarrow Z \},$$

and the set of all trees where only leaves are Z-labeled is

$$Z^{\tilde{\Delta}} = \bigcup_{\alpha\in\Delta}\{\alpha-L(\alpha)+m \mid m : L(\alpha) \rightarrow Z \}.$$

Example. Fig 4.1 is an example of a tree α.

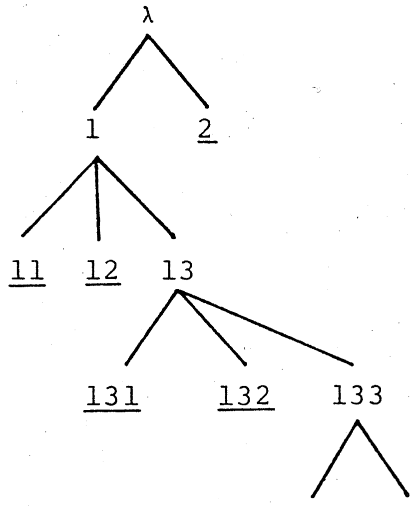

Fig 4.1

Leaves are marked by underlines, that is, $L(\alpha) = \{11, 12, 131, 132, \dots\}$

Fig 4.2 is an example of Z-labeled tree.

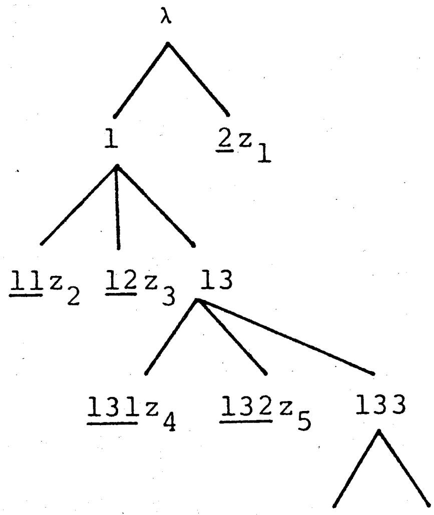

Fig 4.2

For example $2z_1$ means that leaf 2 is labeled by z_1.

Now we define a pair $(T,\#)$ as follows:

(1) $T = \Delta_t + \{\perp\}^{\tilde{\Delta}}$,

(2) $\#$ is a mapping from T^+ to T such that

$$\#(\alpha_1,\ldots,\alpha_n) = \{\lambda\} + 1\alpha_1 + \ldots + n\alpha_n.$$

Then we get:

Theorem 4. $(T,\#)$ is a semantic space.

proof. Clearly $\#$ is a bijection. So it suffices to show the uniqueness and existence of the mapping s satisfying the following commutative diagram:

$$\begin{tikzcd}
X \arrow[r, "s"] \arrow[d, "D"'] & T \\
(X+A)^+ \arrow[r, "\tilde{s}^+"'] & T^+ \arrow[u, "\#"']
\end{tikzcd}$$

where we define for any a in A, $\tilde{s}^+(a) = \{\perp\}$.

First we show the uniqueness of the mapping s. Let t, s be the semantic mappings satisfying the above commutative diagram. Then for any x in X with $D(x) = (x_1,\ldots,x_n)$, we get

$$t(x) = \{\lambda\} + 1\tilde{t}(x_1) + 2\tilde{t}(x_2) + \ldots + n\tilde{t}(x_n),$$

$$s(x) = \{\lambda\} + 1\tilde{s}(x_1) + 2\tilde{s}(x_2) + \ldots + n\tilde{s}(x_n).$$

As $t(x)$ and $s(x)$ are the trees with the same root $\{\lambda\}$, we can say that $t(x)$ and $s(x)$ are the same trees if all the children of the root $\{\lambda\}$ are the same. Indeed if x_i is in A, then by the definition of $\tilde{s}$, $\tilde{s}(x_i) = \tilde{t}(x_i) = \{\perp\}$. Otherwise, that is, if x_i is in X, then $\tilde{t}(x_i) = t(x_i)$ and $\tilde{s}(x_i) = s(x_i)$. But $t(x_i)$ and $s(x_i)$ are the trees with the same root $\{\lambda\}$, so all the children are the same.

Next we show the existence of the mapping s. We consider the following commutative diagram:

$$\begin{tikzcd}
X \arrow[r, "s_m"] \arrow[d, "D"'] & T_X \\
(X+A)^+ \arrow[r, "\tilde{s}^+_{m-1}"'] & T_X^+ \arrow[u, "\#_X"']
\end{tikzcd}$$

where (1) $T = \Delta_t + {}_{+\{\perp\}}\tilde{\Delta}$,

(2) $\#_X$ is a bijection from T^+ to T (like #),

(3) $\tilde{s}_m$ $(m \geq 0)$ is a mapping from X+A to T such that

(i) $\tilde{s}_m | X = s_m$,

(ii) For any a in A, $s_m(a) = \{\perp\}$,

(iii) $s_0(x) = \{\lambda\} + 1\{x_1\} + 2\{x_2\} + \ldots + n\{x_n\}$, where $D(x) = (x_1, \ldots, x_n)$.

Thus for any x in X with $D(x) = (x_1, \ldots, x_n)$, the following expression holds for any $m \geq 1$:

$$(*) \quad s_m(x) = \{\lambda\} + 1\tilde{s}_{m-1}(x_1) + \ldots + n\tilde{s}_{m-1}(x_n).$$

By the way we consider the order in T in the following way.

Notation. $\alpha(x_1,\ldots,x_n)$ denotes the tree α with the leaves $x_1,\ldots,x_n$, and $\omega = \alpha(x_1,\ldots,\alpha_i,\ldots,x_n)$ denotes the tree replacing x_i with the tree α_i.

For any ω, α in T_x, we define the order $>$ by:

$\omega > \alpha$ iff $\omega = \alpha(x_1,\ldots,\alpha_i,\ldots,x_n)$ for some x_i and α_i in T_x.

Then it is easy to show that $(T_x,>)$ is a poset. By this definition and the equation (*), we have $s_m(x) > s_{m-1}(x)$ for any $m \geq 1$. T_x has a closure property, so we get

$$s(x) = \sup s_m(x) \text{ in } T.$$

Therefore by taking the sup of (*), we get

$$s(x) = \{\lambda\} + 1\tilde{s}(x_1) + \ldots + n\tilde{s}(x_n). \qquad \square$$

5. Concluding remarks

We have given a mathematical framework of a dictionary in which semantics of words are specified by elements of our semantic space. Our basic idea was based on the observation that any word x in a dictionary is expressed with a finite sequence $x_1,\ldots,x_n$ of words in it and that semantics s(x) of a word x is obtained by combining the semantics $s(x_1),\ldots,s(x_n)$ of words $x_1,\ldots,x_n$.

This approach may be the first trial on dictionary semantics and will also support the semantic description of data models. We will use this framework for the semantic description of our data model called bottomless data model [4].

The next target of our study will be a rational property of dictionary semantics [5]. Another interesting mathematical problem is to solve the dictionary domain equation: $X^n + \ldots + X \cong X$ [6].

Acknowledgements

The author would like to express his sincere thanks to Dr. Setsuo Arikawa for his valuable suggestion, constructive comments and encouragement during the course of this study. The author wishes to thank Mr. Satoru Miyano for critical reading the manuscript and helpful comments. The author also wishes thank Prof. Setsuo Ohsuga for his helpful comments.

References

[1] Borkin, S.A. DATA MODELS: A SEMANTIC APPROACH FOR DATABASE SYSTEMS, MIT press, Cambridge, (1980).

[2] Winograd, T. Understanding natural language, Academic Press, New York, (1972).

[3] Stein, J and Urdang, L. (Eds.) THE RANDOM HOUSE DICTIONARY of the ENGLISH LANGUAGE, Random House, New York, (1967).

[4] Shiraishi, S. A bottomless data model, in preparation.

[5] Sato, M. Algebraic Structure of Symbolic Expressions, TR-82-05, Fac. of Sci., Univ. of Tokyo (1982).

[6] Shiraishi, S. The solution of the dictionary domain equation, to appear.