Research into Denotation of Linguistic Units by Ensemble

An Song Duk

Faculty of Korean language and Literature

The great leader Comrade **Kim Il Sung** said as follows:

"Long-term research should be conducted with a view to opening up new scientific fields and introducing the latest developments in science and technology widely in the national economy." ("**Kim Il Sung** Works" Vol.35, p.313)

It is one of important tasks of the linguists to introduce advanced science and technology and open up new branches of science and technology. This necessitates further concretizing the applied linguistics which is making rapid progress on the basis of modern science and technology and gradually expanding the range of its application.

To clarify the structure and system of the language by introducing mathematical techniques and computer aid into the study of language which is one of the social phenomena, and to put the analysis and treatment of linguistic data on a scientific and precise basis—this is posed as the vital demands in opening the borderline science between social sciences and technical engineering.

With the developments of science and technology and further intensification of the research into individual scientific fields, the mathematical methods are widely introduced into different fields such as natural and social sciences including linguistics.

The purpose of introduction of mathematical methods is visually to change formulated linguistics matters into one or more simple, logistic and algorithmically-settled mathematical ones.

Mathematical linguistics, one of technical sciences, establishes the system and expands achievements using mathematics, especially the theory on ensemble.

First of all, let us consider the definition of ensemble in view of mathematical linguistics.

Ensemble, one of the main conceptions in modern mathematics, is a primary conception which can hardly be defined by any simpler conceptions.

In general, an aggregate of some things is called an ensemble, and each object in the ensemble is called element of the ensemble. Likewise, in the language, the alphabet can be seen as ensemble of a certain number of letters, the word as ensemble of a limited sequence of well-ordered letters, and the sentence as a feasible combination of inflected words.

When the association of Korean alphabet is A,

A={ㄱ, ㄴ, ㄷ, ㄹ, …, ㅃ, ㅆ, ㅉ, ㅏ, ㅑ, ㅓ, ㅕ, …, ㅐ, ㅒ, ㅔ, …}, i.e., A is an ensemble of 40 letters.

When B is an association of Korean consonants, B is an ensemble of 19 consonants.

When C is an association of Korean vowels, C is an ensemble of 21 vowels.

Ensembles are represented by capital Latin letters like A, B and C, when objects are denoted by small Latin letters like a, b and c.

a(object)∈ A(ensemble)

When there is a clear-cut definition whether a(object) is an element of ensemble A, A is called a definite ensemble.

Ex. When M is an ensemble of Korean nominative particles, M={가, 이, 께서} is a definite ensemble.

When A is a definite ensemble thanks to the definition,

1(a∈ A)=a∈ A

0(a∈ A)=a∉ A

When it is not a definite ensemble, it is called an unclear ensemble.

For example, ensemble A is given as:

A={babies, children, boys and girls, unmarried young people, young people, middle-aged people, elderly people}

In view of the field of meanings of nouns, the meanings of these words do not have definite limits but cross one another. Referring to a ten-year-old, 47 of 100 people called them "child" when 53 called then "a boy or girl." In other words, the conception of a ten-year-old belonged to either "children" or "boys and girls" at about 50 per cent.

An ensemble is made by certain methods. There are two methods, that is, enumeration and collection of elements that satisfy a certain quality.

Enumeration is a method of making an ensemble by arranging the elements in order. It is usually used for a limited ensemble of a small number of elements.

For example, when an ensemble of Korean consonants is B, B is an ensemble made by enumeration.

Ensembles are also made by the method of arranging the elements that satisfy a certain quality P (x), i.e., M={x/P(x)}.

Ex. When all the words whose length is no more than three are collected from the Modern Korean Dictionary,

M={x| "x"≦ 3}

When there are two ensembles A and B and all the elements of A belong to ensemble B, ensemble A is called a subset of ensemble B, and is denoted as A⊆ B.

In case of ensemble A of the Korean alphabet and ensemble B of the Korean consonants, formula B⊂ A is established. In case of A⊆ B and B⊆ A, the two ensembles A and B are called the same and denoted as A=B. And in case of A⊆ B and A≠B, A is called a true subset and is denoted as A⊂ B.

Ensemble B of the Korean consonants is a true subset of ensemble A of the Korean alphabet. For example, when A is an ensemble of students in the linguistic department and B is an ensemble of honor students at the college of literature, the formulas above give us the following meanings.

A⊂ B: All the students in the linguistic department are honor students.

A=B: All the honor students at the college of literature study in the linguistic department and they all are honor students.

B⊂ A: There are honor students among them in the linguistic department.

A∩B=Φ: There isn't any honor student in the linguistic department.

Many of definitions of a set theory give logic symbols. Here are some of the important logic

symbols.

∀: arbitrary, any, all

→: come from

∃: exist

Φ: empty set

~: equal

∨: or

∧: and

The definition of subset can be represented by logical symbols.

When there is an arbitrary x, suppose "x belongs to ensemble B." is given by the proposition "x belongs to ensemble A"

When this is presented by a logic symbol, $\forall x, x \in A \Rightarrow x \in B, A \subset B$

Next, let us consider the calculation of ensemble and its rules.

Union of ensemble

An ensemble which consists of all the elements which belong to ensemble A or B denotes A∪B.

In general, Aα, $\alpha \in I \rightarrow \bigcup_{\varepsilon \in I} A_{\alpha}$.

Product of ensembles

An ensemble which consists of all the elements which belong to both ensemble A and B is given by $A \cap B$ or $A \bullet B$.

In general, $A_{\alpha}, \alpha \in I \rightarrow \bigcap_{\alpha \in I} A\alpha$.

Ensemble of remainder

An ensemble of the elements that belong to A but not to B is called remainder of the deduction of B from A, and is denoted as A\B. In this case of $B \subset A$, A\B is called the remainder of the deduction of ensemble B from ensemble A and is denoted as $\overline{B} \vee B^{c}$.

Rules of calculation of ensembles:

1) $A \cup A = A, A \cap A = A, A \setminus A = \varphi$
   $A \cup \phi = A, A \cap \phi, A \setminus \phi = A$

2) $A \cup B = B \cup A, A \cap B = B \cap A$ (commutative law)

3) $\left.\begin{array}{l}(A \cup B) \cup C = A \cup (B \cup C) \\ (A \cap B) \cap C = A \cap (B \cap C)\end{array}\right\}$ (associative law)

4) $\left.\begin{array}{l}(A \cup B) \cap C = (A \cap C) \cup (B \cap C) \\ (A \cap B) \cup C = (A \cup C) \cap (B \cup C)\end{array}\right\}$ (distributive law)

5) $\overline{(A \cup B)} = \overline{A} \cap \overline{B}, \overline{(A \cap B)} = \overline{A} \cup \overline{B}$ (negative law)

Next, let us consider the denotation of the theory of ensemble about linguistic units.

General alphabet is a set of finite elements like letters.

For example, alphabets of Korean, English, Japanese, etc are

{0, 1, 2, …, 9}-alphabet.

Word is a finite sequence in alphabetical order among the general alphabet.

The numbers from 1 to n can be attached to a word.

$$\alpha = \alpha_1 \alpha_2 \ldots, \alpha_n$$

$$\beta = \beta_1 \beta_2 \ldots, \beta_m$$

The number of consonants and vowels that contains the word α is called the length of $\bar{\alpha}$.

$$\|\alpha\| = n, \|\beta\| = m$$

The total number of words whose length is k among the general alphabet A which is consisted of n letters, is equal to n^k.

Sentence is a finite sequence in which words of general alphabet are written in order in compliance with a certain rule. In other words, sentence can be seen as a set of the finite number of words.

A certain ensemble of words in the general alphabet A is called the language of alphabet A. In other words, language can be seen as a true subset of A^* which comes from the alphabet ensemble A.

We should further accelerate the scientific and theoretical progress of linguistics by making deeper studies of mathematical linguistics, and thus support the implementation of Supreme Leader **Kim Jong Un**'s plan of building a thriving socialist country.