Title	Ontology of language : linguistic types and tokens
Sub Title	
Author	飯田, 隆(Iida, Takashi)
Publisher	Centre for Advanced Research on Logic and Sensibility The Global Centers of Excellence Program, Keio University
Publication year	2008
Jtitle	CARLS series of advanced study of logic and sensibility Vol.1, (2007.) ,p.285- 295
JaLC DOI	
Abstract	
Notes	Part 4 : Philosophy and Anthoropology
Genre	Research Paper
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO12002001-20080331-0285

16 Ontology of Language: Linguistic Types and Tokens

Takashi Iida[1]

[1] Department of Philosophy, Keio University

I. Ontology and Epistemology of Language

In everyday talk, as well as in scientific discourse, we usually assume that there are entities that are called "languages". There seems to be no reason to doubt such an assumption. There are various languages now present across the world, some are very familiar to us (such as our native tongue), and others whose very names are unknown to us. But what do we mean exactly when we say that such a language "exists"? For example, we are frequently told that the number of the languages that existed previously exceeds the number of languages that are in present use. This assertion suggests that for a language to "be" or "exist", it is necessary that there are people who use it at the present time. If someone says that even a language with no current users still exists, so long as there remain books or other documents written in it, you may reply that these written documents are merely traces of the fact that a language used to exist.

However, the matter is not so simple. Let us consider a small linguistic unit, such as a word, instead of a bigger unit like the entire Japanese or English language. Take the word "rain", for example. You have probably used this word countless times, and have seen or heard it used in many different places. What makes it possible for you to know that you have encountered this word

on different occasions? You believe that you have encountered the same person or car (which you had seen on different occasions) because you believe that such a person or car continues to exist somewhere without being seen or heard by us. Can we imagine that the word "rain" continues to exist in the same way that a person or a car does? If you say that it exists, even when the word is not being actively used by anybody, where does it exist while nobody is using it? Does a word then exist only when somebody uses it? It doesn't seem so, because when we say that English contains the word "rain", we will not say that the word exists only at particular times and in particular places. Instead, we are inclined to think that the word exists in a way that cannot be located in some particular time or place. If we think a word exists in such a way, should we also think that the language to which the word belongs exists in the same way? Should we not think that a language is something that exists independently of the fact that there happens to be some people who currently speak it?

Such questions belong to a philosophical discipline called "ontology". Two of its typical questions are the following: What kind of entities are there among those that exist? What relations hold true between different kinds of entities? Among the questions asked in the beginning of this article, there are at least two ontological questions, one of which asks whether the existence of a language depends on its speakers, and the other of which asks what kind of entity a language is. It was also asked what kind of an entity a word like "rain" is, and in order to answer this question, it was asked further what it meant to say that we had encountered the same word again. All of this suggests that what kind of an entity something is closely connects with what it means to say that we have "the same" thing of that kind. The problem of the identity of a language arises frequently and in various ways. Should we say that the language used in the Man'yo era is the same language as the one we use now? In the case of modern Japanese, there are numerous differences in vocabulary and accent among its speakers. Is it correct to say that all Japanese people speak the same language? Such questions themselves do not belong to ontology of language, but rather to the linguistic study of Japanese. The question that belongs to ontology of language is a more general one, one which asks to identify the general conditions for the very identity of a language.

Another philosophical discipline is epistemology. To this discipline belongs the question how the understanding of a language is possible. Everyday

linguistic exchanges involve more than addressing and receiving a predetermined number of remarks; we can understand long speeches and intricate paragraphs. How is such a feat possible? It is not because we have heard or seen such speeches and paragraphs. We can know or think many things that are new to us through using a language and, for the most part, we do so without any conscious effort on our parts. What kind of epistemological mechanism should we assume exists, in order to explain such abilities?

This question is closely related to another important question that also belongs to the epistemology of language, namely, the nature of our cognitive relation to the native tongue. When a child reaches a certain stage of development, she understands a language spoken around her, and uses it herself. Such an achievement is usually regarded as an intellectual one, unlike the earlier achievement of becoming able to walk. But why is that so? Is it because there is some piece of knowledge that makes it possible for a child to become a speaker of a language? If so, what is it?

Let us assume that there is something that we should know in order to become a speaker of a language. If this is true, every speaker of a language should know this something. However, this piece of knowledge must have properties that are much different from those properties typically possessed by a usual sort of knowledge. For one thing, usually we cannot say what we know about our native tongue; though a speaker of Japanese can use correctly the particles "te-ni-wo-ha" most of the time, she will be at a loss if she is pressed to explain exactly what rules she follows in using them. Linguistic knowledge is both fundamental to our cognitive lives and unique in character, and these facts make linguistic knowledge a very important subject to epistemology in general.

Our focus in this report is on the ontology of language, not the epistemology of language. However, as the various philosophical disciplines are closely connected to each other, we cannot discuss the problems in the ontology of language without also encountering those in the epistemology of language.

II. Types and Tokens of Linguistic Expressions

How do we proceed to specify a language in an explicit way? First, we should specify a vocabulary, a set of words that form the fundamental elements of the language. In many cases, words are classified into several classes, which are called categories according to the role they play in sentences. A vocabulary must be a finite set. Otherwise, we could not decide whether a given expression belongs to a given language or not. After having specified a vocabulary, we should state the rules for forming complex expressions from the words in the vocabulary. We form a phrase by combining words, and a sentence by combining phrases. Let us call a word a simple expression, and a phrase or sentence a complex expression.

If a recursion can be repeated without restrictions in a language, it is possible to form an infinite number of complex expressions, and hence, an infinite number of sentences. Of course, it is impossible that the whole infinite variety of these sentences is actually produced; on the contrary, it is certain that there are sentences that will never be pronounced or written. Even though each sentence is finite in length, its length can be arbitrarily large if the language allows unrestricted recursion. Thus, if the current assumption that our universe is finite in both space and time is correct, it happens that we don't have time or space enough to present certain very long sentences. It is not difficult to imagine such sentences. Just take a simple rule for forming logical conjunctions in a language for propositional logic, and apply it many times over to just one propositional letter, say, P.

Even in a language where only a finite number of sentences are possible, there may be many complex expressions that are permitted by grammar and yet never encountered in reality. Suppose that Japanese were a finite language where recursion is not unrestricted. If this were true, the language would still be complex enough to contain many grammatically correct expressions that have never been used, and will never be used in the future. Moreover, there may be Japanese sentences that happen never to be used by anybody, simply owing to pure chance.

Thus, in any language, many of the sentences that are judged to be grammatically correct will be never used, and hence, never produced in any way whatsoever. It seems only natural to say that these "unrealized" sentences

have only a possible existence (rather than a real one). One role of the grammar of a language is to give the totality of the possible sentences in the language. As even such huge sentences that cannot be contained in our universe are among the "possible" sentences, the possibility we are concerned with here should not be physical possibility, but one much more akin to logical possibility.

To hold that an unrealized sentence has only a possible existence is to think that existence for a sentence is the same as existence in space and time, and hence, that a sentence is a spatiotemporal object. If we make the distinction between possible existence and actual existence for sentences, we must hold that the actual existence of a sentence is nothing but its being produced in a particular time and place.

In general, our way of talking about sentences or linguistic expressions suggests that we are not treating them as spatiotemporal entities. This can be observed from the way we use the phrase "the same sentence". Please look at the following example:

> We had a terribly hot summer this year.
> We had a terribly hot summer this year.

What are we going to say about this? Here we have two sentences that are the same. But how can we say that? If there are two things of the same kind, they should be different from each other, and it is therefore impossible that they are the same. Is it not a contradiction to say that here are two dogs which are the same dog?

The key to solving this puzzle is in the fact that the word "sentence" can be used in two different ways. As is frequently the case, the apparent contradiction we have here is caused by ambiguity. When we say that we have two sentences, "sentence" refers to some spatiotemporal entity, whereas the word "sentence" in the phrase "which are the same sentence" could not refer to a spatiotemporal entity. In the latter case, a sentence is construed as something which does not exist in space and time, and is called "a sentence as a type". In the former case, a sentence is construed as something that exists in space and time, and is called "a sentence as a token". It is not difficult to see that a similar distinction applies to linguistic expressions in general.

A sentence as a token is a concrete entity that is located in a particular time and place. It can be an event consisting of a series of sounds or gestures, or a

material object like the stains of ink on a paper. Although an expression as a token is an entity which is bound to one particular time and place, we regard a linguistic expression as something that can be used repeatedly in different times and places, just like a tool. Here, we are thinking of expressions as types.

When we scold our child for her bad writing, or complain to someone else of not catching what she said, we are concerned with tokens. However, most of the time when we are talking about a language, we are concerned with types. If someone says that he missed the name mentioned, and goes on to ask what the name the other person mentioned was, it is obvious that what he wants is not a name as a token but a name as a type.

If you wish to know what sort of things constitute a given kind, you can do so by examining the conditions under which these things can be identical. Our considerations on what we understand by "same sentence" or "same expression" show that we regard sentences, or linguistic expressions in general, as types that are not located in space and time. As types can be counted and named, they should be regarded as objects. Let us call those objects which exist in space and time concrete objects, and those which do not exist in space and time abstract objects. Therefore, types are abstract objects.

III. Ontology of Linguistic Types

Typical abstract objects are mathematical objects, such as numbers and sets. These objects exist necessarily if they exist; they are not contingent beings. Is this also true of linguistic types? That is, do they exist necessarily if they exist? The answer to this question wholly depends on what the existence of a linguistic type consists in.

At least for a simple type like a word, it seems correct to say that the existence of its type amounts to the same as the existence of its tokens. It does not make sense to insist that a word has existence as a type, in spite of the non-existence of its tokens. Among the possible combinations of Japanese syllables, some are already realized as Japanese words, and some are not. Those that do not constitute actual words in Japanese must be regarded only as possible word types, and hence, have no existence.

As a token is a concrete entity, its existence is always a contingent matter. Therefore, a simple type like a word, whose existence depends on the existence

of its tokens, is a contingent abstract object. Although you might find this conclusion very strange, it is not so strange as it seems on the surface. Take any word in Japanese. It is entirely contingent that it belongs to the current vocabulary of the language. The matter is more obvious for a type entity other than linguistic expressions; there is no doubt that a film such as *Tokyo Story* or the symphony known as *Jupiter* came to exist only contingently.

As a type does not exist in time (unlike a token), it would be senseless to say that a type begins to exist at a certain time or ceases to exist at a certain time. Nevertheless, we tend to think that a word ceases to exist as a type if nobody will use it. Similarly, we tend to think that the film *Tokyo Story* itself will no longer exist when all the prints, DVDs, scenarios, etc. are lost.

Though it seems natural to think in this way, the "naturalness" of this way of thinking derives from the fact that our conception of existence is heavily biased towards existence in time. Compared to tensed existential predicates like "existed", "exists" and "will exist", an untensed existential predicate always sounds unnatural. In particular, we are inclined to think that the existential predicates which are appropriate for a type entity are tensed ones, because the existence of a type entity depends on the existence of tokens (which are spatiotemporal entities). However, as a type entity is not in space or time, a tensed predicate cannot apply to it.

What might be helpful for us is a distinction between real and unreal: This distinction applies to both temporal existents and non-temporal existents. Thus, we can use the predicate "is real" as an untensed existential predicate. As tokens are temporal entities, for them to be real means the same as one of the tensed existential predicates applying to them. For types, which are not temporal entities, their reality cannot be the same as the applicability of such tensed existential predicates.

In what circumstances is a word or a film as a type a real word or film? It is real when there are corresponding tokens; that is, when a tensed existential predicate applies to them. If such a token is going to exist in the future, we may not always know with certainty that a type in question is real. Such a limitation in our knowledge is only an epistemological problem, which is independent of ontological concerns. What is obvious is that a type is real insofar as there are some past tokens or future tokens, even if there are no tokens of it at present. Even though every token of *Tokyo Story* is lost, this fact does not make *Tokyo Story* an unreal film; similarly, even though one of

our words now in use shall disappear completely in the future, and thus the future generations will have no clue as to its existence, the word will not become an unreal word.

I hope our conclusion that a simple type such as a word is a contingent abstract object has now become less strange, by the considerations we have offered here. The next thing we should consider is what constitutes the existence of a complex type, such as a phrase or a sentence.

The most straightforward answer to this question is that the existence of a complex type is embodied in the existence of its tokens, just as in the case of a simple type. According to this answer, a grammatically correct phrase or sentence that is never to be produced in the entire history of the universe does not exist. Though this seems to be a perfectly reasonable conclusion, there is one problem here.

Let us remember that a word is a contingent being. Hence, there are possible words that are not in fact real. It is easy to imagine a possible word which might have been a Japanese word but never becomes one. For example, the expression "agoya" is a possible combination of Japanese syllables but it is not a Japanese word. Let us suppose that this combination of syllables never becomes a Japanese word in the future, also. Still this expression might be a Japanese word, say, an adverb. Thus, this expression is a type entity which exists only in possibility. Moreover, we can form a possible phrase or a sentence like "agoya waratta" or "Taro wa agoya waratta", using this possible word. If you compare these possible expressions with those grammatically correct expressions which happen to have no tokens, you might have the impression that the latter kind of expressions are "much more real" than the former. There is a reason for such an impression: The possibility of the latter kind of expressions is already contained in an actual language, whereas the possibility of the former kind is not.

One way to be faithful to such an impression is to hold that a complex type such as a phrase or a sentence exists not only possibly but actually, if the grammar of the language allows its formation. More formally, this can be stated thus:

> A complex type α exists if and only if all simple types which are components of α exist.

Here it is assumed that the term "simple type" covers not only words but also basic operations, like concatenating words. For such an operation its existence is the same as the existence of its tokens. As the necessary and sufficient condition for the existence of a simple type is the existence of its tokens, the following is true.

> A complex type α exists if and only if all simple types which are components of α have their tokens.

Our claim that a complex type like a phrase or a sentence has existence in spite of the nonexistence of its tokens agrees with our way of speaking when we wish to state something general about linguistic matters. Consider one such generalization: In Japanese a new noun phrase can be formed from two noun phrases by using the particle "no". There are two things worth noting in the way this generalization is stated. Firstly, no one thinks that this generalization applies only to those noun phrases which happen to have tokens. Secondly, no one thinks that this generalization applies even to those possible noun phrases that contain words that are not part of actual Japanese. In other words, the quantificational domain for such a generalization over types consists of all of the expressions which can be formed from the actual words, and it does not matter whether these expressions have their tokens or not. In the light of the close tie between existence and being in the domain of quantification, this means that the existence of tokens is not required for a complex type to exist.

In sum, there are three ways for a linguistic type to exist.

(a) A linguistic type exists as a type with a token.
(b) It exists as a type without any token.
(c) It exists only possibly as a possible type.

A simple type like a word always belongs to category (a). A complex type such as a phrase or a sentence belongs to either (a) or (b). To category (c) belongs a word that is not actual or a complex expression which contains such a word. As we argued above, category (a) consists of contingent abstract objects. The same can be said of category (b), because the existence of a complex type depends on the existence of simple types, which in turn depends

on the existence of their tokens. Lastly, category (c) consists of the contingent abstract objects that happen not to exist.

IV. The Identity of a Language

I have been arguing that a language is a system consisting of type entities, both simple and complex. Among simple types, there are those that operate on other types and return new types, and others that are operated on. Take some language, L. If we designate the totality of the types of the former kind in L by C, and the totality of the types of the latter kind in L by V, then the totality of the complex types of L can be designated as C(V). If we think C always contains a constant operation, that is, the operation which returns the input itself, then C(V) designates the totality of the simple and complex types of L. But there is more to a language. A language is not only a syntactic structure, but also a semantic system. So, let us consider a relation M, which relates each element of C(V) to its meanings. As some elements of C(V) may not have meaning by themselves, M need not relate meanings to every element of C(V). M is not a (partial) function but a relation, because some elements of C(V) may have more than two meanings. There may be two different meaning relations, M and M′, for the same C(V); in such a case, we have two different languages. Thus, a language L can be represented as a pair of a system of types C(V) and a meaning relation M.

$$L = \langle C(V), M \rangle$$

This characterization of a language is very crude. In particular, it is certainly unrealistic that the semantic aspect of a language could be adequately represented by the meaning relation M. However, apart from such defects, there seems to be a fundamental difficulty in characterizing a language in this way. The difficulty is caused by the fact that both of the elements which constitute the language are abstract objects. Abstract objects do not exist in time. Therefore, it is meaningless to talk about their changes. Obviously, this conflicts with our belief that a language changes and does so constantly.

You might say that our way of characterizing a language is appropriate for only artificial languages, such as the languages of logic, and is not suitable to

natural languages like Japanese and English. However, there is no difference in ontological status between an artificial language and a natural language. Both of them consist of words and sentences, which are type entities. As type entities are abstract entities, it should follow that a language is also an abstract entity. As an abstract entity, a language must be incapable of change. How can we talk about the various changes in English or Japanese without any hint of conflict or difficulty?

We have come to such an impasse because our present way of talking about a language is misleading, in two respects. First, what is called a change in a language is not a change in the language itself, which is impossible, but a replacement of the language by a new language that is similar to it. Second, the names of natural languages like "Japanese" and "English" are not proper names that designate a single language, but common names that apply to numerous different languages. If these two points are appreciated, it will be obvious that there is no conflict between the fact that a language is an abstract entity and the reality of linguistic changes.