The "beginning paradox":

An empirical and conceptual problem in the current Usage-based Model of language[*]

Masato YOSHIKAWA

1. Introduction

In a field called *Cognitive Linguistics*, the process of language learning and the linguistic representation in memory are usually structured from an empiricist view of language, or more specific, language acquisition: namely, the Usage-based Model of language (UBM). This model assumes that language is or can be acquired through linguistic experiences and our general cognitive ability such as attention, categorization, and reasoning. This contrasts sharply with a rationalist and nativist theory of language known as *Generative Grammar*, which assumes that language is a product of an innate built-in knowledge, that is, Universal Grammar (UG).

Since the proposal of UBM by Ronald Langacker in 1987 (Langacker 1987), a large number of scholars have advocated the model and applied it to various phenomena related to language, such as morphophonological system (e.g., Bybee 1985, 1995, 2001), grammatical constructions (e.g., Boas 2003; Goldberg 2006), and the process of language acquisition in general (e.g., Tomasello 2003). Owning to the studies by those scholars, UBM has become more and more elaborated and obtained larger amount of supporting data.

In spite of the progress just described, however, it should be pointed out that UBM possesses one crucial problem in accounting for the process of learning language: that is, it cannot account for the very beginning of the language acquisition. Put differently, UBM does not tell us how to begin acquiring language. We shall call this discrepancy the "beginning paradox."

This paper reveals the detail of the paradox and presents the way to solve it. Specifically, 1) UBM's account for the process of learning linguistic elements will be reviewed in detail and 2) the deficiency of it will be pointed out; after that, 3) the alternative model of language learning will be presented: the model is named *Exemplar and Indices Model* (EIM).

[*] I would like to express my gratitude to the following people for their comments on my presentations or written papers related to the topic in this paper: Professor Ippei Inoue (Keio University), Professor Yukio Tsuji (Keio University), Associate Professor Isamu Takahashi (Keio University) and Fuminori Nakamura (Keio University). In addition, I would also like to thank Kow Kuroda (National Institute of Information and Communications Technology) for inspiring me to think of the basic idea of this paper. However, needless to say, the remaining deficiencies are, if any, due to my lack of ability.

2. What is the Usage-based Model

Usage-based Model of language (UBM) was proposed by Ronald Langacker in the research program named *Cognitive Grammar* (Langacker 1987). He claims that linguistic representation we have in mind is full of specific forms including inflected nouns such as *toes*, *beads*, and *walls* (examples employed by Langacker 1987: 46) together with what we call *base forms* (such as *toe*, *bead*, and *wall*) from which inflected forms are made (Langacker 1987: 46). General properties of language such as the pluralization of a noun, which can be represented as "N + -*s*," are thought to be extracted from the specific forms as *schemas*. Therefore, in UBM, general schemas (or rules) and specific items are considered to *coexist* in our memory. Langacker regards theories which assume that rules and items are separated entities (e.g., Pinker 1991: what is called *dual-process model*, see Bybee 1995; Tomasello 2003: 237-239) as committing a *rule/list fallacy* (Langacker 1987: 27-29).

2.1. Langacker's Dynamic Usage-based Model (DUBM)

Langacker characterizes his theory of Cognitive Grammar as being usage-based theory. By the term *usage-based*, he means "maximalist, nonreductive, and bottom-up character of the general approach" (Langacker 1990: 264), which expresses an oppositional stance of minimalist, reductive, and top-down theory of language structure (as adopted by Generative Grammar). This characterization can be said to be an *attitude* toward description of linguistic phenomena or a *methodology*.

As for the psychological phenomena related to the language learning, Langacker (2000) summarizes several phenomena in order to explain the process of language acquisition. He presents the model named *Dynamic Usage-based Model* (DUBM), which focuses more on the dynamic process of learning language (Langacker 2000). The phenomena discussed in the model include, among others, *entrenchment*, *schematization* and *extension* of schemas.

Entrenchment is a process of, say, automatization or routinization. Through experiencing repetitive occurrence (or recurrence) of some event, executing or understanding the event becomes a routine. This is the process of entrenchment. The entrenched complex structure as a "pre-packaged" assembly is called a *unit* (Langacker 2000: 3-4).

Schematization is claimed to be a special case of *abstraction*, which is attained by eliminating away the difference between two or more distinct items or structures and extracting a structure shared by the two. Schematization differs from a mere abstraction in that it contains varying levels of "granularity," ranging from being coarse-grained to fine-grained. Therefore, schematization is the process through which we can construct a hierarchical knowledge with varying degree of generality. A product of schematization is called *schema* (Langacker 2000: 4).

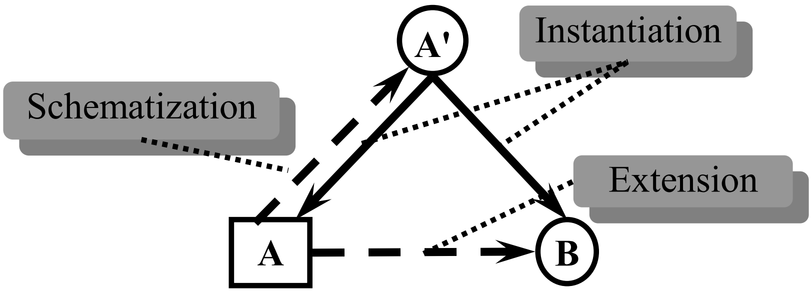

Fig. 1: Extension and Schematization
(from Figure 3 in Langacker 2000: 13, with modification)

Extension of schema is the process in which a certain *super-schema*, a hierarchically higher schema, newly instantiates a novel *sub-schema*, a lower schema, so that the category including such schemas is extended (Langacker 2000: 12-13). When a previously-entrenched schema *A* is further abstracted into super-schema *A'*, the process is schematization; when the super-schema *A'* newly instantiates another sub-schema *B*, the process is extension. Schematization can be said to be "upward" growth, and extension, "outward" growth (Langacker 2000:12). Fig. 1 depicts the process of extension and schematization.

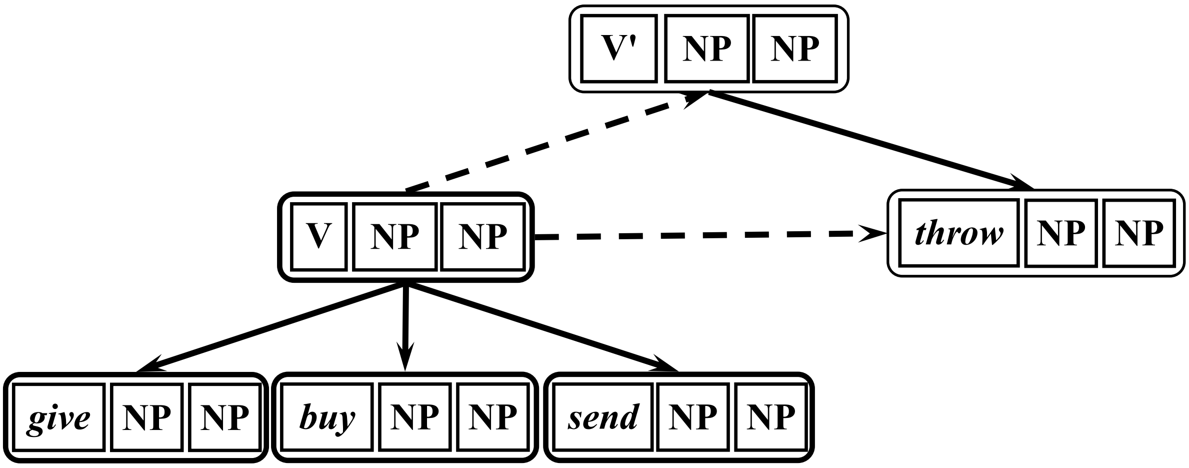

Fig. 2: A simple example of schematization and extension

Here let us see a simple example. Suppose we, through entrenchment, have learned a schema [V NP NP] where *V* denotes a verb and *NP* a noun phrase, with verbs *give*, *buy*, and *send*. This is a schema of a verb phrase of *Ditransitive Construction* (e.g., *give me a book*, *buy her a ring*, *send him a letter*, etc). At this time, if we newly hear an instance in which a verb *throw* is used in the same schema [V NP NP] (i.e., [*throw* NP NP]), we extend the current schema in such a way that the schema includes [*throw* NP NP] as an instance of [V

NP NP]. In consequence, we obtain an extended schema [V' NP NP] (see Fig. 2).[1]

In acquiring a language, it is necessary for us to learn such an abstract structure as depicted above in Fig. 2. However, due to the very abstractness, we cannot learn those structures in a single step. In fact, language-learning infants cannot abstract grammatical knowledge until they get to a certain stage of development (Bates & Goodman 1999; Caselli, Casadio & Bates 1999). Then, what process do we go through in the course of language acquisition until we have mastered language?

2.2. Tomasello's acquisition model

In order to know what the specific process of language acquisition is like, it can be profitable to see the detail of Tomasello's (2003) discussion. He describes the process from the usage-based point of view and, based on the description, also presents a usage-based developmental model of language acquisition.

The basic scenario of the acquisition Tomasello describes is, in a word, a gradual and piecemeal process of the *rote learning* and the *abstraction of schematic patterns* from rote-learned items. Based on close observation of the data related to infants' behavior found in natural settings or in experiments and that found in corpora, he concludes that language acquisition can be accounted for in the usage-based way: that is, language is *learnable* via linguistic experience in the true sense. The remainder of this subsection presents the brief sketch of the acquisition scenario and its developmental model presented by Tomasello (2003: 139-140).

Children begin with the most concrete type of expressions, known as *holophrases*, gained by rote learning from the speech of surrounding adults (Tomasello 2003: 36-40). The holophrases include such one-word expressions as *Phone*, *Towel*, *Bath*, *Game*, and *Make*, and, in addition, unanalyzed sentences such as *I-wanna-do-it*, *Lemme-see*, and *Where-the-bottle* (Tomasello 2003: 38).

Holophrases are thought to be only products of rote-learning. From this it follows that in order to acquire a holophrase, what children need are the following two: to segment *communicative intentions*, which make holophrases independent linguistic expressions, and to gain enough input to *remember* and to *reproduce* the phrase (Tomasello 2003: 38-40, 174). The latter is practically equal to the process *entrenchment* presented by Langacker (2000).

---

[1] Strictly speaking, we should *know* in advance that *throw* is an instance of the category *V*(erb) and, in addition, should also be able to *recognize* it as an instance of *V* in the on-line process of pursing the sentence in a certain way.

In the course of development, children grow to combine one item with another at around one and a half years old. First, infants begin to use multi-word utterances with the mere combinations of words with no syntactic relations between them (*word combinations*, Tomasello 2003: 114, 123), such as *Ball table*.

At around the same age, infants also begin to employ combinations which are more systematic in the sense that one constituent of the combinations takes the status of, as it were, "pivot," and the other behaves like slot fillers. This kind of combinatory expressions are called *pivot schemas* (Tomasello 2003: 114-117, 123-125), such as *More* ___ as in *More juice*, *More milk*, and the like.

Around their second birthday, children begin to use patterns with syntactic relations such as "subject-of" and "predicate-of." In this period, children do have "syntax," but it is not totally alike to that of adults. The major difference of syntax between children and adults is generality. Children's syntax in this period does not have general character; it is still based on lexical items, especially, verbs. This stage of syntax is called *item-based constructions*. Of the constructions, most are centered at verbs; such constructions are called *verb-island constructions* (Tomasello 2003: 117-122, 125-126), which is most important to achieve further abstraction. An example is *Throw* ___ as in *Throw ball*, *Throw can*, and *Throw pillow*.

The acquisition of these three types of patterns needs the ability of *schematization* (Tomasello 2003: 122-126). This can be identified with the process given the same name by Langacker (2000). Children should find what is shared by several holophrases and extract the shared structure such as *More* ___ and *Throw* ___.[2]

After acquiring lexically-bounded simple grammar, children start to bind many verb-islands together based on relational (not concrete) sameness. In this period, children have adult-like abstract and general constructions, which have high productivity (Tomasello 2003: Chapter 5). The abstract kind of constructions includes, for instance, *Ditransitive Constructions* (e.g., *John gave me a book*, *Mary sent him the letter*, etc), *Resultative Constructions* (e.g., *He hammered the metal flat*, *I laughed myself silly*, etc), *Middle Constructions* (e.g., *The book sells well*, *The shirt washes easily*, etc), and so on.

As is evident, abstract constructions need not have anything concrete in common among the instances of the same construction. There is nothing shared, for example, by the two sentences *John gave me a book* and *Mary sent him the letter*, both of which are instances of Ditransitive Construction. What is shared is said to be the *relationship* between the

[2] Almost the same steps described so far are assumed by Lieven, Pine, & Baldwin (1997: 190-192).

participants in the construction (Tomasello 2003: 163-166). The relation found in Ditransitive Construction in question is like a causal relation in *the transfer of gift to receiver by giver*. The transfer is encoded by the verb (i.e., *give* and *send*) and *gift*, *receiver*, and *giver* correspond to the last NP (i.e., *a book* and *the letter*), the middle NP (i.e., *me* and *him*) and the first NP (i.e., *John* and *Mary*), respectively (Tomasello 2003: 166; see also Goldberg 1995).

Finding the relational structure shared by several item-based constructions is claimed to be achieved through the ability called *Analogy* (Tomasello 2003: 164-169). Analogy in this sense is also called *Structure Mapping* (e.g., Gentner & Markman 1997). This ability is somewhat different from *schematization*, which is the ability to find *concrete* properties shared by two or more items or the like and extracting structures with varying abstractness (Tomasello 2003: 164).

To summarize, language acquisition can be said to proceed through the following process: first, children segment phrases based on communicative intention from fluent speech and learn the segmented phrases as *holophrases* by rote (*entrenchment*); second, they find similarity between the rote-learned holophrases and form several item-based concrete patterns, that is, *word-combinations*, *pivot schemas*, and *item-based constructions* (*schematization*); third and finally, they bundle several item-based patterns together into the abstract kind of grammatical constructions such as Ditransitive Constructions based on the relational sameness (learning with *analogy*).

3. The “beginning paradox”

As reviewed so far, the current Usage-based Model seems to succeed in accounting for how the language learning proceeds. However, several specific problems aside, we cannot ignore at least one crucial problem in the model: it cannot account for how the language acquisition begins. In this section, the detail of the problem is revealed through the critical investigation of the acquisition process assumed by the current UBM.

This section consists of three subsections: the first one provides a preliminary discussion to the heart of the problem; the second and the third offer the heart of the problem, the former of which presents a strong argument incompatible with UBM’s assumption, and the latter of which, as a consequence of the former, reveals the fundamental deficiency of the current UBM, i.e., the “beginning paradox.”

3.1. Preliminary

Before looking into the heart of the problem, it should be pointed out that even the most concrete type of expression, i.e., *holophrase*, cannot be learned without abstraction or schematization. This fact is indirectly manifested by Tomasello (2003) in the statement that acquiring holophrases needs segmentation of communicative intentions (Tomasello 2003: 38-40, as reviewed in section 2.2. of this paper). Fluent speech spoken by surrounding adults given to infants is, needless to say, *concrete* and never tells infants what abstract structures or the like it has. Consequently, in order to obtain any abstract structure, infants need to find similarity among vague speeches they have heard by themselves. It is true that one holophrase has only one abstract pattern of sounds; in other words, one and the same holophrase is *phonologically constant*. However, the phonological sameness does not mean the concrete sameness: phonemes are without doubt abstract elements.

There are, however, a few *a priori* factors to find abstract properties in concrete speeches: among them is *perceptual similarity*. Speech is composed of *sounds* produced by a vocal cord and other vocal apparatus. We humans cannot differentiate infinitely varying properties lying in the external world; our perception is innately navigated to construct a certain kind of structure in mind. Therefore, it is highly reasonable to assume that if there are some concrete properties shared by a phoneme, children can find abstract sound patterns, that is, *the sequences of phonemes*, based on perceptual similarity found in fluent speech. In fact, this is probably an assumption widely taken by the theory of language acquisition.

However, is this really true? As will be seen in the next subsection, it has to be said that the assumption is highly dubious. It is dubious because it possesses the following two problems: 1) it does not seriously take into consideration the matter of *segmentation*; 2) as a consequence of the first problem, it cannot explain the mechanism of *similarity judgment* necessary for learning linguistic expressions or patterns. The next subsection presents the detail of the two problems.

3.2. The problem of segmentation

As for the first problem mentioned just above, it is profitable to review the findings in the field of *Artificial Language Learning* (ALL) studies. ALL studies investigate human ability of *statistical learning* employed in language learning, using experimental paradigm (Saffran 2003). Among several interesting findings in the field, the best-known one is probably that concerning the *word-segmentation task* (e.g., Saffran, Aslin, & Newport 1996; Saffran, Newport, & Aslin 1996; Aslin, Saffran, Newport 1998).

Adults' fluent speech infants hear has no *a priori* perceptual cues indicating word boundaries such as pauses between words. Therefore, language-learning infants should find word boundaries based on information which is not purely perceptual. ALL studies reveal that this task can be accomplished by using statistical information found in fluent speech (Saffran, Aslin, & Newport 1996; Saffran, Newport, & Aslin 1996; Aslin et al. 1998).

In English, there are several biases in the alignment of phonemes, i.e., *phonotactics*, within and across words. For example, no English word begins with the phoneme /ŋ/ and ends with /h/; there are few words begin with the syllable /ti/ while quite a few words end with the syllable: *beauty*, *duty*, *guilty*, *kitty*, *naughty*, *pretty*, *treaty*, *witty*, and so on. Therefore, if children can adequately find such biases, they can employ those biases in segmenting words from fluent speech.

The results of experiments in ALL studies show that children can actually find biases in fluent sound sequences. The input stimuli employed in the experiments are such non-existent "linguistic" sequences as *bidakupadotigolabubidaku…*, which contains several recurrent units, i.e., "words," such as *bidaku* (Saffran, Newport, & Aslin 1996).

It is not fair to make no reference to the fact that Tomasello also reviews the findings of ALL studies in regard to the speech segmentation (Tomasello 2003: 28-30, 60-62). Therefore, Tomasello is surely aware of the importance of the word segmentation task on language acquisition (but never is Langacker).

However, even if infants *can* find statistic biases in fluent speech in order to segment words, it is not necessarily true that they *do* use those biases in learning language. The experiments in ALL studies employ *phonological sequence* as input stimuli which are digitally recorded human voices. Therefore, input sequences infants hear have acoustically highly constant character, which is hardly the case with those in natural settings. Speeches surrounding infants are various: phonologically identical sound patterns may vary from speaker to speaker and, moreover, even within one and the same speaker. Consequently, infants should first be engaged in the task of *normalization*, in which acoustically various sounds are clustered into one category (Cf. Goldinger 1996; Johnson 1997).

Furthermore, there is an argument for the impossibility of acoustic definition of phonemes (e.g., Port 2007b). For example, Port (2007b) shows the following fact as straightforward evidence for the impossibility: a spectrogram displays that there is no invariant feature in the sounds as instances of one and the same phoneme (Port 2007: 151; see also Cole et al. 1997). He presents a spectrogram of English syllables /di/ and /du/, which share the same consonant /d/ (Port 2007b: Fig 3). Fig. 3 is a spectrogram of Japanese syllables /ka/ and /ki/, which also shows high dissimilarity between two versions of the same consonant /k/ (The spectrogram was created using WaveSurfer 1.8.5). If acoustic definition is impossible, the segmentation task can never be performed based on the acoustic factors, which are perceptual.

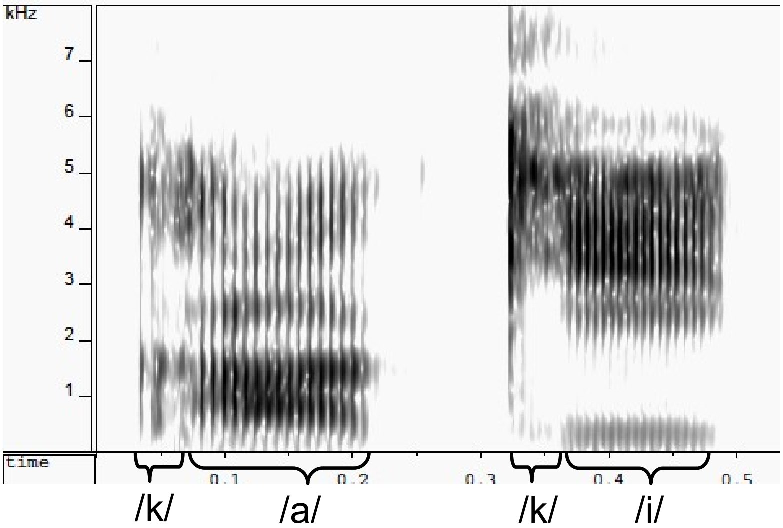

Fig. 3: Spectrogram of Japanese /ka/ and /ki/

Then, how do infants segment words from fluent speech? Port provides an answer for this question that they successfully employ *suprasegmental factors* such as prosodies: words as units in spoken language can be characterized by "circles in time," also called *rhythm* or *meter* (Port 2007a). In fact, young infants cannot use abstract sound properties such as phonemes and, instead, they efficiently employ prosodic factors (Cf. Masataka 2003). In addition, Thiessen, Hill & Saffran (2005) report that prosodically characteristic speech employed by adults (i.e., *infants-directed speech*) facilitates infants' finding of abstract patterns from the sequences.

If the first task in language acquisition, that is, word segmentation, depends crucially on prosodic factors, it may follow that language-learning infants do not extract abstract schemas such as phonemes from memorized speeches. Prosody can only tell infants *where* to segment the sequence; it never tells them *what* to segment from the sequence. Consequently, the segmented information would be stored in memory retaining highly concrete properties. This is an assumption held by Port (2007b) as a hypothesis called "rich memory," which assumes that highly concrete and episodic information is stored in memory and therefore there are no substantial abstract representations such as phonemes and lemmas. This hypothesis will be reviewed in detail in the next section (4).

3.3. The problem of similarity judgment

The segmentation problem just described leads to a more fundamental problem: the problem of *similarity judgment*. Under the current Usage-based Model, extraction of abstract properties or schemas from two or more items needs similarity judgment because abstraction is assumed to be based on the process of finding sharing point in two or more items (see 2.1). If infants can abstract schematic representations from concrete speeches and store them in memory, similarity judgment can be easily performed: infants have only to bundle the "same" abstract properties together. Therefore, the model predicts that once segmentation is accomplished, segmented expressions or patterns are, if given frequently enough, automatically entrenched in memory.

This is the story Langacker assumes. He remarks, as to the neuro-cognitive process of schematization or abstraction, as follows (Langacker 2009: 629):

> [T]he abstraction of units does not require any special mechanism; it is essentially automatic given the general Hebbian principle that occurring patterns of neural activation leave traces (in the form of strengthened synaptic connections) which facilitate their own recurrence.

He illustrates the process of abstraction or entrenchment using diagrams like Fig. 4. In the figure, the thickness of line represents the degree of entrenchment; the downward arrows denote the process of *instantiation* where items as smaller rectangles on the top of the arrows are schemas, and larger ones containing the smaller ones and arrows are instances; the arrow from left to right below the rectangles represents the time course; different characters correspond to different items.

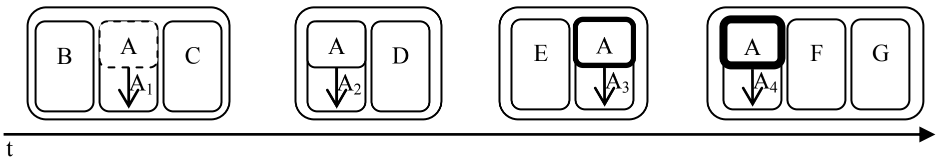

Fig. 4: The abstraction process of an expression which never occurs alone (adopted from Langacker 20009: Figure 5)

However, unfortunately, it has to be said that such a rosy story is highly gratuitous, given the argument for the impossibility of acoustic definition of phonemes and the rich memory hypothesis. The abstraction process like that in Fig. 4 above is never gone through unless the every occurrence (i.e., A_i where i = 1 to 4 in Fig. 4) of the recurrent item such as A in Fig. 4 is known as independent and as having the property identical to any other occurrences of the same item (say, A). In other words, similarity cannot be found without knowing in advance that there are one or more sharing properties.

This is surely impossible when there are no abstract sharing properties, i.e., phonemes, in recurrent items, hence a paradox: we can never learn any linguistic expressions or units until we know they are so. The process of similarity judgment assumed in the current UBM is based on the fact that we can find similarity between two or more items sharing perceptual or other kinds of properties. This necessarily leads to the “Homunculus fallacy,” or, the problem of infinite regression. The plausible model of similarity judgment should not be based on *the fact that* we can find similarities, but on *how* we can.

Then, naturally there arises a question: how is it that we can find similarities? The next section provides a probable answer to the question.

4. Alternative: Exemplars and Indices Model (EIM)

If, as Port (2007b) assumes, our memory consists of highly rich exemplar fragments, there remains a possibility that similarity judgment is performed based on several kinds of *features* assigned to cloudy fragments stored in our memory. In other words, similarity-based abstract entities such as schemas should not be fixed as representations of, say, images or concepts; instead, they could be derivationally computed in the course of a certain mental process, which is exactly what the similarity judgment is.

In order to make the possibility realistic, we need to formulate a realistic model of *feature assignment* and *similarity computation* based on features. In this section, one candidate for the model is presented: the candidate is named *Exemplars and Indices Model* (EIM). The remainder of this section offers the detail of this model.

4.1. Rich memory hypothesis

The EIM is not a novel idea. As will be reviewed in the next subsection, it is what Kuroda (2007) proposed as *Extremely Usage-based Model* (EUBM). Moreover, Kuroda’s (2007) proposal is based on the argument by Port (2007b), which presents a hypothesis called *rich memory hypothesis* mentioned above. Therefore, here let us see some details of the hypothesis.

Port (2007b) claims that abstract descriptions of sounds (such as phonemes) are only the artifacts and therefore in our memory there are no phonemes. Abstract properties are considered to be computed “in the fly” when needed (Port 2007b: 143). It is true that we have an *intuition* about the existence of abstract units like phonemes, but he states that this is no more an intuition: it is considered to be derived from the literacy training (Port 2007b: 153). As a corollary, he argues that phonology should be the study of the patterns in the speech community, not those of mental representations. Psychologically realistic description tools are not useful for the description of phoneme, and vice versa (Port 2007b: 143, 164-165).

Port (2007b) cites a lot of evidence for his hypothesis, including experimental results showing episodic recall, dialect variation, language change, frequency effect (especially of the phenomena of production such as lenition), and acoustic (i.e., physical) inconsistency

(Port 2007b: 149-152). The last one was reviewed in section 3.2 above. Here, we shall review an experimental study cited by Port (2007b), in which the first evidence was offered. The study was performed by Palmeri, Goldinger & Pisoni (1993).

Palmeri et al. (1993) claimed that in memorizing a spoken word highly concrete features such as speaker's voice property were also stored, according to the result of experiments. In the experiments, subjects were exposed to a series of words from loudspeaker and asked to judge whether a word provided last was equivalent to a word provided first. The same subject listened to several sets of words with varying number (1, 2, 4, … 64) of intervening words between the fist and the last. The different subjects were exposed to the word sets with different number (from 2 to 20) of speaker who pronounced them; that is, the number of speakers differed between subjects.

As a result, the performance was significantly higher when the first and the final words were produced by one and the same person than when produced different ones, and, although the performance got worse as the number of talkers who pronounced the words increased, the difference in performance between the 'same speaker' condition and the 'different speaker' condition was stable; that is, the performance in the 'same speaker' condition always exceeded that in the 'different speaker' condition by almost the same amount. This result strongly suggests that the process of memory recall is greatly constrained by concrete sound property such as speaker's voice property and therefore the structure of stored memory has highly concrete character.

Based on these pieces of evidence, Port proposes the hypothesis of exemplar-based, episodic memory (Port 2007b: 163), in which he stated as follows:

> [T]he dimensionality of this memory will vary from speaker to speaker but it is surely far richer than linguists have ever considered in the past. Of course, these memories may include many prototypes and abstractions as well [...]. This memory makes possible the perception of the identity of phonological fragments based on some similarity measure. Stored information includes the categories that each utterance fragment (e. g., word, morpheme, etc.) might belong to. This memory, because of its redundancy, can differentiate fragments based on their frequency of occurrence. Turning to the production problem, the speaker also uses frequency implicit in the memory to determine details of how to pronounce a fragment in any particular situation. Somehow apparently, the database of tokens of individual speech fragments (such as words) is able to influence a speaker's choice of pronunciation decisions, since speakers (especially younger ones) modify their pronunciations to be more similar to what they hear others say.
>
> (Port 2007b: 163-164)

4.2. Extremely Usage-based Model (EUBM)

Based on the argument by Port (2007b), Kuroda (2007) proposes a general model of language acquisition and linguistic memory, which is named Extremely Usage-based Model of language (EUBM). He sketches what the ontogenetic acquisition process of syntax will be like if Port's rich memory hypothesis is true (Kuroda 2007: 27). That is to say, he explores the possibility that we memorize all the expression we hear or read. Under the model, the fact that we can understand and produce the strings with complex structure, i.e., strings composed of more than one items such as words, is considered due not to our ability to combine two or more items (e.g., *Merge*, Chomsky 1995, 1999), but to the architecture of our memory in which complex expressions like sentences are stored in their integrity (Kuroda 2007: 29).

In his model, abstract or schematic units like phonemes or morphemes are reinterpreted as the "indices" by which we access the concrete exemplar memory of language (ibid). The scenario of ontogenetic acquisition of linguistic knowledge Kuroda assumes is as follows:

(1)
a. Based on his/her rich memory, a language learner develops a database in which vast number of concrete forms, *f*, and the situation where the forms are employed, *s*, are recorded in pairs (*f*, *s*).
b. A word *w* and its meaning *m*(*w*) is only an index of *f* and *s* paired with *f* (=*m*(*f*)) in the database.
c. A novel expression can be interpreted by transferring the semantic and phonological information of another expression which is the most similar to, i.e., having maximal overlaps with, it.
d. For a language user *X*, a sentence such as *Colorless green idea sleep furiously* is not acceptable or interpretable (though grammatical) because it is judged to be similar to none of the expressions he/she knows.

(Kuroda 2007: 31-32)

He radically denies the "principle of compositionality," that is, an idea that a meaning of a complex unit can be obtained by composing the meanings of the sub-units composing the unit. Instead, he considers that a meaning of a complex unit is a meaning of the unit, not composed by smaller units nor something comes from somewhere else (Kuroda 2007: 32). The (minimal) semantic unit of language is informally defined as "a minimal unit which satisfies the argument structure of a word" (Kuroda 2007: 31), though the exact nature of the unit is not clarified yet. The unit is highly similar to what is called "sentence," but, as is evident from the studies of discourse analysis or conversational analysis, in real conversations it is impossible to identify the unit called sentence and if possible, it is quite difficult (ibid).[3]

---

[3] In the original version of Kuroda (2007) (available at: http://clsl.hi.h.kyoto-u.ac.jp/~kkuroda/papers/la-with-rich-memory-full.pdf), it is mentioned that the unit in question may be equivalent to that Chafe

4.3. How are indices and exemplars at work

In this model, exemplars are thought to be fragments of memory and indices are identified with features assigned to the exemplars. It is assumed that indices are first only perceptual and hence innately available. It follows that very young infants' linguistic memory should consist of concrete perceptual features, connected with rich exemplars. This consequence is compatible with the argument in 3.2.

In the course of development, however, indices are assumed to *grow* so that abstract kinds of indices can appear: that is, some indices should be *acquired*. Acquired indices can be identified with *schemas*.[4] Therefore, under EIM schematization is reinterpreted as indices formation and, as mentioned above, schemas as indices. The growth of indices can be attained by bundling or composing lower-level features together.

It is considered to be possible that the growth is accomplished in a relative fashion. Since all the linguistic experiences are stored, abstraction can be done *retrospectively*: similarity between concrete features may be detected long after memorizing them. Put differently, the criterion of similarity judgment is allowed to be relative. Relative similarity can organize cloudy exemplars into a complex loosely tied network. This seems to be an only way to avoid the "beginning paradox."

Very simplistically, this is to say that, for example, we may not recognize that apples are similar to bananas at first; we may find them similar on the ground that bananas are more similar to apples than knives, stones, cats, and so on. This relative process of similarity judgment can be depicted in the figure like Fig. 5.

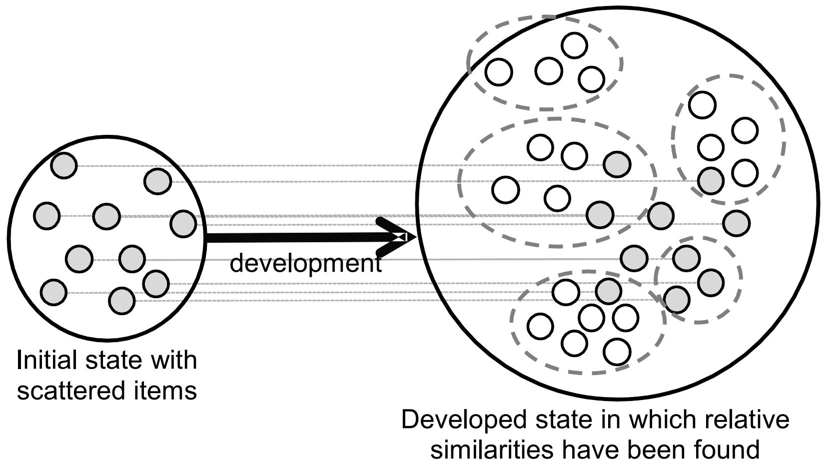

Fig. 5: Similarity judgment in a relative way

(1994) calls an "intonational unit."

[4] In fact, innately available features can also be regarded as *innate schemas*, which have been acquired in the course of human evolution.

Note that indices are only indices, that is, they cannot be substantial entities of memory. Therefore, no instances can be licensed by schemas. Interestingly, as a consequence of this assumption, EIM can avoid the problem of *overgeneralization*. EIM assumes that novel expressions are licensed based on *transfer* of form and/or meaning of stored exemplars (as seen in (1)d).

5. Concluding remarks

This paper reveals that the current version of Usage-based Model of language (UBM) possesses a crucial problem: it cannot explain how to begin acquiring language. The problem is named the "beginning paradox." The paradox is paradoxical in that the model faces with a dilemma that linguistic expressions can never learned unless they are already known as such and such expressions.

The paradox arises from the fact that the current model is unable to explain how to find similarity among linguistic items. The current UBM assumes *schematization* as primary factor to language acquisition, in which a certain similarity is found among two or more items having some properties in common. Recent studies of linguistic memory, however, offer findings incompatible with the assumption: our linguistic memory consists of far richer fragments than linguists have ever considered. These findings strongly suggest that similarity cannot be found in a way assumed by the current UBM.

Instead, if linguistic memory is assumed to be exemplar, that is, all the linguistic experience we have had are stored in its integrity as exemplars, the paradox can be avoided, which is what this paper presents as a solution to the paradox. In a word, under the assumption, abstraction is made unnecessary when language acquisition begins. Similarity judgment is thought to be done retrospectively based on the indices assigned to the stored exemplars. This is a model named *Exemplars and Indices Model* (EIM). It seems that EIM is an only way to solve the beginning paradox.

It is true that the validity of EIM should be verified by, in addition to linguistic studies, psychological and neurological research which investigates what human memory, linguistic or not, is like. Therefore, we should be sensitive to findings to be revealed in those studies, which should lead to further development of EIM.

Bibliography

Aslin, Richard N., Jenny R. Saffran, and Elissa L. Newport. 1998. Computation of conditional probability statistics by 8-month-old infants. *Psychological science* 9. 321-324.

Bates, Elizabeth, and John Goodman. 1999. On the emergence of grammar from the lexicon. *The emergence of language*, ed. by Brian MacWhinney, 29-79. Mahwah, NJ: Lawrence Erlbaum.

Boas, Hans. 2003. *A constructional approach to resultatives*. Stanford: CSLI Publications.

Bybee, Joan. 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam; Philadelphia : John Benjamins.

——. 1995. Regular morphology and the lexicon. *Language and cognitive processes* 10. 425-455.

——. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.

Caselli, Cristina, Paola Casadio, and Elizabeth Bates. 1999. A comparison of the transition from first words to grammar in English and Italian. *Journal of child language* 26. 69-111.

Chafe, Wallace. 1994. *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago Press.

Chomsky, Noam. 1995. *The minimalist program*. Cambridge, MA.: MIT Press.

——. 1999. *Derivation by phase*. Cambridge, MA.: MIT Press.

Cole, Ronald A., Joseph Mariani, Hans Uszkoreit, Annie Zaenen, and Victor Zue. (eds.) 1997. *Survey of the state of the art in human language technology*. Cambridge: Cambridge University Press.

Johnson, Keith. 1997. Speech perception without speaker normalization: An exemplar model. *Talker variability in speech processing*, ed. by Keith Johnson and John W. Mullennix, 145-166. San Diego: Academic Press.

Gentner, Dedre, and Arthur B. Markman. 1997. Structure mapping in analogy and similarity. *American psychologist* 52. 45-56.

Goldberg, Adele E. 1995. *Constructions: A construction grammar approach to argument structure*. Chicago; London: University of Chicago Press.

——. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

Goldinger, Stephen D. 1996. Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of experimental psychology: Learning, memory, and cognition* 22. 1166-1183.

Kuroda, Kow. 2007. Tettei-shi-ta yoho-kiban-shugi no moto de no bunpo-kakutoku: “Kyokutan ni yutaka-na jirei-kioku” no kasetsu de egaku atarashii suji-gaki. [An extremely usage-based model of grammar acquisition: A new scenario sketched under the hypothesis of “extremely rich exemplar memory.”] *Gengo* 36 (11). 26-34.

Langacker, Ronald. 1987. *Foundations of cognitive grammar Vol. 1: Theoretical prerequisites*. Stanford: Stanford University Press.

——. 1990. *Symbol, image, and concept*. New York: Mouton de Gryter.

——. 2000. A dynamic usage-based model. *Usage-based models of language*, ed. by Michale Barlow and Suzanne Kemmer, 1- 63. Stanford: CSLI Publications.

——. 2009. A dynamic view of usage and language acquisition. *Cognitive linguistics* 23. 627–640.

Lieven, Elena V, Julian M. Pine, and Gillian Baldwin. 1997. Lexically-based learning and early grammatical development. *Journal of child language* 24. 187-219.

Masataka, Nobuo. 2003. *The onset of language*. Cambridge: Cambridge University Press.

Palmeri, Thomas J., Stephen D. Goldinger, and David B. Pisoni. 1993. Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of experimental psychology: Learning memory, and cognition* 19. 309-328.

Pinker, Steven. 1991. Rules of language. *Science* 253. 530-535.

Port, Robert. 2007a. The problem of speech patterns in time. *The Oxford handbook of psycholinguistics*, ed. by Gareth Gaskell, 503-514. Oxford; New York: Oxford University Press.

——. 2007b. How are words stored in memory?: Beyond phones and phonemes. *New ideas in psychology* 25. 143-170.

Saffran, Jenny R. 2003. Statistical language learning: Mechanisms and constraints. *Current directions in psychological science* 12. 110-114.

Saffran, Jenny R., Richard N. Aslin, and Elissa L. Newport. 1996. Statistical learning by 8-month-old infants. *Science* 274. 1926-1928.

Saffran, Jenny R., Elissa L. Newport, and Richard N. Aslin. 1996. Word segmentation: The role of distributional cues. *Journal of memory and language* 35. 606-621.

Thiessen, Erik D., Emily A. Hill, and Jenny R. Saffran. 2005. Infant-directed speech facilitates word segmentation. *Infancy* 7. 53-71.

Tomasello, Michael. 2003. *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA.: Harvard University Press.