The Shape and Function of Phonology in Evolutionary Linguistics:
Why We can Explore Language Origins from Extant Languages, and How

Shin-ichi Tanaka*

SUMMARY: The study on language origins and evolution had been considered a pipe dream for more than a century. Even in the 1990s, there were external and internal barriers to academic research on this topic, which caused some serious clashes between linguists and biologists. However, the Merge-only hypothesis of human language evolution has opened a door to resolve the clashes and establish scientific methods for modern evolutionary linguistics. This article introduces such a new scenario and offers its empirical evidence from phonology. Specifically, I argue that the “Third Factor” played a key role in the shift from Merge and the SM interface in proto-language to human phonology. This view makes explicit the reason why we can explore language origins, the hardest problem in science, from languages at hand, and helps us establish specific methods for exploring it in phonology. Empirical tests for this view involve the typology of palatal phonotactics in Japanese and English.

Key words: precursor, recursive Merge, SM, Third Factor, Identity Avoidance, Implicational Law

1. Introduction

Since the research on the origins of language was banned by la Société de Linguistique de Paris in 1866 and by the Philological Society of London in 1872, this topic had been a taboo for more than a century. It was considered a myth, not an academic study, because it was difficult to test hypotheses empirically due to the lack of direct evidence, i.e., due to the absence of “language fossils”.

In addition to this external barrier, internal barriers also contributed to the stagnation in the study on language origins and evolution, even a century after the ban in Europe. Such barriers were internal to the whole academic community that concerns this topic. One difficulty was the lack of interdisciplinary efforts. The lack of commonalities between disciplines, such as goals, methods, technical terms, culture, and ideology, discouraged interdisciplinary communication, which in turn impeded the synthesis of significant findings from various fields. Another difficulty was the absence of platform theories for language origins and evolution, partly because the Zeitgeist in the linguistic community considered this topic premature and partly because linguistic theories in the 1980s, namely, the principles-and-parameters approach, considered Universal Grammar (henceforth, UG) too rich. This was also true for Optimality Theory that appeared in the 1990s, which had too many universal constraints. If UG was indeed so rich, language should be species-specific. From these two internal barriers, there derived another internal barrier: a conflict in which linguists claimed human language was a product of some salutation that had evolved on a unique, species-specific basis while biologists claimed human language was a product of phyletic gradualism that had evolved from similar traits or organs of other species. In this way, the competition between a saltational view and a gradual view further discouraged any interdisciplinary communication. In addition, even among biologists, there was another conflict in which some said human language was a product of homoplasy (analog) with songbirds’ vocal learning while others claimed it was a product of synapomorphy (homology) with chimpanzees’ serial learning.

However, the paradox between saltationism and...
gradualism can be resolved by making the Strong Minimalist Thesis (Chomsky 2004; henceforth, SMT) a potential platform theory for language origins and evolution in which UG is quite simple and involves only recursive Merge. The goal of this article is to resolve this apparent paradox between saltationism and gradualism by illustrating how the emergence of recursive Merge is both saltational and gradual at the same time. Evidence for that explanation will be offered from phonology.

The reasoning is organized as follows. First, I will introduce Fujita’s (2016a, 2016b) Merge-only hypothesis of human language evolution as a general scenario of evolutionary linguistics (section 2.1). I will then demonstrate how the saltational and gradual nature of recursive Merge can resolve the apparent conflict by using another hypothesis of Fujita’s (sections 2.1 and 2.2). It will also be shown there that the controversy over whether language is a product of homoplasy (analogy) or synapomorphy (homology) can be resolved as well by this hypothesis. Following the given hypotheses, I will make explicit the role of phonology in evolutionary linguistics and propose some methodologies in phonology to empirically test the hypotheses in the given scenario (section 2.3). Finally, I will provide specific evidence from Japanese and English for the hypotheses by invoking the typology of palatal phonotactics as a case study (sections 3.1 and 3.2). A well-known alternative view on the phonotactics of palatals in English will be refuted by adducing arguments against it (section 3.3).

2. The Scenario of Evolutionary Linguistics as a Platform Theory

2.1 Background

As mentioned in section 1, reducing UG only to recursive Merge has been a crucial step toward devising a new scenario and establishing specific methods for evolutionary linguistics in the 21st century. This is because a simple content of UG is a natural feature by which gradualism from other species to humans is maintained and the so-called “logical problem of language evolution” or “Darwin’s problem” can be resolved. Since the advent of the SMT, the grammar of a particular language has been considered as recursive Merge, in conjunction with the interfaces of the Sensory-Motor (SM) system for externalization or phonetic interpretation and of the Conceptual-Intentional (CI) system for internalization or semantic interpretation.

Based on the SMT, Fujita (2016a, 2016b) proposes a scenario for evolutionary linguistics as shown in (1), elucidating the origins of human language in macro-evolution, a change that separated humans from other species.

(1) Merge-Only Hypothesis of Human Language Evolution (Fujita 2016a, 2016b)

Hierarchical Grammar

Human Language

<table>
<thead>
<tr>
<th>Homoplasy</th>
<th>SM</th>
<th>Lexicon</th>
<th>CI</th>
<th>Synapomorphy</th>
<th>Autapomorphy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Merge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proto-SM</td>
<td></td>
<td>Proto-lexicon</td>
<td>Proto-CI</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Linear Grammar

e.g. vocal learning in songbirds
e.g. serial learning in chimpanzees

Note: Homoplasy (analogy) is a similarity between traits of species that is not derived from an immediate common ancestor, like wings of birds and bats. Synapomorphy (homology) is a similarity between traits of species that is due to their immediate shared ancestry, like limbs of vertebrates. Autapomorphy is a derived trait that is unique to a given species, like loss of limbs of snakes.
Here, the evolution from proto-SM to human SM is gradual by way of Merge only, and it is a case of homoplasy that is analogous to songbirds’ vocal learning. On the other hand, the evolution from proto-CI and proto-Lexicon to human CI and human Lexicon is gradual by way of Merge only, and it is a case of synapomorphy that is homologous to chimpanzees’ serial learning. In either way, the only difference from other species is the emergence of recursive Merge, which is thought of as a case of autapomorphy.

What we call “recursive Merge” is not simply an operation related to sentence complementation or category percolation. Recursion is a mathematical term, which is defined as a process of repeating a function, each time applying it to the result of the previous stage. That is, the output of an operation becomes an input of the next operation. Thus, if we call a simple operation of pairing two discrete elements into an unordered set “core Merge”, following Fujita (2009, 2014), then recursive Merge, which makes the output of the previous operation an input to the next as in (2a), is indispensable for constructing hierarchical structures like the ones in (2b).

(2) Crucial Nature of Merge
a. core Merge: 
  \[(a, \beta) \mapsto \{a, \beta\}\] (one operation)
  iterative Merge: 
  \[(a, \beta, \gamma) \mapsto \{a, \beta, \gamma\}\] (two operations)
  iterative Merge: 
  \[(a, \beta, \gamma, \alpha) \mapsto \{a, \beta, \gamma, \alpha\}\] (three operations)
b. hierarchical structures

c. core Merge: 
  \[(a, \beta) \mapsto \{a, \beta\}\] (one operation)
  iterative Merge: 
  \[(a, \beta, \gamma) \mapsto \{a, \beta, \gamma\}\] (two operations)
  iterative Merge: 
  \[(a, \beta, \gamma, \alpha) \mapsto \{a, \beta, \gamma, \alpha\}\] (three operations)
d. non-hierarchical linear structures

What is contrastive with the operations in (2a) is “iterative Merge”, which is a simple repetition of core Merge and does not make the output of the previous operation an input to the next as in (2c). The difference lies in whether the combination is binary or non-binary. The structures in (2d) result from repeating core Merge once, twice, and thrice, respectively, which are not hierarchical but flat and linear. These are typical structures observed in the finite-state grammar of songbirds like Bengalese finches, which repeat sets of sound chunks in some linear order, as illustrated in (3).

(3) Bengalese Finch Song (Berwick et al. 2011)

a. sets of sound chunks

b. finite state grammar

c. possible songs
  ab-cde-fg, ab-cde-ab-cde-fg, ab-cde-fg-ab-cde-fg, etc.

This grammar is not one that concerns hierarchical structure but one that creates linear structure by combining the sets of chunks sequentially in a certain order. So other species may have iterative Merge, but only humans obtained recursive Merge in the process of evolution.

Now this scenario works out a happy and reasonable compromise for the two conflicts mentioned in section 1. The speciation-related change from proto-language to human language is gradual in the sense that it can be characterized only by the emergence of recursive Merge, but it is saltational at the same time because the emergence of recursive Merge occurred only in humans. It may also be saltational in that the emergence of recursive Merge resulted from preadaptation or exaptation, as we will see in section 2.2. On the other hand, the speciation-related change from proto-language to human language is a product of both homoplasy and synapomorphy since human SM is analogous to proto-SM in songbirds and human CI and human Lexicon is homologous to proto-CI and proto-Lexicon in chimpanzees.

2.2 The Emergence of Recursive Merge as a Preadaptation

The next question to be addressed is how gradually recursive Merge emerged in humans, if it did not appear out of nothing, as well as why that was possible only for us. I assume, following Fujita (2009, 2014, 2016a,
b), that the emergence of recursive Merge resulted from preadaptation or exaptation, the transfer of a certain function in the same trait (behavior or organ). As a classic example, bird feathers had initially evolved for temperature regulation, but later were adapted for flight. In the case of language, its precursor (or its trait of the previous function) was what Greenfield (1991) called “Action Grammar”. Considering a developmental and evolutionary connection between action and language, Greenfield distinguishes three methods of object combination, as in the case of cup nesting given in (4a–c), and focuses especially on the crucial step from the pot-type (4b) to the subassembly-type (4c) of manipulations of objects. Analogous structures in language are given on the right, in which $X$ means the host or head of that combination as compared to a larger cup.

(4) Action Grammar and Language
   a. Pairing Strategy
   b. Pot Strategy (single target)
   c. Subassembly Strategy (multiple targets)

Both the pot and subassembly strategies may be said to be recursive operations in that the combined objects as a result of one nesting are carried on to become the input to the next nesting with another object. Then, both strategies can create hierarchical structures in the case of language, and the final results are the same in the case of cup nesting. However, in the pot strategy there is only a single target (host or head) that is combined with other objects, whereas in the subassembly strategy there are two targets (hosts or heads): a previous subunit (=subassembly) with one target is combined with another. In that sense, the subassembly strategy requires more complex chunking and more working memory than the pot strategy (see Aboitiz et al. 2006) for the connection between combination type and working memory). In fact, it is reported from a comparative cognitive experiment by Maynard Smith and Szathmáry (1995) that only the pot strategy is observed in chimpanzees while the subassembly strategy is used by human infants as young as 20 months old. So only the subassembly strategy is uniquely human.

Given the discussion so far, the whole picture of the parallelism between action and language can be summarized as in (5). This hypothesis proposed by Fujita (2009, 2014, 2016, a, b) shows that the uniquely-human subassembly strategy in action was somehow transferred by preadaptation or exaptation into sub-Merge in language, which is exactly the origins of human language in macroevolution. We call it here the hypothesis of the motor control origin of Merge (HOMCOM).

(5) The Hypothesis of the Motor Control Origin of Merge (HOMCOM)
We now understand from this perspective that the emergence of recursive Merge is actually not “saltational” but gradual again and that autapomorphy lies in the subassembly use of Merge in humans.

2.3 The Role of Phonology in Evolutionary Linguistics

From the discussion in sections 2.1 and 2.2, it follows that the shift from proto-SM to human SM is characterized as the emergence of Merge (as in (1)) and that the shift from proto-Merge (pairing/core Merge and pot-Merge) to human Merge is characterized as the emergence of sub-Merge (as in (5))

Then the next question will be how to test these hypotheses empirically and how phonologists can contribute to such empirical tests. In addition, since phonology is part of a grammar of externalization, or a set of operations for externalization in a particular language, the role of SM in the shift from proto-Merge to sub-Merge will also be at issue. So specific questions to be addressed in phonology are those as illustrated in (6).

(6) Research Questions to be Addressed in Phonology

a. Question in Biological Evolution

How did SM interact with, and have an influence on, the shift from proto-Merge (pairing/core Merge and pot-Merge) to sub-Merge?

b. Question in Developmental and Cultural Evolution

How can the HOMCOM be tested empirically from a viewpoint of phonology, given that a grammar is created by recursive Merge and the SM/CI interfaces?

Let me propose some ideas for answering these two significant questions. My proposal is to make explicit the role of phonology in evolutionary linguistics and establish specific methods for phonology to make empirical claims.

First, concerning question (6a), I propose that the shift from proto-Merge to sub-Merge was governed by “the Third Factor”, in the sense of Chomsky (2005), which made the emergence of Merge in macroevolution the way it was. The Third Factors are principles of the natural or physical law by which some traits (behaviors or organs) in animals are necessarily governed. Thus the Third Factors must be the principles in the SM system exactly because the traits of animals (especially their behaviors) are formed through the SM system. The gist of this proposal is illustrated in (7), where “Precursor” refers to some Merge-related traits in other species, including Action Grammar above, and “Merge” involves various forms of its operation in human language, such as core/iterative/recursive applications and pairing/pot/subassembly strategies.

(7) Third Factors in Macroevolution and Microevolution

What is crucial is that the Third Factors enter into the shift from Merge to the grammar of a particular language as well as the one from the precursor to Merge. It is precisely for this reason that we can have indirect access to hard issues in macroevolution (i.e., phylogeny) by investigating the grammars of individual languages directly as issues in microevolution (ontogeny=language acquisition and glossogeny=language change).

Now what exactly is “the Third Factor”? What is an example of a law of “the Third Factor”? One such case, I propose, is “Identity Avoidance” (Tanaka 2016a, b) or “Competitive Exclusion Law” (Fujita, Tanaka and Ikeuchi to appear). In the behavior of animals, complete competitors (e.g., large carnivores or alpha males) cannot coexist in a certain domain (i.e., niche or territory),
which is known as Gause’s Law or the competitive exclusion principle. This leads either to the extinction or demise of a competitor with less advantage (the competitive species in phylogeny or the competitive male in ontogeny), or to its evolutionary or behavioral shift toward a different ecological niche or territory. In the case of cup nesting like (4), the same size of cup cannot be paired by any means. So it is also a physical law.

In language, the law of Identity Avoidance works for the internalization of syntactic elements in the CI system for semantic interpretation, specifically for labeling by minimal search. As Chomsky (2008, p. 145) states, the algorithm for labeling proceeds as “in \{H, α\}, H an LI, H is the label” (H=head, LI=lexical item), meaning that the category of the head becomes the label of the whole merged pair. That is, a merged pair like \{*H, H\} is uninterpretable in the CI system and hence prohibited in principle (except for dvandva compounds) because it is unknown which head becomes the label. To put it simply, if an internalization statement \{α, α\} is generated, the derivation crashes at CI (because it is uninterpretable). The same is also true for the externalization or linearization of syntactic elements in the SM system. For example, the Distinctness Condition proposed by Richards (2010) states that if a linearization statement \{α, α\}' is generated, the derivation crashes at SM (because it is unpronounceable). In phonology, this condition is known as the Obligatory Contour Principle, which bans adjacent identical or similar sounds within a certain domain (McCarthy 1986, Odden 1986, Yip 1988). What two phonological elements are prohibited within what domain depends on the grammar of a particular language (for specific examples, see section 3.3). That is why the law of Identity Avoidance extends exactly to the shift from Merge to the grammar of a particular language as well as the one from the precursor to Merge, as shown in (7).

Another example of the Third Factor is the Implicational Law: a complicated trait, either behavior or organ, is derived from, and thus implies the existence of, the original simple counterpart of the same trait. This is obvious and universally true for evolution (phylogeny) and development (ontogeny). Bipedalism in humans implies quadrupedalism in other species, and the former in adults implies the latter in infants. Similarly, in the cup nesting mentioned above, the subassembly strategy implies the pot and pairing ones, and the pot strategy in turn implies the pairing one. This is true for humans and chimpanzees on the one hand and for adults and infants on the other.

In language, one example of this implicational relation is compounding. As Roeper and Snyder (2005) report, English allows both one-targeted and two-targeted operations of compounding while Swedish allows a one-targeted operation only.

(8) The Pot-Merge and Sub-Merge in Compounding
a. Swedish: only pot-merge

\[
\begin{array}{ccc}
\text{barn} & \text{buk} & \text{klub} \\
*\text{barn} & \text{buk} & \text{klub}
\end{array}
\]

b. English: both

\[
\begin{array}{ccc}
\text{child} & \text{book} & \text{club} \\
*\text{child} & \text{book} & \text{club}
\end{array}
\]

In other words, sub-Merge implies pot-Merge, but not vice versa. In this way, the Implicational Law applies to the shift from the precursor (=action Merge) to linguistic Merge, which results in Greenfield’s Action Grammar in cup nesting and so-called “implicational markedness” in compounding, respectively.

To sum up, the schema in (9) illustrates a proposal of how the Third Factors enter into the shift from action Merge to linguistic Merge. “Action Merge” here refers to some behavioral traits in other species.

(9) Third Factors in the Evolution of Merge

One thing to note is that the presence of sub-Merge in human language does not necessarily mean that the operation of Merge is always recursive and multiple-targeted; instead, it only means human language is unique in the sense that Merge is allowed to operate in a multiple-targeted fashion as well as in a single-targeted fashion. In other words, just as what identical sounds are prohibited in what domain depends on the grammar of a particular language in the case of the Obligatory Contour Principle, whether a phonological operation involves sub-Merge or pot-Merge depends on the grammar of a particular language. This leaves intriguing consequences in language acquisition and language typology, as given in (10).
(10) Developmental and Typological Consequence of the HOMCOM
a. A linguistic operation that involves sub-Merge in a certain language implies the presence of pot-Merge for the same operation; therefore, the use of the simple operation, pot-Merge, first develops and then the complex one, sub-Merge.
b. A linguistic operation that involves sub-Merge in a certain language implies the presence of pot-Merge for the same operation; however, the same operation that involves only pot-Merge in another language allows that use of Merge only.

Language acquisition concerns ontogeny, and language typology results from glossogeny, so we see again that the Third Factors operate on ontogeny and glossogeny as well as on phylogeny, as was emphasized in (7).

Given the discussion so far, we are now in a position to answer the questions raised in (6) at the beginning of this section and make explicit the role of phonology in evolutionary linguistics.

(11) Methods for Answering the Questions in (6)
a. Question in Biological Evolution
   → Uncover the Third Factors of the SM that was involved in the functional transfer from some precursor to Merge. They are natural or physical laws that hold for some trains (behaviors or organs) of other species and for some linguistic structures of humans.
b. Question in Developmental and Cultural Evolution
   → Uncover the phonological operations of the SM that leave developmental or typological consequences that are in agreement with the HOMCOM. They are operations that are reduced to Merge with the SM interface and involve combination of some three phonological elements.

In the next section, we will focus on the typology issue in (11b) and demonstrate that one such operation is the phonotactics of palatals in language.

3. Case Study in Typology

3.1 The Definition of Palatal Phonotactics

In pursuing an empirical test of the HOMCOM from the typology of palatal phonotactics, there are three points to be clarified in its procedure. First, the empirical test must involve the triplet CyV, a combination of the onset consonant, the palatal /y/, and the nucleus vowel, in which the palatal /y/ must be present underlyingly. That is, it may not be an emergent and transitional articulation as a result of palatalization like /CV/ → CyV. As Roeper (2011, p. 58) states, “Merge is a binary recursive operation that is invoked as soon as more than two words are combined. Therefore all languages with 3 word combinations are examples of recursion over two binary acts of Merge.” This point is clear from the compounding case in (8) and holds true for the empirical test that was mentioned in (11b).

(Note also that this statement of Roeper’s involuntarily admits that either pot-Merge or sub-Merge is recursive to the extent that they exhibit 3 word combination and create hierarchical structure as in (8), contra Fujita’s (2009, 2014, 2016a, 2016b) interpretation as mentioned around (4) and in note 2.)

Second, palatal phonotactics and so-called “palatalization” are two different things. Palatalization is a phonological process that changes the place of articulation (and sometimes the manner of articulation) of an onset consonant to the palatal region when combined with a high and/or front vowel. In this sense, palatalization typically occurs in the two-way combination or doublet C+V, as in kat-e ‘Win!’ / ka-fi ‘victory’ and kas-e ‘Lend!’ / ka-fi ‘loan’ in Japanese, and Egypt / ‘loan’ in English. In contrast, palatal phonotactics is a set of co-occurrence restrictions on the combination of C+y or y+V. I am claiming that if one uses this as an empirical test of the HOMCOM, it must be a three-way combination or triplet like C+y+V.

Third, as a consequence, it does not matter whether the phonological change involved is a case of secondary palatalization or full palatalization. That is, it is not relevant here whether the onset consonants of tulip, during, and super are [ty] or [ʃ], [dy] or [dʒ], and [sy] or [ʃ], respectively, in a dialect that allows palatalization. What matters is the fact that the onset consonant co-occurs with the palatal /y/ and undergoes secondary or full palatalization in this dialect.

Given these premises, the HOMCOM predicts there to be two types of languages that accommodate the underlying triplet C+y+V: pot-Merge type and sub-Merge type as illustrated in (12).
The Shape and Function of Phonology in Evolutionary Linguistics

(12) Pot-Merge and Sub-Merge in Palatal Phonotactics

a. pot-Merge type

\[
\begin{array}{c}
\text{C} \\
\text{2nd} \\
\rightarrow \text{y} \\
\text{1st} \rightarrow \text{V}
\end{array}
\]

b. sub-Merge type

\[
\begin{array}{c}
\text{C} \\
\text{2nd} \\
\leftarrow \text{y} \\
\text{1st} \rightarrow \text{V}
\end{array}
\]

Whether the palatal /y/ is first combined with the nucleus or the onset is tested by the presence or absence of a co-occurrence restriction applied to the combination. A co-occurrence restriction stems from the reason for motor control or perceptual distinctness in the SM system. So if there is a certain co-occurrence restriction applied to the pair, then it forms solid evidence showing that the two sounds are actually combined by Merge with the SM interface.

3.2 Empirical Tests of the HOMCOM

Now let us examine in detail how palatal phonotactics applies in Japanese and English. First, it is evident from the following minimal pairs that the palatal /y/ is underlying in Japanese.

(13) The Underlying Status of /y/

\begin{itemize}
  \item u ‘cormorant’ / yu ‘hot water’
  \item o ‘tail’ / yo ‘night’
  \item bou ‘certain person’ / byou ‘second’
  \item ta ‘rice field’ / tya ‘tea’
  \item sa ‘subtraction’ / sya ‘company’
  \item zou ‘elephant’ / zyou ‘high grade’
  \item en ‘empty’ / kyuu ‘sphere’
\end{itemize}

Words with the CyV structure are basically non-native since the native vocabulary allows /y/ to stand only in an exclusive onset like yu ‘hot water’ and yo ‘night’. Other words with the CyV structure here are Sino-Japanese. Crucially, a co-occurrence restriction on the yV sequence agrees with one on the CʴyV sequence, as is clear from (14) (I assume here, following Tanaka (2009), that the underlying form of [h] is /p/).

(14) Co-Occurrence Restrictions on CʴyV in Sino-Japanese

That is, /y/ can be followed only by back vowels but can be preceded by any consonant, which means that the combination concerned in Sino-Japanese is the pot-Merge type in (12a). Note also that the absence of sequences like *hyu (=/pyu) and *myu as well as *byu is due to another restriction banning *Labial+/y/ and the morpheme byuu in gobyuu ‘misunderstanding’ is used only for this idiomatic expression unlike the free morphemes in (14). Since the sequence of Labial+/y/ is acceptable with other back vowels, this restriction must apply to the whole combination of the labial onset and the previously-combined chunk /yu/, reinforcing the pot-Merge analysis (words with *wyV or *ywV are excluded by yet another restriction, which is irrelevant here).

The mimetic vocabulary also supports the pot-Merge analysis. As Mester and Ito (1989) argue, there is a co-occurrence restriction between the onset consonant and the palatal /y/. A morpheme with the palatal /y/ adds to the base various meanings of ‘uncontroledness’ such as childishness, instability, excessive energy, noisiness, and impoliteness, and clings to the rightmost coronal except for the rhotic /r/ or otherwise to the leftmost non-coronal.

(15) Co-Occurrence Restrictions on CʴyV in Mimetics

a. Search for a coronal from the right.

\begin{itemize}
  \item suko-suko / syuko-syuko ‘the way of stroking a slender bar’
  \item seko-seko / *syeko-syeko ‘the way of hurrying or flattering’
\end{itemize}
pata-pata / patya-patya ‘the sound of flapping or slapping’
pati-pati / *patyi-patyi3 ‘the sound of clapping hands’
b. If there is none, cling to the left consonant; if both are coronals, cling to the right.
poko-poko / pyoko-pyoko ‘the way of round things appearing or hopping’
peko-peko / *pyeko-pyeko ‘the way of bowing repeatedly’
nita-nita / nitya-nitya ‘the way of grinning hideously’
neti-neti / *netyi-netyi3 ‘the way of clinging to things tenaciously’
c. The rhotic /t/ is excluded though it is coronal.
koro-koro / kyoro-kyoro ‘the way of round things moving or rolling’
kero-kero / *kyero-kyero ‘the sound of frogs croaking’
turu-turu / tyuru-tyuru ‘the sound of sipping or slurping’
tiri-tiri / *tyiri-tyiri3 ‘the way of small flames burning things’

Crucially, this restriction applies to the whole combination C+yV and not to C+y, because, as in (15), it presupposes another restriction mentioned above that prevents front vowels from appearing after the palatal /y/. In fact, the palatal can only appear before back vowels even if there is no consonant in base-initial position, as shown in (16). That is, the restriction on yV applies first, followed by the one on C+yV.

(16) The Palatal /y/ with the Zero Onset
a. /y+/back vowels
ota-ota / yota-yota ‘the way of being flurried or staggering’
ororo / yoro-yoro ‘the way of being fluctuated or staggering’
ura-ura / yura-yura ‘the way of the air shimmering in daylight’
uruuru / yuru-yuru ‘the way of tears welling up in eyes’
b. /y+/front vowels
ira-ira / *yira-ira ‘the way of being irritated or impatient’
iki-iki / *yiki-yiki ‘the way of acting or smiling lively’
egu-egu / *yegu-yegu ‘the way of bursting into tears or weeping bitterly’
ero-ero / *yero-yero ‘the way of being obscene or indecent’

Interestingly enough, the form yota-yota (not *otya-otya) in (16a) shows that appending the palatal to the zero onset is favored over appending it to the coronal consonant, which enforces some revision of Ito and Mester’s generalization on mimetic palatalization.

In the same way, the foreign vocabulary, mainly borrowed from English, also supports the pot-Merge use in the C+yV combination.

(17) Co-Occurrence Restrictions on C+yV in Foreign

<table>
<thead>
<tr>
<th></th>
<th>φ</th>
<th>k / g</th>
<th>s / z</th>
<th>t / d</th>
</tr>
</thead>
<tbody>
<tr>
<td>ya</td>
<td>yangu</td>
<td>young’</td>
<td>kyampu ‘camp’</td>
<td>ayai ‘shy’</td>
</tr>
<tr>
<td>yaao</td>
<td>yarei</td>
<td>‘year’</td>
<td>gyanu ‘gang’</td>
<td>zyanu ‘genuine’</td>
</tr>
<tr>
<td>yi</td>
<td>*yi</td>
<td>*yii / *yii</td>
<td>*yi / yii</td>
<td>*yi / yii</td>
</tr>
<tr>
<td>yu</td>
<td>yauza</td>
<td>‘wee’</td>
<td>kyutu ‘cute’</td>
<td>ayuto ‘shh’</td>
</tr>
<tr>
<td>yuutopia</td>
<td>utopia</td>
<td></td>
<td>gyanutaubu ‘gustav (crocodile)’</td>
<td>zyuumu ‘jules (French name)’</td>
</tr>
<tr>
<td>ye</td>
<td>*ye</td>
<td>*yee / *yee</td>
<td>*yee / *yee</td>
<td>??yee / ??yee</td>
</tr>
<tr>
<td>yo</td>
<td>yotto</td>
<td>‘yacht’</td>
<td>kyuusii ‘gongsi (goblin)’</td>
<td>syokku ‘shock’</td>
</tr>
<tr>
<td>youtaruto</td>
<td>yogur’t</td>
<td></td>
<td>gyoumu ‘guillaume’</td>
<td>zyoa ‘joe (in French)’</td>
</tr>
</tbody>
</table>

One thing that arouses curiosity here is that only the sequence *yee is allowed freely as in tyekku ‘check’, *yee ‘chess’, and *yee ‘cherry’, even though /y/ is followed by a front vowel. However, other *yee sequences are prohibited or, as in ??yee, ??yee, and ??yee, tend to be avoided especially for speakers in the older generation;
and crucially, */y+e/ is vocalized and turned into [i+e], which means that the restriction against */ye/ is basically carried over to */C+ye (except for the idiosyncrasy of tye and the tolerable change in progress).

(18) Avoidance of */ye and */C+ye

yesu ‘yes’ → isu
syepaado ‘shepherd’ → sepaado
dyeri ‘jelly’ → zerii

Again, we witness here that the phonotactic restriction on yV stands on */C+ye/ as well. So we can safely conclude that the restriction in question holds true for all the lexical strata in Japanese.

The examples so far count solid arguments in favor of the pot-Merge use with some phonotactic restrictions in the SM interface. But the recent foreign vocabulary exhibits a somewhat different change in progress. The original bisyllabic Ci+u sequence undergoes glide formation of /i/ and then syllable fusion, leading to secondary or full palatalization, as illustrated in (19).

(19) Glide Formation and Syllable Fusion of Ci+u→Cyuu

a. glide formation and secondary palatalization
berusaiu ‘Versailles’: [i.u]→[y.uu]→[yuu]
purotoniumu ‘protonium’: [ni.u]→[ny.uu]→[nyuu]
erubiumu ‘erbium’: [bi.u]→[by.uu]→[byuu]
b. glide formation and full palatalization
ritimu ‘lithium’: [ti.u]→[ty.uu]→[tyuu]
karusiumu ‘calcium’: [si.u]→[sy.uu]→[syuu]
sinpoziumu ‘symposium’: [zi.u]→[zy.uu]→[zyuu]

These sequential phonological processes show a change in progress where the Cy+V chunk, a sub-Merge type of combination, is emerging in at least some part of the foreign vocabulary, although this is not a canonical use of Merge in Japanese.

All such words that undergo these processes are borrowed from English, which suggests an analysis of English CyV with sub-Merge, in which Cy are combined first and then Cy+V. This is true as argued in the following way.

But before discussing palatal phonotactics, let us examine labial phonotactics since /w/ and /y/ form the natural class glide and have direct relevance. First, the underlying status of /w/ is illustrated by the minimal pairs in (20a), and the labial /w/ can be followed by any vowel, lax or tense, as in (20b).

(20) Labial Phonotactics: Cw+V

a. the underlying status of /w/

win / in axe / wax eek / week eel / wheel
eight / eait ow / wow oak / woke alter / Walter
tin / tin dell / dwell soup / swoop coat / quote
b. no restriction between /w/ and a vowel

win wood wet won wag
weep woo wait woke water
c. restrictions on Cw: */Labial+w and */Sonorant+w

twin dweeb Dwanye Dwayne
thwack quit quote swoop swung

But it can only be preceded by a consonant that is not labial or sonorant. This shows that the restrictions of */Labial+w and */Sonorant+w are observed in the Cw combination in the onset position, which is then syllabified with the following vowel. The result is a sub-Merge type of organization of Cw+V.

In the same vein, palatal phonotactics also suggests this type of organization. The minimal pairs in (21a) show the underlying status of */y/ and the forms in (21b) guarantee the absence of any restriction between */y/ and the following vowel.
(21) Palatal Phonotactics: Cy+V
a. the underlying status of /y/

- ear / year: S / yes, up / yup: earn / yearn
- east / yeast: eight / yate: oak / yoke: ah / yah
- poor / pure: moot / mute: coot / cute: whom / Hume

b. no restriction between /y/ and a vowel

- yid: yuan: yet
- yeast: youth: yate
- poor / pure: moo: mut: coo: cute
- whom / Hume

c. restriction on Cy: *Coronal+y (in some dialect)

- c[y]ute: c[y]ure: c[y]ue

- ʃ: wan: tʃ
- dy: ʤ: sen
- sy: ʃ: mis

- (s)Cy rejects subsequent vowels other than /u/.

- It follows then that palatal phonotactics forms solid evidence for the sub-Merge analysis of the (s)Cy+V cluster in English.

However, there are some restrictions on the onset clusters Cy and sCy. In a certain dialect, especially General American, *Coronal+y and *s+Coronal/Dorsal+y are known to be prohibited as in (21c, d), and hence the palatal /y/ undergoes deletion, leading to the neutralization of contrasting pairs such as due / do and scute / scoot. What is crucial here is that, as in (21e), the resultant Cy and sCy clusters combine with only the vowel /u/ given the restriction against *(s)Cy+−u (the cluster (s)Cy rejects subsequent vowels other than /u/).

It follows then that palatal phonotactics forms solid evidence for the sub-Merge analysis of the (s)Cy+V cluster in English.

Now, if this is true, then the typological consequence of the HOMCOM in (10b) predicts that there should be some evidence for the pot-Merge use of the same operation. Postlexical palatalization in (22) forms such evidence; here, the lexical heterosyllabic C.yV sequence is resyllabified postlexically into tautosyllabic CyV.

(22) Pot-Merge Use of C+yV in Resyllabification

a. ty−ʤ: want you last year
b. dy−ʤ: send you would you
c. sy−ʤ: miss you kiss you

That is, it involves “left-capture” resyllabification of the coda consonant into the following onset position, leading to secondary or full palatalization. It is generally said in Lexical Phonology that some sound-related processes get grammaticalized (interacted with other components of grammar) in the course of historical change from phonetic ones in the postlexical stratum to phonological ones in the lexical stratum (see, for example, “Part III: Applying theory to historical change” in Hargus and Kaisse (eds.) 1993). Language change should be partly a process of elaboration in grammar. In that sense, it is quite natural that the primitive use of pot-Merge like (22) is retained at the phonetic, mechanical level and the derivative use of sub-Merge like (21c–e) is elaborated into the deep core of the phonological lexicon. That is, the sub-Merge use is a later development in phonology after elaborating the pot-Merge use in phonetics.

Finally, the arguments developed so far based on the specific uses of Merge in Japanese and English do not contradict with the common assumption of syllable structure and, crucially, with the putative evolution from the primitive proto-syllable CV to the derivative complex syllable CCV. For the pot-Merge use in Japanese, there are two possible views concerning the organization of syllable structure. One view is (23a), under which the onset /y/ adjoins to the nucleus /V/ followed by C adjunction to the onset /y/. An alternative view is (23a’), under which /y/ is an on-glide element adjoining to the nucleus /V/ and then the onset C adjoins to the whole nucleus. I will take the former view here, because there is no conceivable evidence showing that the yV sequence is a complex nucleus in Japanese. In general, whether /y/ is vocalic or consonantal relies solely on its position in the nucleus or elsewhere. It then follows that /y/ is a consonantal element in Japanese.
(23) Merge and Syllable Structure

a. pot-Merge

\[
\begin{array}{ccc}
\text{Onset} & \text{Nucleus} & \Rightarrow & \text{Onset} & \text{Nucleus} \\
\ y & \rightarrow & \ V & C & \Rightarrow & \ y & \rightarrow & \ V \\
\end{array}
\]

The structure in (23a) might at first appear to be a somewhat distinct internal organization from the one in (12a); however, if you think of C, /y/, and V as small, medium, and large cups, respectively, then you will easily see that the medium one is first put into the large one and then the small one into the medium one. This is exactly a pot strategy. Moreover, the evolution of syllable structure is considered as a shift from the primitive, unmarked proto-syllable CV to the derivative, marked syllable CCV. In fact, historically, the native vocabulary of Japanese originally had allowed only CV until the Asuka period, but through the influence of Buddhism, the vocabulary of Classic Chinese was introduced into the Japanese lexicon around the Nara period, which established CyV until the Muromachi period (see Hashimoto (1942), for example, for such a traditional view). Thus, the synchronic view in (23a) precisely concurs with the diachronic fact and better reflects cultural evolution than the one in (23a'). An onsetless syllable V with the branching nucleus is nothing but a very marked one for a proto-type primitive.

On the other hand, the sub-Merge use of English can be accounted for in a straightforward manner: it is based on the syllable organization in (23b), where the palatal /y/ first adjoins to the host C in the onset and then the complex onset adjoins to the host V.

From the discussion so far, we now clearly see the difference between Japanese CyV and English CyV in terms of Merge. After the process of externalization or linearization, the CyVs in Japanese and English might appear the same on the surface, specifically in their linear order and syllable structure: Cy and V form an onset and a nucleus, respectively, in both languages. However, from the perspective of palatal phonotactics in the SM system, the two CyVs are completely different in a deep and hidden way of their combination.

It is important that combination by way of Merge constitutes the vital source of typology in human language.

However, there is a further issue that remains to be examined: the internal organization of English CyV may be not uncontroversial since some researchers claim that it has a structure like (23a'). This claim amounts to saying that /y/ is a vocalic and not a consonantal element, leaving an undesirable consequence that shows the pot-Merge use in the palatal phonotactics of English. Let us turn to this topic in the next section.

3.3 Alternative View for English CyV

Contrary to the idea so far, there is a well-known counter proposal by Davis and Hammond (1995) that the CyV sequence of American English is actually /C+V/ in the underlying representation, which means that /y/ is underlyingly a semi-vowel that, together with the following vowel, belongs to the nucleus.

The main reason for such a proposal is to maintain the parallelism of co-occurrence restrictions between Cw and Cy, as summarized in (24). They happen to be all specific versions of the Obligatory Contour Principle mentioned in section 2.3. As seen in (20c), the Cw sequence is subject to the two restrictions of *Labial+Labial and *Sonorant+Sonorant, the latter of which incorporates a sub-restriction against *Glide+Glide. Similarly, if the same kinds of restrictions on homorganicity and sonority held true for the CyV sequence, a problem would arise: */my/ would not be prohibited by *Sonorant+Sonorant, as the examples in (24c) show.
This is why Davis and Hammond thought of CyV as /C+iV/ underlyingly, making /m+iu/ free from the restriction against *Sonorant+Sonorant holding in the complex onset. Then why do the on-glides in tune, due, and Sue undergo deletion? They account for this by assuming a rule that deletes an on-glide /i/ when it is tautosyllabic with a preceding coronal. That is, they are arguing that deletion is not motivated by *Coronal+Coronal in the complex onset but by another homorganic co-occurrence restriction holding over the onset and the nucleus.

However, there are some sets of counterevidence against this claim. First, under this view, there is an underlying assumption following Steriade (1988) that a co-occurrence restriction on homorganicity can hold between the onset and the nucleus whereas a co-occurrence restriction on sonority holds exclusively within the complex onset (Davis and Hammond 1995, pp. 160–161). This assumption is indispensable for accounting for the deletion facts in (21c) while accommodating the examples in (24c). Moreover, this assumption itself is considered an established insight in English phonology. But unfortunately, the systematic gap of *Glide+y sequences like *wy and *yy suggests that their analysis with this assumption would fail, because the required co-occurrence restriction should hold between the onset and the nucleus as in */w+iu/ and */y+iu/ even though it must be a sonority restriction. Instead, in order to maintain this valid assumption, *Glide+y must be applied to the complex onset, just like *Glide+w, under the more general version of *Glide+Glide. In addition, we claim that there is no such parallelism in restriction between Cw and Cy as suggested in (24). In fact, in varieties of English dialects, there are many which lack the restriction *Coronal+Coronal and accommodate /y/ after coronals. Likewise, *Sonorant+Sonorant does not hold for the CyV sequence, in order to accommodate /myV/ in general. So the only parallelism left between Cw and Cy is *Glide+Glide, which excludes *yw, *ww, wy, and *yy and holds for all varieties of English dialects.

Second, Davis and Hammond (1995, p. 174) neatly account for the presence or absence of deletion in (25a, b) based on stress-sensitive resyllabification. In each pair, the former member with stress on [iú] undergoes deletion as expected but the latter without any stress on [iu] does not. Since a stressed vowel attracts the following onset consonant to the tautosyllabic coda position by so-called “right capture”, deletion necessarily applies to the forms in (25a). But in the case of (25b), this right capture deprives [iu] of its onset coronal, blocking tautosyllabic deletion. That is why stressless CyV generally preserves /y/ even in American English.

(25) Stress-Sensitive Resyllabification

<table>
<thead>
<tr>
<th>Type</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. vo.[iú]m.i.nous</td>
<td>con.ti.n[iú].i.ty</td>
</tr>
<tr>
<td>b. vol.[iú]me</td>
<td>con.tin.[iú].e</td>
</tr>
<tr>
<td>c. áltr.u.is.tic</td>
<td>áltr.u.ism</td>
</tr>
<tr>
<td></td>
<td>álff.u.en.cy</td>
</tr>
</tbody>
</table>

However, the words in (25c) do undergo deletion, even though right capture would deprive [iu] of its onset coronals just like (25b). Davis and Hammond (1995, p. 174) attempt to evade this problem by stipulating that “these would lack /iú/ underlyingly.” which means that the /y/ of Cyu is exceptionally absent in the first place for these cases. But I claim that a more consistent, better account is to consider that deletion applies to (25c) because they are resyllabified as áltr.u.is.tic, áltr.u.ism, álff.u.en, and so on. What matters is actually the tautosyllabicity of the Cyu sequence, i.e., (25a) c. vs. (25b), and not whether it is stressed as in (25a) or

(24) Parallel Co-Occurrence Restrictions on Cw/y

<table>
<thead>
<tr>
<th>Type</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Cw+V</td>
<td></td>
</tr>
<tr>
<td>iii. *Glide+Glide in CC:</td>
<td>*yw, *ww</td>
</tr>
<tr>
<td>b. Cy+V</td>
<td></td>
</tr>
<tr>
<td>iii. *Glide+Glide in CC:</td>
<td>*wy, *yy</td>
</tr>
<tr>
<td>c. C+iV</td>
<td></td>
</tr>
<tr>
<td></td>
<td>music, mute, mule, muse, mural, mucus, etc.</td>
</tr>
</tbody>
</table>
stressless as in (25b, c). If so, CyV does not have to be /CiV/ underlyingly.

Third, if the CyV sequence was actually /CiV/, then co-called CiV-lengthening would apply to the words in (26a), but the fact does not support the prediction.

(26) CiV-Lengthening
a. Dř ác.[iů].la  cáp.[iů].la  per.pét.[iů].al
b. Lrán.ian  stúd.io  ré.méd.ial
c. lrán  stúd.y  rém.é.dy

This rule should lengthen an underlyingly short or lax vowel before the CiV sequence, just like the examples in (26b) as compared to the ones in (26c). Davis and Hammond are not concerned with the relevance to CiV-lengthening but may evade this problem by the rule of glide formation /CiV/ˠCyV, which, they assume, converts the original nucleus /i/ into the onset [y] (Davis and Hammond 1995, p. 168). This rule is ordered after stress-sensitive resyllabification and deletion, changing /i/ to [y] in the examples in (25b) and (26a). However, the rule should apply to the forms in (26b) as well, so they actually cannot evade this problem since the distinction between (26a) and (26b) is not captured by glide formation. Instead, if the underlying forms in (26a) have /CyV/ and the ones in (26b) have /CiV/ then the distinction is quite straightforward.

Fourth, coronal stops /t,d/ may undergo affrication when followed by approximants /r, w, y/, as in (27b), in a certain dialect of English. Full palatalization in (27biii-iv) is considered part of such affrication, which is ordered after postlexical “left-capture” resyllabification as is clear from (27biv).

(27) Affrication in the Complex Onset

Crucially, the domain of affrication is the complex onset, because the type of vowels following does not matter. If so, the internal structure of tune, deuce, and the like in (27biii) must be the one in (27a) in which the palatal /y/ belongs to the onset and not to the nucleus, just like the other approximants /r, w/.

Fifth, if the CyV sequence was actually /CiV/, then the /i/ would count as a mora and the whole syllable would be heavy so as to attract stress. English nouns and suffixed adjectives are known to undergo Latin Stress Rule, by which stress falls on the heavy penult (28a) or otherwise on the antepenult (28b). (Note here that stress-sensitive resyllabification naturally does not apply at this stage because its application is later after stress assignment. Syllabification is simply based on the Maximal Onset Principle.)

(28) Latin Stress Rule

However, as Davis and Hammond (1995, pp. 164–165) admit, this is not true for /CiU/ in (28c), which is treated as a light syllable by Latin Stress Rule, so they stipulate /CiU/ as monomoraic, unlike (28a), without any clear reason (although there are sporadic examples like Bermúda and Epicírus). But if it were monomoraic, the coherent moraicity of all rime elements including nuclei and codas in English would fail. The proposal of treating CyV as /CiV/ just to accommodate the myV sequence would have to sacrifice or miss this important generalization. Instead, if /CiU/ is monomoraic at all and the /i/ substantially does not count as a mora, then a simple analysis is to regard it as CyV. In fact, the forms in (28b) all contain the CwV sequence, and it is better to analyze it as CyV in order to maintain the parallelism in syllable weight with CwV.

Sixth, when the C of CyV happens to be absent as in udal, UFO, Ukrainian, unity, uranic, usual, utopia, uvula, and so on, the indefinite article that precedes them is its allomorph a and not an. This means that the initial syllable of these words is /yV/ and not /iV/, because it goes without saying that the indefinite article a is solely compatible with consonant-initial words. So there is no reason for assuming an underlying difference in /y/ between udal and feudal.

Finally, Davis and Hammond’s evidence for an
analysis with /CiV/ is also based on possible dialectal variations of a language game called Pig Latin. I will not discuss it in detail here, but an analysis of CyV with sub-Merge does not contradict with the facts if we assume a rule of vocalization /CyV/ → CiV, which reorganizes the original onset /y/ as the nucleus [i], the opposite process of Davis and Hammond’s glide formation mentioned earlier.

4. Concluding Remarks

We have so far discussed the shape and function of phonology under a new general scenario in evolutionary linguistics and have proposed some specific research questions, hypotheses, and methods for unveiling language origins by observing particular languages at hand. To retail the plot of the story I have developed, we have taken the following course.

(29) The Roadmap we have Followed

a. The general scenario of the evolution from proto-language to human language
   → The Merge-only hypothesis in (1)

b. The emergence of human-specific Merge by pre-adaptation or exaptation
   → The hypothesis of the motor control origin (HOMCOM) in (5)

c. The reason that (1) and (5) in macroevolution can be tested empirically
   → The Third Factors in SM governing the shift of precursor → Merge → grammar in (7) and (9)

d. The specific hypotheses to be tested in phonology
   → Consequences of the HOMCOM in acquisition and typology in (10)

e. The specific methods for testing the hypotheses (10) in phonology
   → Exploring the Third Factors and/or phonological operations combining three elements in acquisition or typology in (11)

f. The specific cases for such an operation matching up to the HOMCOM in typology
   → The pot-Merge and sub-Merge uses of Japanese and English in palatal phonotactics in (12) and (23)

This is a summary of the key ideas in the present article. However, there remain some issues to be explored, according to the statement in (11). First, we have to seek more Third Factors other than Identity Avoidance and Implicational Law and adduce evidence that they are working in precursor, Merge, and grammar. Second, we have to seek more phonological operations other than palatal phonotactics to test the HOMCOM empirically in acquisition and/or typology. Finally, even for palatal phonotactics, we have to look for languages other than Japanese and English with the three-way combination of underlying C+y+V and test the implicational relation in the two subtypes of Merge. The door is now open for phonetics and phonology to contribute to evolutionary linguistics. Let us concentrate all our efforts on the issue of language origins by integrating multidisciplinary knowledge in science.

Acknowledgements

This research is supported by a Grant-in-Aid for Basic Science (B) of the Japan Society for the Promotion of Science whose specific research topic is “Internal and External Evidence for Phonological Theory: A Transdisciplinary Approach” (subject number: 16H03427) and by a MEXT Grant-in-Aid for Scientific Research on Innovative Areas #4903 (Studies of Language Evolution for Co-creative Human Communication) whose specific branch is “Theoretical Frameworks for Studying the Origins and Evolution of Language” (subject number: 17H06379). Part of the contents of this article was publicized at an invited talk in the colloquium organized by the department of linguistics at Simon Fraser University, BC, Canada held in March 2, 2017, and in a research meeting of the above mentioned project held in March 19, 2017 at Iwaki Meisei University, Fukushima. I would like to express my deepest gratitude for valuable comments and feedback from the audience, especially to John Alderete, Nancy Hedberg, Keir Moulton, Toshiyuki Tabata, Takeru Honma, and Kohei Nishimura. I’m also grateful to Queenie Chan for stylistic improvement and to the two anonymous reviewers for useful suggestions in content and style. Any errors left are the sole responsibility of the author.

Notes

1) As for the relevance to the terms in (2), proto-Merge involves an iterative application of core Merge (2c) as well, while sub-Merge is equivalent to a recursive application of core Merge (2a), or what we have called recursive Merge so far. This is basically an idea by Fujita (2009, 2014, 2016a, 2016b).

2) But the fact that human Merge is characterized by sub-Merge does not necessarily mean that it consists only of sub-Merge: in fact, it contains core-Merge and pot-Merge
as well. Thus, there is a possibility that recursive Merge can be pot-Merge as well as sub-Merge by definition, and I take this position here, as seen in the discussion below (4), which means a disparity from Fujita’s interpretation of ‘recursion’. Fujita considers only the subassembly strategy recursive and hierarchical, the pot one iterative and sequential. But I claim that both are recursive and hierarchical, the only difference being whether it is single-targeted or multiple-targeted. The reasons for taking this interpretation are three-fold. First, in the pot strategy, the output of one nesting as in (4bi) is an input of the next nesting, which combines it with as in (4bii), and this process is no different from the definition of recursion. In fact, Fujita (2016a) illustrates the pot strategy with hierarchical structure like the one in (4b). Second, the final results in (4b) and (4c) are the same nesting, hierarchical structure . The difference is the number of targeted cups in the process of nesting. Third, even chimpanzees may get to know the subassembly strategy when they are captive-bred with intensive linguistic training (cf. Matsuzawa 1991, Hayashi 2007). This suggests that they may have the potential to use multiple-targeted recursion and that the potential is hidden in single-targeted recursion.

3) The contrast in (un)grammaticality of pati-pati³/patyi-patyi, neti-neti²/netyi-netyi, and tiri-tiri/tiryi-tiryi in (15a–c) may be confusing. However, given the independently-motivated allomorphic rule that palatalizes coronals before high vowels (see the discussion at the beginning of section 3.1), it is evident that only the former examples are grammatical and undergo this rule, surfacing as pa[ʧ]i-pa[ʧ]i, ne[ʧ][i]-ne[ʧ]i, and [ʧ][iri]-[ʧ][iri], respectively. See also (16b) for the incompatibility of the /y/ sequence.

4) But again, sub-Merge implies the presence of pot-Merge here, and as expected, there is some more evidence for the pot-Merge use of the C+yV sequence in foreign words. Note that the myu sequence is strictly prohibited in Sino-Japanese and mimetics, which means there was originally no vocabulary with this sequence in Japanese. Although some words borrowed later may have this sequence as in myuu.zik.ku ‘music’ and myuu.to ‘mute’, the restriction of myu somewhat happens to work for the foreign vocabulary (when myu is a high vowel /i/) and causes metathesis to avoid the difficulty of its pronunciation: si.myu.relu,syon → su.mi.ree.syon ‘simulation’ and ko.myu.mi.kee,syon → ko.mi.nyu.kee.syon ‘communication’. In this metathesis, yu and i are substituted for each other, which clearly shows that the m+yu sequence is a pot-Merge use of combination.

5) The point here is that the deep and hidden combination of the C+y+V is distinct from their surface linear order or syllable structure. As an anonymous reviewer points out, Scheer (2004) and Samuels (2009) assume that the structure of C+y+V is not hierarchical but just linearly represented. But I am claiming that the discussion above shows evidence for their deep and hidden hierarchy, independently of their linearization or syllabification. Also, according to the reviewer, in Government Phonology, C and V of CyV are first merged and then /y/ is added to the right-hand side of C of the already-constructed CV; alternatively, C and /y/ are first merged and then the chunk and V are concatenated to form CyV. Thus, the construction process for Japanese /y/ (first y+V and then C+yV) is not employed in this framework, because /y/ of Cy cannot be the head of the onset constituent (cf. Harris 1994). This view does not matter here since we are assuming that combination by way of Merge and syllabification in a specific theory are different things. But I am grateful to the reviewer for this valuable suggestion: different views from mine may exist as to the hierarchy and syllable structure of CyV. However, a different view does matter in some cases, which will be discussed in section 3.3.

6) There may be another difference between the two languages as to whether CyV involves full or secondary palatalization (i.e., [ty]ube or [ʧ]aibu). This is also a matter of the SM system, but we ignore it here.

7) As for the absence of [y] in a word-initial stressless syllable that begins with a coronal as in neut.roc.ic, tu.tör.ial, tu.milt.u.ous, Teu.tón.ic, etc., Davis and Hammond (1995, p.172, p.174) just say again that “these would lack /iu/ underlyingly.” However, I claim that this can also be accounted for by the tautosyllabicity of the CyV sequence, just in the same way as their corresponding source words like nér.ón, tít.or, tu.múlt.últ, Tétu.on, etc. Here again, the presence or absence of stress on the relevant syllable is irrelevant.

8) This phenomenon can also be analyzed as sonorant devoicing, a type of voicing assimilation by which sonorants lose their voicing after voiceless obstruents, as in plan, crow, slip, shrimp, sneeze, fleet, and so on (Iversen and Salmons 1995). In this analysis, /i/ in try, for example, does not undergo affrication but /r/ simply becomes voiceless: [ʧr]y and not [ʧt]y. In fact, Cruttenden’s (2014: sections 9.2.6 and 9.7.2) description amounts to saying that the word initial consonants in try and chay are phonetically different. However, it is quite natural that their phonetic details are different because of their complex/simplex status in the onset. Thus, they are substantially not comparable on the same conditions, because lexical /ʃ/ cannot be followed by /e/ in English and there is no word like *chray. In addition, as long as the same change also occurs after voiced obstruents in dry, dwell, and deuce as shown in (27b), it should be better to regard it as affrication. More importantly, the phenomenon in (27)
occurs in the complex onset, irrespective of whether it is affrication or sonorant devoicing from a phonetic point of view. There is no doubt that this point shows evidence for the constituency of a voiceless obstructant and a sonorant within the onset, but special thanks go to the reviewer for bringing the possibility of sonorant devoicing to my mind.

References


(Received Feb. 22, 2017, Accepted May. 18, 2017)