On Lenneberg Conjecture: Syntax as Calculus of F

ARIKAWA Koji

Abstract

Lenneberg conjectured that syntax is the calculus of functional categories \([F]\). What insight can we gain from a “soft-mathematical” syntax? The Lenneberg Conjecture (LC) predicts that syntax is the calculus of \(F\) with no set parameter, i.e., \(2^x=1\) genotype (the initial state \(S_o\) of CHL yielding “Homosapienses”). The Borer–Chomsky Conjecture (BCC) predicts that \(F\) is parametrized, i.e., \(2^x=4996\) phenotypes. BCC and LC are connected by symmetrical exponential function \(y=e^x\). \(y=2^x\) here, where growth rate gets close to the formed structure. Feature checking controls derivative \(\frac{dy}{dx}\) (growing speed), looking into an infinitely small structure \(\Delta y\) at infinitesimal time (step) \(\Delta x\).

1. Introduction: Lenneberg Conjecture (LC) and Borer-Chomsky Conjecture (BCC)

In this paper, we use Lenneberg’s insight, reproduced below, as our starting point.

(1) “Syntax is the calculus, so to speak, of functional categories, and the categories are arranged hierarchically from the all-inclusive to the particular” (Lenneberg 1967: 292).

Functional categories \(F\) include \(v, v^*, T, C,\) and \(D\). What is calculus? “Calculus is all about growth rates” (Strang 2010a: 2). “Calculus is about pairs of functions” (ibid. 3). A car is a good example, as it has many pairs of functions, e.g., the speedometer and trip meter.\(^1\) The speedometer is a derivative (i.e., speed or growth rate at an instant)\(^2\) and tells us what our speed or growth rate

---

1) Refer to Strang (2010a: vii) for examples of pairs of functions.

2) This is the first derivative \(\frac{dy}{dx}\) (read as [di: wat di: eks], not [di: wat ouv di: eks]), which is the speed (growth rate) at an instant in time. \(dy\) stands for an infinitesimal distance, whose limit is 0 (it infinitely approaches 0), i.e., \(dy \to 0\). \(dx\) is an instant, the limit of which is 0 (it infinitely approaches 0), i.e., \(dx \to 0\). Note that a derivative \(\frac{dy}{dx}\) is not a division. The second derivative \(\frac{d^2y}{dx^2}\) (read as [di: sek tand wal di: eks skweard]) is acceleration, describing whether the growth is speeding up or slowing down, i.e., the rate of changing speed (Strang 2010b).

Keywords: Borer–Chomsky Conjecture, calculus, feature checking, Homosapienses, Lenneberg Conjecture
is at a particular instant in time, whereas the trip meter is an integral or total mileage. In other words, the trip meter tells us how the total mileage (i.e., sentential structural distance) grows at each instantaneous point of time. “What calculus finds is the speed at each separate moment—the whole history of speed from the whole history of distance. Your car has a speedometer to tell the derivative. It has a trip meter to tell the total mileage. They have the same information, recorded in different ways. From a record of the speeds we could recover a lost trip meter and vice versa. *One black box is enough, we could recover the other one:* The derivative (speedometer) tells how the distance is changing. The integral (odometer) adds up the changes to find the distance. This is the ‘Big Picture’...” (Strang 2010a: vi–vii). Here, the keywords of calculus include “instant,” “growth rate,” “continuous,” “changing,” and “recover (inverse).”

Next, we ask what Lenneberg’s intuition was and what we can learn from it. If the computational system (i.e., procedures) of human natural language (CHL) is a car, the speedometer is a derivative that indicates the growth rate at every instant of structure building, while the trip meter is an integral that shows us the total architecture of a sentence structure at some instant of time. Does syntax calculate growth rates of sentence trees? What is the derivative (i.e., the growth rate or speed at an instant of time) of the sentence-structure building? What is the integral (i.e., the total structure at each instant) of the structure growing process? Lenneberg’s intuition can be qualified as the conjecture below.

(2) LC: Syntax is the calculus of $F$ (Lenneberg 1967: 292).

LC is concerned with the initial state $S_0$ of CHL and the principle of minimal computation (MC). Call the single natural language of Homo sapiens as “Homosapienses.” At this stage, $F$ is similar to a stem cell hiding all potential bifurcations. No parameter is set. LC deals with the genotype of CHL. In contrast, BCC accounts for language variation.

(3) BCC: $F$ is parametrized (e.g., Chomsky 1981, Borer 1984, Baker 2008).3)

3) Chomsky (1981: 27) suggested that inflectional head (INFL) causes a parametric distinction between the obligatory and optional presence of a subject. For example, English and French contain base rule $S \rightarrow NP INFL VP$, whereas Semitic languages such as Hebrew and Arabic contain the base rule $S \rightarrow (NP) INFL VP$. Naming optional subject (NP) in the latter as “pro,” i.e., the unpronounced pronoun, the pro parameter is ON for the latter type of languages and OFF for the former. When the pro parameter is OFF in a language, the language uses expletives such as “it,” “there,” and “il.” Borer (1984: 29) extends Chomsky’s suggestion and proposes “a system that reduces all interlanguage variations to the properties of the inflectional system.” The functional categories such as $v$, $v^*$, $T$, $C$, and $D$ build the inflectional system.
BCC is concerned with language phenotypes (the steady state $S_0$ of CHL). Assume that the parameter is binary. If $\text{CHL}$ contains 12 parameters, then the variation is $2^{12}=4096$, which is approximately the closest number of phenotypes at this point. Therefore, an idealized number of particular languages or dialects of Homosapienses is 4096 at an appropriate level of abstraction. If $\text{CHL}$ contains 11 switches, then $y=2^{11}=2048$ languages, which is too small an approximation; if $\text{CHL}$ has 13 switches, then $y=2^{13}=8192$, which is too large an approximation. Given the above, an equation for BCC is exponential function $y=2^x$; however, a much more difficult problem is what the switches actually are and why. See section 1.3.

Regarding Lenneberg’s statement that “the categories are arranged hierarchically from the all-inclusive to the particular,” we translate it as $\text{CHL}$ has properties of discrete infinity and yields fractal structures, which leads us to the Fibonacci sequence.

Jenkins (2000) encourages us to make the most of “soft mathematics” in the sense of Devlin (1996), “unrigorous mathematics” in the sense of Fourier, or “dirty mathematics” in the sense of Heisenberg (Crease and Mann 1987: 428) to “gain new insight” into the study of CHL. More specifically, Jenkins provides us with the following constructive viewpoint.

(4) “In the early stage of the study of any scientific discipline, whether physics or the study of mind, we try whatever works, whether that is ‘soft mathematics’ in Devlin’s sense, to gain new insight, unrigorous mathematics à la Fourier, ‘dirty mathematics’ in Heisenberg’s sense, or ‘hard mathematics.’ We know that physics moved through all of these stages, in one area or another throughout its history. We also know that unification proceeded slowly in some areas, more rapidly in others and sometimes piecemeal” (Jenkins 2000: 49).

From this brief introduction, we organize the reminder of our paper as follows. In Section 2, we sketch out changes in the human brain; here, if calculus is about change, it should be able to describe brain development. Next, in Section 3, we introduce an exponential function where LC and BCC are naturally placed. In section 4, we propose that feature checking controls a derivative (growing speed), looking into an infinitesimal structure and time. In Section 5, we conclude our paper.

---

4) An equation such as this can yield only an idealized number here and not the exact actual number. As a related example, the formula for population growth is $P=P_0e^{rt}$, where $P$ represents the final population, $P_0$ represents the initial population, $e$ is the natural logarithm (a mathematical constant approximately equal to 2.71828), $r$ represents the growth rate, and $t$ represents time. This formula is used to predict an idealized final population, not the actual population which fluctuates according to unpredictable events. Refer to Algebra → Exponentials and Logarithms → Population Growth at Coolmath.com (2017) for a good introduction.
2. CIL development as a change in growing speed of the brain

The key phrase that Lenneberg uses is “syntax development,” which directly relates to his insightful intuition below that syntax is the calculus of F.

(5) “In the absence of systematic research on children’s understanding of adult sentences, and hence of their developing ‘analytic equipment’ for syntax, we can only make educated guesses at how grammar actually develops. The study of adult syntax makes it clear that discourse could not be understood, and that no interpretable utterances could be produced, without syntax development pari passu with lexical and phonological development” (Lenneberg 1967: 292).

Further, Polya encourages us to draw pictures, although we initially have no idea as to what these pictures may signal and how they interact.

(6) “Draw a figure” (Polya 1945). “Figures are not only the object of geometric problems but also an important help for all sorts of problems in which there is nothing geometric at the outset” (ibid: 103).

Given the above, in Figure 1, we draw approximate graphs of changes known about human brain development.

![Figure 1: Change in growing speed of human brain development](image-url)
The number of neurons in the cerebrum peaks at 33 weeks after fertilization (i.e., one month before birth) at approximately 30 billion neurons, and then undergoes drastic apoptosis (i.e., genetically programmed cell death) to bring the number of neurons down to 14 billion (Tominaga and Mogi 2006: 140). The number of neurons in the cerebrum cortex peaks at 17 weeks after fertilization (i.e., six months before birth) at 14 billion neurons, and then undergoes slow degeneration (Mizutani 2006: 48). The synapse density in the cerebrum peaks at eight months old at six synapses per cubic millimeter (ibid. 81). The number of neurons and synapse density in the cerebrum show lognormal distribution: “a continuous distribution in which the logarithm of a variable has a normal distribution” (Weisstein 2017).

The apparent development of natural language showing the “Basic Property of human language” seems to spurt at around three years old. Likewise, the development of m-LAD is fully working at three years old. The m-LAD becomes inactivated at around ten years old when secondary sexual characteristics become dominant. The dramatic apoptosis of neurons and synapses precedes the apparent spurt in the emergence of language. The m-LAD and secondary sexual characteristics seem to be a tradeoff in development. We, therefore, speculate that Lenneberg had an intuition that the process by which these interacting graphs change can be investigated using the calculus and that F is responsible for this process of change.

3. Exponential function \( y = 2^x \) produces language phenotypes

The laws of nature appear to respect the exponential function \( y = e^x \), a function that calculus created (Strang 2010c). A mathematical constant \( e \) here is named after Euler, also called Napier’s number, and is the base of the natural logarithm (\( \ln \)) (approximately 2.71828) defined as \( e = \lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n \). In this paper, we focus on \( y = 2^x \) because the binary ON/OFF-parameter setting of switch box connected to CHL-operating system (OS) yields the phenotypes. The corresponding graph is shown in Figure 2.

---

5) The “Basic Property of human language” is a natural phenomenon as follows. “[E]ach language yields a digitally infinite array of hierarchically structured expressions with systematic interpretations at interfaces with two other CHL-external systems, the sensorimotor system for externalization and the conceptual system for inference, interpretation, planning, organization of action, and other elements of what is informally called “thought” (Berwick and Chomsky 2016: 89–90).

6) Leonhard Euler (1707–1783) was a Swiss mathematician, who found \( e \) in a search for a derivative (growing speed) of logarithmic function \( y = \log x \) (Horiba 1991: 91). John Napier (1550–1617) was a Scottish mathematician, who first proposed the idea of logarithm (Horiba 1991: 42). \( \lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n \) means \( \left(1 + \frac{1}{n}\right)^n \) with \( n = 1, 2, 3, \ldots \) to infinity \( \infty \). An alternative definition is \( e = \sum_{n=0}^{\infty} \frac{1}{n!} = \frac{1}{1} + \frac{1}{1} + \frac{1}{1 \cdot 2} + \frac{1}{1 \cdot 2 \cdot 3} + \ldots \).
If we start with /TypeOne0 \( /TypeOne0 \) at initial state /TypeOne1 \( /TypeOne0 \) of CHL in which parameters are absent or unset, we find that the number of languages is exactly one, i.e., there is a single natural language of Homo sapiens, which, as noted above, is called “Homosapienses.” The mirror image of the exponential function (i.e., vertical axis) is the logarithmic function (i.e., horizontal axis). Given \( y=2^x \), the mirror image is \( x = \log_2 y \). Equation \( y=2^x=1 \) expresses that no parameter setting yields one language: Homosapienses. Thus, exponential equation \( y=2^x \) predicts that CHL yields Homosapienses without parameter setting. No set parameter is expressed as \( 0 = \log_2 1 \). Equation \( y=2^{12}=4096 \) expresses that setting of 12 parameters yields 4096 languages. The number of switches is expressed as \( 12 = \log_2 4096 \). If the number of phenotypes is 7000, then \( 2^{13}=8192 \) is not too large. Calculating the exact number of phenotypes is impossible here since many languages continuously undergo synchronic and diachronic change and become endangered languages.

The most important property of the exponential function \( y=e^x \) is that the function equals the growth rate. Equation \( y=e^x \) expresses “no change,” i.e., it realizes invariable and symmetry (Horiba 1991: 7). An instant point contains the whole. \( y=e^x \) hides fractal property and realizes perfect symmetry. In other words, when a system grows, the growth rate speeds up.

---

7) The derivative (growth rate) of exponential function \( y=e^x \) is \( \frac{dy}{dx} = e^x \). The integral of \( y=e^x \) is \( \int y \, dx = \int e^x \, dx = e^x + C \), where \( C \) is integral constant (Horiba 1991: 7).
(7) The exponential function $y = e^x$ is the solution of $\frac{dy}{dx} = y$ that starts from $y = 1$ at $x = 0$.

The exponential function equals its slope (i.e., the growth rate or growing speed) (ibid). Slope of $y = e^x$ is $\frac{dy}{dx} = e^x$. Slope of $y = a^x$ is $\frac{dy}{dx} = (\ln a) a^x$, where $\ln a$ stands for natural logarithm of $a$, i.e., the power of $e$ that produces $a$: $e^{\ln a} = a$. In our case, the growth rate $\frac{dy}{dx}$ of $y(x) = 2^x$ at $x = 12$ is $(\ln 2)2^{12} \approx 0.693 \times 4096 = 2839$. When we observe language variations on the order of 1000, we are simultaneously looking at a derivative or growth rate 2839, which equals approximately 70% of $2^{12} = 4096$. The fact that $\text{Cth}$ (i.e., a natural object) exhibits the same variations obeying the exponential function is consistent with the fact that the exponential function is frequently observed in life sciences. In other words, the exponential function is a life function.

(8) “The laws of nature are expressed by differential equations, and at the center is $e^x$. Its applications are to life sciences and physical sciences and economics and engineering (and more—wherever change is influenced by the present state)” (Strang 2010a: 283).

The unchanging exponential function does a good job at representing human language variation.

---

8) Strang (2010a: 15). The zero power of any positive number is 1. To illustrate this, note that $2^0$ instructs us to multiply 1 by 2 zero times, i.e., do not multiply 1 by 2 at all, which leaves us with 1. Next, $2^1$ instructs us to multiply 1 by 2 exactly once, which yields $1 \times 2 = 2$. Continuing this pattern, $2^2$ instructs us to multiply 1 by 2 exactly twice, which yields $1 \times 2 \times 2 = 4$, and so on. When $y = e^x$ solves $\frac{dy}{dx} = y$, all other functions $Ce^x$ solve it too, because $Cy = Ce^x$ and constant $C$ on both sides of the equation cancel each other out (ibid).

9) The key is to connect $2^x$ with $e^x$. Let $\ln 2$ (natural logarithm of 2) be the power of $e$ that yields 2. Then $e^{\ln 2} = 2$. $\ln 2$ is about 0.693. Express $2^x$ as $e^{x\ln 2}$. Slope (derivative) of $y = 2^x$ is $\frac{dy}{dx} = \frac{d}{dx} 2^x = \frac{d}{dx} e^{x\ln 2}$.

“$2^x$ also grows exponentially, but not as fast as $e^x$ (because 2 is smaller than $e$). Probably $y = 2^x$ could have the same graph as $e^x$, if I stretched $x$ axis. That stretching multiplies the slope by the constant factor $\ln 2$” (Strang 2010a: 19). Slope of $y = 2^x$ is $(\ln 2)e^{x\ln 2} = (\ln 2)2^x$.

10) According to Stewart (2011: 64), Malthus (1826) “asserted that populations of living creatures, if their growth is not restrained by lack of food or predation, grow ‘geometrically’ the population size at successive instants of time is multiplied by the same fixed amount. ... The numbers grow very rapidly—the modern term is ‘exponentially’.” Discussing “the role of mathematics in unification,” Jenkins (2000: 45) cites Davis and Hersh (1981: 199), who points out that the “exponential emerges as trigonometry in disguise, and vice versa,” i.e., $e^x = \cos x + i \sin x$, where $i = \sqrt{-1}$ (i.e., Euler’s formula). When $x = \pi$, we obtain $e^\pi + 1 = 0$, where imaginary number $i$ connects four independently needed numbers, i.e., the most basic natural number 1, made-in-India zero, the ratio of the circumference of a circle to its diameter $\pi \approx 3.14\ldots$, and made-in-calculus natural-logarithm base $e = 2.71828\ldots$. Richard Feynman called it “our jewel.” Exponential and trigonometry appear unrelated in real-number world but they are tightly connected in imaginary-number world (Mizutani 2015: 60–63). An equilibrium (i.e., symmetrical; exponential) state is oscillating (i.e., trigonometry). The Cth-growth equation $y = 2^x$ hides oscillation.
Suppose that CHL contains $x=12$ parameters. Linear function $y(x)=2x$ yields output $y(12)=24$ languages whereas squaring function $y(x)=x^2$ yields output $y(12)=144$ languages and exponential function $y(x)=2^x$ yields output $y(12)=4096$ languages. Since language variations are expressed by an order of 1000, exponential function $y(x)=2^x$ is a good candidate for the formula describing the CHL parameter.

Given that brain computation is conveyed electrically (i.e., via digital ON or OFF values) as well as chemically with analog threshold values, our conclusion supports the principles and parameters approach to CHL, where these parameters are set to binary values. The idealized formula of language variation is therefore $y(x)=2^x$ with $x=12$, i.e., $y(12)=4096$.\(^{11}\) As noted above, a much more difficult problem is what these parameters are and why.

4. Feature checking as derivative and structural growth as integral

Lenneberg speculates that the first stage of m-LAD has the rule as follows.

\[(9) \quad S \rightarrow W\]

Here, “sentence $[S]$ is formed by the use of any word that belongs to the class $W$, and all of the child’s words do belong to it” (Lenneberg 1967: 292–293). The second stage, as illustrated in Figure 3, develops a binary structure that rotates in three dimensions (adapted from Braine 1963; Lenneberg 1967: 293).

![Figure 3: Illustrating the second stage (i.e., two-word stage) of m-LAD](image)

Further, “[t]he entire utterance seems to turn around them” (ibid.), which represents the emergence of Merge.

\(^{11}\) Assume that CHL exhibits approximately 5000 language variations. Suppose that the language variation equation were $y(x)=2x$. Since $5000=2\times2500$, CHL needs 2500 parameters to yield 5000 language variations. Suppose instead that the language variation equation were $y(x)=x^2$. Since $5041=71^2$, CHL needs 71 parameters to yield 5041 language variations. It remains unclear whether CHL is equipped with 2500 or 71 parameters. The computational cost used in m-LAD must be free because the acquisition is unconscious, quick, and effortless. If so, the number of parameters must be small to yield the language variations on the order of 1000. Here, 12 parameters are simple enough to yield approximately 4000 variations.
Here, Merge is “an operation that takes objects X and Y already constructed and forms a new object Z. The third factor principle of minimal computation dictates that neither X nor Y is modified by Merge (the ‘No Tampering Condition’), and that they appear in Z unordered” (ibid).

Next, as illustrated in Figure 4, the third stage differentiates category $W$ into $m$ and $N$ (adapted from Braine 1963; Lenneberg 1967: 293).

Lenneberg described the essence of language acquisition as “progressive differentiation of syntactic categories” (ibid. 294). His insight is connected with the hypothesis that lexical categories such as N (noun), V (verb), A (adjective/adverb), and P (pre/postposition) are differentiated into complex of lexical features consisting of $[\pm N]$ and $[\pm V]$ (Chomsky and Lasnik 1993), which we illustrate in Table 1.

<table>
<thead>
<tr>
<th></th>
<th>$-N$</th>
<th>$+N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-V$</td>
<td>P</td>
<td>N</td>
</tr>
<tr>
<td>$+V$</td>
<td>V</td>
<td>A</td>
</tr>
</tbody>
</table>

Table 1: Lexical categories formed by complex lexical features

We speculate that lexical categories N and V are differentiated into functional categories such as D, v, v*, T, and C.

Further, Merge is differentiated into External Merge (EM) and Internal Merge (IM). An example of IM is as follows. If Y is part of X, then “the result of Merge is again $|X, Y|$, but in this case with two copies of Y, one the original one remaining in X, the other the copy merged with X” (Chomsky 2013: 40).

Lenneberg’s “progressive differentiation” leads us to a differential calculus of syntactic categories. Lenneberg pointed out that differentiation is a key to any growth system, i.e., “This differential process is not confined to language. In fact, it is the hallmark of all development”
The essence of the idea of the derivative is the process of differentiation. A curve comprises a set of points $P$ that represents the relation between a set of independent variables and a set of dependent variables. For simplicity, let us assume that $P$ relates the single independent time variable $x$ to a single dependent variable $y$ (Figure 5). Given a fixed point $A$ on the curve $P$, the idea is to find the distance and direction, represented by the vector $v$, between $A$ and a neighboring point $B$ as that point $B$ becomes infinitesimally close to $A$. The vector $v$ represents the growth rate of the curve $P$ at the point $A$. The word “infinitesimal” signifies that the magnitude of the vector $v$ approaches zero but never actually reaches it in the differentiation process. Calculus expresses the growth rate of $P$ at $A$ as the limit of the difference ratio $\frac{\Delta y}{\Delta x}$, as the change in $y$ and $x$ become infinitesimally small. If we push $B$ infinitesimally close to $A$, $v$ becomes the tangent line (slope) of $P$ at point $A$. The process compares infinitesimal differentials of $x$ and $v$. We can apply the following image of differentiation (adapted from Mizutani 2011: 37).

Figure 5: Differentiation

---

12) W. G. Leibniz (a German mathematician; 1646–1716) and I. Newton (an English mathematician; 1642–1726) independently contributed to providing mathematical ground to the idea of differentiation and integration. The modern calculus uses symbols such as derivative $\frac{dy}{dx}$ and integral $\int y \, dx$, that are coined by Leibniz.

13) The slope (growth rate; derivative) $\frac{\Delta y}{\Delta x}$ is expressed as $\frac{y(x+\Delta x) - y(x)}{\Delta x}$, where $\Delta x$ corresponds to the infinitesimal change along the horizontal $x$-axis from $x$ at point $A$ to $x+\Delta x$ at point $B$ and $\Delta y$ corresponds to the infinitesimal change along the vertical $y$-axis from $y(x)$ at point $A$ to $y(x+\Delta x)$ at point $B$. Consider, for example, $y=x^3$. $\frac{\Delta y}{\Delta x}$ is $\frac{(x+\Delta x)^3 - x^3}{\Delta x}$. Expanding the numerator, we obtain $\frac{x^3 + 2x\Delta x + \Delta x^3 - x^3}{\Delta x}$, which becomes $2x + \Delta x$. Since $\Delta x$ is infinitesimally small (almost zero but not zero), we can ignore it. The slope (growth rate) $\frac{dy}{dx}$ of $P$ at the point $A$ on the curve $y=x^3$ is limit of $\frac{\Delta y}{\Delta x}$, which is limit of $2x + \Delta x$, which is $2x$. Similarly, the slope of $y=x^n$ is given by formula $y=nx^{n-1}$. The reader is referred to Strang (2010a: 9–14) for details.
Magnify P until we can see an infinitesimal structure. Push the magnification to the limit where we can see changes in \( x(\Delta x) \) and height \( y(\Delta y) \) that are infinitesimally small: almost zero but not zero. The vector \( \nu \) is so short that we can consider the curve P to be a straight line. The growth rate \( \frac{dy}{dx} \) is calculated to be the limit of \( \frac{\Delta y}{\Delta x} \), when \( \Delta x \) approaches zero: \( \Delta x \to 0 \).

Mathematicians battle against the formidable problems of division by zero and infinity using weapons like the limiting process. In the same manner, linguists battle against tough problems of structural growth by employing mechanisms of feature checking (elimination of uninterpretable formal features) with an infinitesimally small steps, i.e., the limiting process. As an example, consider an algorithm yielding a direct wh-question such as follows.

(11) Whom did Mary see?

Here, we ask what the relation is between the two functions of structural growth (i.e., distance) and derivational steps (i.e., speed). We abbreviate the conceptual-intentional system as CI, and the sensorimotor system as SM. Both CI and SM are predecessors shared in animal brains, whereas CHL is a mutant newcomer that has emerged in the human brain. Given this, we offer the following algorithm. AB is antibody (i.e., formal feature in F), AG is antigen (i.e., formal feature such as structural Case in a DP). Refer to Piattelli-Palmarini and Uriagereka (2004) for a hypothesis that CHL is a mutant virus-check (i.e., immune) system suddenly and accidentally evolved in the human brain; AG is a computational virus that lives in symbiosis with CHL.

(12) Algorithm of structure building and feature checking

a. 1. EM (V, Obj) \( \rightarrow \) VP formed
b. 2. EM (v*, VP)
c. 3. F inheritance (v*, V)
d. 4. \( \theta \) (Obj, patient)
e. 5. Probe (v*, V)
f. 6. Match (AB, AG)
g. 7. IM (v*, V) \( \rightarrow \) V adjoins to v
h. 8. Eliminate AG

14) As Edgar Morin (a French philosopher) redefined Homo sapiens (i.e., smart primates) as Homo demens (i.e., insane primates whose instinct is destroyed), CHL has severely affected CI and SM in human’s brain.
i. 9. Prove \( V, \text{Obj} \)

j. 10. Match \( \text{AB}, \text{AG} \)

k. 11. IM \( \text{VP}, \text{Obj} \)

l. 12. Eliminate \( \text{AG} \)

m. 9. Probe \( v^*, \text{Obj} \)

n. 10. Match \( \text{AB}, \text{AG} \)

o. 11. IM \( (v^*\text{P}, \text{Obj}) \)

p. 12. Eliminate \( \text{AG} \) [ACC] checked off

q. 13. EM \( (v^*\text{P}, \text{Subj}) \) \( v^*\text{P} \) formed

r. 14. EM \( (T, v^*\text{P}) \)

s. 15. F inheritance \( (T, v^*) \)

t. 16. \( \theta \) (Subj, agent)

u. 17. Transfer \( v^*\text{P} + \text{VP} \) \( v^*\text{P} \) transferred to CI and SM

v. 18. EM \( (C, \text{TP}) \)

w. 19. F inheritance \( (C, T) \)

x. 20. Probe \( (C, T) \)

y. 21. Match \( \text{AB}, \text{AG} \)

z. 22. IM \( (C, T) \) ............... \( T \) adjoins to \( C \)

A. 23. Eliminate \( \text{AG} \)

B. 24. Probe \( (T, \text{Subj}) \)

C. 25. Match \( \text{AB}, \text{AG} \)

D. 26. IM \( (\text{TP}, \text{Subj}) \)

E. 27. Eliminate \( \text{AG} \) [NOM] checked off; \( \text{TP} \) formed

F. 24. Probe \( (C, \text{Obj}) \)

G. 25. Match \( \text{AB}, \text{AG} \)

H. 26. IM \( (\text{CP}, \text{Obj}) \)

I. 27. Eliminate \( \text{AG} \) [Q] checked off; \( \text{CP} \) formed

Proceed in parallel
J. 28. Transfer CP+TP......CP transferred to CI and SM

In the above algorithm, the two sets of steps \(<i, j, k, l>\) (i.e., Obj-movement to Spec-VP) and \(<m, n, o, p>\) (i.e., Obj-movement to Spec-v*P) proceed in parallel (Chomsky 2008: 147). Furthermore, the two sets of steps \(<B, C, D, E>\) (i.e., wh-DP movement to Spec-TP) and \(<F, G, H, I>\) (i.e., wh-DP movement to Spec-CP) proceed in parallel (ibid.). The following contrast constitutes the empirical evidence for the parallel feature checking.

(13) a. * Of which car did the driver cause a scandal?
   b. Of which car was the driver awarded a prize?

The phase impenetrability condition (PIC) (Chomsky 2001: 13–14) is relevant here. Chomsky defines the guiding principle below, where \(\text{Ph}_1\) represents a strong phase and \(\text{Ph}_2\) represents the next highest strong phase. Note that CP and v*P are strong phases.

(14) The guiding principle

\[\text{Ph}_1 \text{ is interpreted/evaluated at } \text{Ph}_2.\]

PIC is defined below, where “domain” means “complement” and ZP represents the smallest strong phase.

(15) PIC

\[\text{In } [\text{ZP } Z \ldots [\text{ZP } \alpha [H \text{ YP}]])]\], the domain of H is not accessible to operations at ZP; only H and its edge \(\alpha\) are accessible to such operations.

Consider the derivation of (13a). Assume that wh-DP “the driver of which car” EMs at Spec v*P. If the wh-part “of which car” of the wh-DP IMs to Spec CP, IM violates PIC. According to PIC, domain v*P of T [H] is not accessible to operations at CP [ZP]; instead, only T [H] and its edges \(\alpha\) are accessible to such operations. Therefore, to account for the ungrammaticality of (13a), wh-DP must not IM to Spec, TP. We face a problem if wh-DP must IM to Spec, TP for EPP-checking.

Consider the derivation of (13b). The non-transitive light verb v is unaccusative/passive and it does not project a strong phase. Here, weak head v projects weak phase vP. The guiding principle states that a weak phase is not evaluated at the next highest strong phase. In other words, PIC does not care about weak phases. Therefore, the wh-DP in Spec, vP IMs to Spec, CP
without violating PIC. PIC requires that a term in Spec, v*P IM to Spec, TP before C targets the term; however, we lose the distinction if wh-DP IMs to Spec, TP before wh-IM to Spec CP. If it does, C would attract wh-DP in Spec, TP without violating PIC in (13a), thereby incorrectly predicting the example to be grammatical.

We face a double-bind situation here, i.e., the wh-DP in Spec, v*P/vP must reach Spec, TP to check [EPP] off and be probed by C, which is possible in (13b); however, wh-DP must not move to Spec-TP to account for ungrammaticality of (13a). Chomsky proposed the following solution here. T inherits features from C, making T = C, which in turn makes it possible for T to attract the non-wh part of DP “the driver” and for C to attract the wh-part “of which car” in parallel. In both (13a) and (13b), T attracts the non-wh part “the driver” for [EPP]-checking without violating PIC; however, PIC causes a distinction in terms of C’s attraction of wh-part. In (13a), C’s attraction of the wh-part “of which car” violates PIC because wh-DP is contained in strong phase v*P.

Conversely, in (13b), in contrast, C’s attraction of “of which car” does not violate PIC because wh-DP is contained within a weak phase. Chomsky speculated that the same simultaneity takes place in v*-V. Figure 6 presents our approximate translation of the above algorithm into a graph.

Consider an infinitesimal growth $\Delta y$ at a particular instant of time, say, 24. The limit of growth is $\Delta y \to 0$ and the particular instant is $\Delta x$ with limit $\Delta x \to 0$. Derivative or growth rate $\frac{dy}{dx}$ at time 24 is $\frac{dy}{dx} = \text{limit of } \frac{\Delta y}{\Delta x} \to 0$. At time 24, T probes Subj and C probes wh-Obj simultaneously. Times 9 and 24 initiate two-dimensional time. We speculate that parallel feature checking obeys MC: minimization of time. We propose the following.
(16) Feature checking controls derivative $\frac{dy}{dx}$ at an instant in time.

Each step of the feature-checking process is expressed as growth rate $\frac{dy}{dx}$. Syntax focuses on an infinitesimal $\frac{dy}{dx}$, which is the limit of a particular step of the feature checking process.

What is the idea of the integral $\int \frac{dy}{dx} dx$? “Integration is a problem of adding up infinitely many things, each of which is infinitesimally small” (Strang 2010a: 229). “The problem of integration is to find a limit of sums. The key is to work backward from a limit of differences (which is the derivative). We can integrate $v(x)$ if it turns up as the derivative of another function $f(x)$. The integral of $v = \cos x$ is $f = \sin x$. The integral of $v = x$ is $f = \frac{1}{2} x^2$. Basically, $f(x)$ is an ‘antiderivative’” (ibid). The keyword is “work backward.” Recall that calculus deals with pairs of functions.

Here, the structural growth $f(x)$ is function 1, and the growth rate $v(x)$ is function 2. The derivative (i.e., differentiation) moves from the known function 1 to the unknown function 2. The integral (i.e., integration) moves from the known function 2 to the unknown function 1, i.e., the antiderivative.

Surprisingly, CHL seems to simultaneously perform both differentiation and integration; for structure building in CHL, a derivative (i.e., feature checking; AB’s probing, matching, and eliminating AG; a limit of differentiation) is used in parallel with an integral (i.e., IM of copies; a limit of sums). An example of an integral on the SM side is Demerge proposed by Fukui and Takano (1998); they hypothesize the symmetry of derivation, i.e., the computations in the overt (pre-Spell-Out) component and the computations in the phonological component (the interface connected with SM) are symmetric. They propose Linearization, i.e., when applied to the syntactic object $\Sigma$, Demerge yields $[\alpha, [\Sigma - \alpha]]$, where $\alpha$ is an $X^{now}$ constituent of $\Sigma$, and...

15) More precisely, $f = \frac{1}{2} x^2 + C$, where $C$ is a free integral constant. Given a derivative of $x^3$ is $nx^{n-1}$, the integral of $y = 3x^3$ is $y = x^4$, which we write $\int y dx = \int 3x^2 dx = x^4 + C$. We ignore $C$, for it disappears in the derivative. An integral is the inverse (i.e., backward calculation) of a derivative. Refer to Mizutani (2011: 53).

16) We also observe the symmetry of derivation on the CI side. Namely, two different structures yield two distinct meanings, e.g., an expression such as “purple people eater” is ambiguous. A structure $[_{sp} [_{sp} purple people] eater]$ yields the meaning that people are purple, whereas a structure $[_{sp} purple [_{sp} people eater]]$ yields the meaning that the eater is purple. Further, no ternary structure is allowed, thereby disallowing the meaning that both the people and eater are purple.
Concatenate turns \([\alpha, (\Sigma-\alpha)]\) into \(\alpha + (\Sigma-\alpha)\) (ibid).\(^{17}\) Here, function 1 is Linearization (i.e., the phonetic change in time) determined in SM, and function 2 is the structural growth done by Merge. A listener’s CHL uses a derivative (i.e., Merge with feature checking) to move from function 1 to function 2, i.e., rebuilding and computing structures as the listener hears sentences, whereas a speaker’s CHL uses an integral (i.e., Demerge) to move from function 2 to function 1, i.e., linearizing structures as the speaker produces sentences.

Further, we speculate that the growth rate or speed of VP- and CP-buildings is greater than that of v*P- and TP-building. Given this, the slope at \(x\) is the limit of algorithmic steps in CHL spending more time in building v*P (i.e., the first strong phase) and TP (i.e., the second weak phase). Finally, below, Strang explains how the first derivative \(\frac{dy}{dx}\) and second derivative \(\frac{d^2y}{dx^2}\) differ.

(17) “In ordinary language, the first derivative \(\frac{dy}{dx}\) tells how fast the function \(y(x)\) is changing. The second derivative tells whether we are speeding up or slowing down” (Strang 2010a: 11).

The first derivative tells us change of speed of the growth rate, while the second derivative tells us change of acceleration change. After completion of VP, the growth rate slows down. Then, after completion of TP, CHL speeds up again to complete CP.

5. Conclusions

LC predicts that syntax is the calculus of F. Calculus is all about change; thus, the calculus of F calculates the change in F. If we assume that what is relevant here is F at the initial state \(S_0\) (i.e., the baby brain grammar) of CHL, no set parameter exists. Equation \(2^n=1\) expresses this initial state and indicates that we have one genotype that yields the language of Homo sapiens, which we call “Homosapienses.” Further, BCC predicts that F is the locus of parametric difference. If we adopt the number of parameters (i.e., ON-OFF switches) as 12, we have \(2^{12}=4096\) phenotypes (i.e., specific languages), which are realized at steady state \(S_n\) of CHL, which represents an adult brain grammar. Both BCC and LC are potentially expressed by the exponential function \(2^z=y\), where growth rate (i.e., growing speed) equals approximately 70% of structural growth (i.e., structural distance development) at every instant in time; the growth rate is approximately

\(^{17}\) Demerge shares the basic idea with the linear correspondence axiom (LCA; Kayne 1994), which informally dictates that a higher term is pronounced earlier
2867 (i.e., 70% of 4096) at instant in time when CHL contains 12 switches. This high growing speed appears to guarantee the quick and easy mother-language acquisition of our species.

The question remains as to why the relevant logarithmic function is $12 = \log_{4096}$, i.e., why CHL reaches equilibrium with 12 switches. Relying on Polya’s heuristic advice to solve the problem, we sketch calculus graphs of brain change and a feature-checking algorithm without knowing how calculus expresses the graphs. However, without the insight of LC, we cannot speculate that feature checking by F is an algorithm that controls derivative $\frac{dy}{dx}$ (i.e., the speed or growth rate of sentence structure building) at every instant in time. To feature check and eliminate the matching AG, an AB in F microscopically looks into an infinitely small structure $\Delta y$ at infinitesimal time step $\Delta x$.

References


(Accepted on 15 November, 2017)