Symmetry Breaking and Economy in $\mathbf{C}_{\text{HL}}$\textsuperscript{1)}

Koji ARIKAWA\textsuperscript{*}

1. Conceptual Discussion

1.1. What Is Symmetry?

What does it mean to say that a physical law has symmetric properties? Suppose we get the same result when we transform an equation. We then say that the equation has symmetric properties. Suppose we have a state in which we cannot tell whether an observer A is right or an observer B is right in describing a situation. Then we say that the relevant physical law has symmetric properties (Feynman 1965). Consider a parabola as in the following.

\begin{equation}
(1)
\end{equation}

\[ y = x^2 \]

Jenkins (2000: 161) explains the basic notion of symmetry: "The graph of the equation for the parabola \((y=x^2)\) has symmetry under reflection about the \(y\)-axis. If we substitute \(-x\) for \(x\) in the equation \(y=x^2\), the form of the equation remains unchanged. The equation is "invariant" under the following reflection transformation: \(x \rightarrow -x\). The equations possess symmetries."

We will see how built-in principles as symmetry breaking and economy interact each other in the computational system of human natural language ($\mathbf{C}_{\text{HL}}$).

This paper is divided into two parts: (i) conceptual discussion, and (ii) empirical discussion. In the first half of (i) from Section 1.1 to Section 1.6, various key words are introduced such as symmetry, cost (economy), entropy, spontaneous symmetry breaking, Fibonacci numbers, chirality, L-PS (laevorotatory phrase structure), D-PS (dextrorotatory phrase structure), LCA (Linear Correspondence Axiom), and BPS (Bare Phrase Structure). In 1.7, the problems of LCA and BPS as originally given are introduced. In 1.8, LCA is modified so as to become compatible with BPS. In 1.5–1.8 particularly, we offer a simple explanation of phrase structure without employing the notions as label and projection. In 1.9, we argue for the extremely general axiom called Polarization Axiom (PA), which determines the order and structure of $\mathbf{C}_{\text{HL}}$.

\textsuperscript{*} Faculty of Letters, St. Andrew's University

\textsuperscript{1)} This study is supported by (i) Research Grant provided by St. Andrew's University (Momoyama Gakuin University) for 2000~2001, and (ii) Grant-in-Aid for Scientific Research (C) provided by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) for 2001~2002 (Project #13610669). The project title is "Symmetry Breaking ~ Word Order and Information Generation ~."
(the computational system of human natural language). In 1.10, we explain what it means to say that symmetric structures shows unstable states in $C_{HL}$. In 1.11, the nature of information generation in 1-dimension (1-D) is discussed. In 1.12, we argue for the existence of imperfect asymmetry in $C_{HL}$. In 2.1–2.3, we point out some problems of the analysis of passive sentence proposed by Nunes (2001) and propose a simpler explanation. In 2.4–2.5, we introduce the Hybrid Model and the Two-Chain Hypothesis to explain orders and structures in $C_{HL}$. We argue that the former can account for the data better if it incorporates the latter. We also modify the Law of Conservation generally assumed for $C_{HL}$. In 2.6–2.8, we offer evidence that supports the Hybrid Model + the Two-Chain Hypothesis. In 2.9, the interaction among MLC (Minimal Link Condition), PBC (Proper Binding Condition), UCA (Uniformity Corollary of Adjunction), Economy principles, and symmetry breaking is shown. In 2.9.1–2.9.2, we take up the superiority effect, wh-island effect, HMC effect, and superraising cases. We offer new evidence for the existence of the superiority effect in Japanese. In 2.9.3, we argue against UCA and propose that UCA be eliminated. Section 2.9.4 offers a simpler account of the subject condition effect, CNPC effect, and adjunct condition effect found in Japanese. Section 3 summarizes the main points of this paper.

1.2. Symmetry Breaking and Cost

Citing Gell-Mann (1994: 194), Jenkins (2000: 154) states that “symmetries may be spontaneously hidden or “broken.” That is, even though the physical law itself may exhibit perfect symmetry, its realization in nature may be asymmetric.” So even though the physical law itself (e.g., Linear Corresponding Axiom (LCA) for natural language (Kayne (1994)) may exhibit perfect symmetry, one of its realizations in nature (e.g., the human brain), namely, the effects (e.g., word orders and phrase structures) of the computational system of human language ($C_{HL}$) may be asymmetric.

We propose that the notion of entropy is relevant. Entropy is the amount of mixedupness, disorderness or randomness in a given system. The symmetric state contains low entropy. The symmetric state requires the maximum degree of energy to maintain the highly ordered state. Since it contains the maximum energy and more information, it is unstable, and it is costly to maintain such symmetric state. An asymmetric state has a high amount of entropy. Asymmetric state requires the least degree of energy: no energy is needed when you just let things go. Since it contains the least energy and less information, it is stable, and it is costless

2) "The essence of spontaneous symmetry-breaking lies in this very circumstance: equations with a particular symmetry can have solutions that individually violate that symmetry, although the set of all solutions is symmetrical" (Gell-Mann, 1994: 194).

3) An American theoretical physicist/chemist J. Willard Gibbs (1839–1903) and an Austrian theoretical physicist Ludwir Boltzmann (1844–1906) used the word entropy in this sense in statistical mechanics. The word entropy was used by a German theoretical physicist Rudolf Clausius (1822–88) in 1865 to indicate amount of non-reversible nature of thermal phenomenon, which first had been found by a French physicist/mathematician Sadi Carnot (1796–1832).

4) The more entropy, the less information. We are indebted to the works by an American mathematician/electronic engineer Claude Elwood Shannon (1916–) about the relationship between entropy and information.

5) "... (T) he symmetric state is unstable, whereas the asymmetric state turns out to be stable..." (Jenkins (2000: 164))
to maintain such state. The least-effort (energy) flavor of physical laws in general is in conformity with the laws of thermodynamics, roughly stated as follows\textsuperscript{6}.

(2) \textit{The First Law of Thermodynamics}

Energy is conserved in a closed system.

(3) \textit{The Second Law of Thermodynamics}

In a closed system, entropy always increases.

So things always go wrong (toward disorder). Let us put this in a diagram.

(4)

\begin{center}
\begin{tikzpicture}
\draw[->] (-3,0) -- (3,0) node[right] {symmetric \quad asymmetric};
\draw[->] (0,-3) -- (0,3) node[above] {entropy};
\draw (0,0) -- (3,3) node[right] {symmetry-breaking \rightarrow};
\draw (0,0) -- (0,3) node[above] {High-energy};
\draw (0,0) -- (3,0) node[right] {Low-energy};
\end{tikzpicture}
\end{center}

The diagram in (4) shows what happens in a closed (ideal / equilibrium) system. The Second Law holds for a closed system. The natural language system, a realization of the emergent and self-organizing properties of the human brain, is not an equilibrium system\textsuperscript{7}. In a non-equilibrium system, there are two types of energies involved: system-internal energies (interacting with each other) and the free energy of entropy. The further interaction of these two energies determines the present state of the system. In reality, the human language faculty is an open (real / non-equilibrium) system, in which internal and external information dynamically causes its structural change. What actually is happening then is the following.

(4')

\begin{center}
\begin{tikzpicture}
\draw[->] (-3,0) -- (3,0) node[right] {symmetric \quad asymmetric};
\draw[->] (0,-3) -- (0,3) node[above] {entropy};
\draw (0,0) -- (3,3) node[right] {symmetry-breaking \rightarrow};
\draw (0,0) -- (0,3) node[above] {High-energy};
\draw (0,0) -- (3,0) node[right] {Low-energy};
\end{tikzpicture}
\end{center}

\textsuperscript{6} For the correlation between the least-effort principle in physics and the economy principle in C\textsubscript{HL}, see Fukui (1996).

\textsuperscript{7} Human brain is the typical case of a complex system. A complex system has the following characteristics: (a) it shows the butterfly effect, (b) it is sensitive to the initial-state conditions, and (c) iteration of a simple operation plays an important role. C\textsubscript{HL} in human brain seems to satisfy these requirements: (a) the final state of C\textsubscript{HL} shows remarkable variation in the phonetic and formal aspects (the butterfly effect), (b) C\textsubscript{HL} is sensitive to the initial-state (UG) conditions, (c) a simple operation Merge is crucial. A complex system has boundary conditions, within which the system allows variations. We take legibility conditions imposed on C\textsubscript{HL} to be these boundary conditions.

\textsuperscript{8} Deletion/Erase distinction is crucial.
Things go wrong (toward disorder) until they reach the point \( P \) where some degree of order is obtained. In our context of \( C_n \), the process up to a point \( P \) is the structure-building derivation, and \( P \) is the point of convergence (every uninterpreted features are deleted\(^9\)), and every feature gets interpreted at relevant interfaces. Put differently, symmetric state contains more information, i.e., undeleted uninterpretable features. The process of entropy-decrease is the process of uninterpretable-feature deletion. From the point of view of ontogeny, the process up to \( P \) is the process of language acquisition (the initial state of language faculty (UG or \( S_0 \) undergoes change), and \( P \) is the point where UG attains its relatively steady final state \( (S_0)^3 \)).

1.3. Fibonacci Numbers and Word Order

If we count the number of clockwise-running spirals and counterclockwise-running spirals in the sunflower, typically we have 21 clockwise spirals and 34 counterclockwise spirals. Or in a larger sunflower, we find 34 clockwise spirals and 55 counterclockwise spirals. If we have still larger one, we find 55 clockwise spirals, and 89 counterclockwise spirals. These are numbers which we call Fibonacci numbers. The definition of Fibonacci numbers is the following\(^{10}\).

\[
(F_n) = \text{def. } F_1 = F_2 = 1, \quad F_{n+1} = F_n + F_{n-1}, \quad \text{where } n \geq 2
\]

The following is a sequence of Fibonacci numbers.

\[
1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 377, 610, 987, 1597, 2584, \ldots
\]

We find Fibonacci numbers everywhere in the nature, e.g., sunflower spirals, pinecone spirals, tree branching, generative spirals in primordia, and nautilus spirals (logarithmic spirals), to mention a few. Jenkins (2000: 157) introduces the importance of the Fibonacci numbers as follows. The divergence angle, which is the angle formed at the center of the spiral between successive primordia, is 137.5°. Auguste Bravais, one of the founders of modern crystallography, and his brother, Louis, found this angle. The Fibonacci numbers and the divergence angle are closely connected: the ratio of successive Fibonacci numbers is $34:55$ ($\cong 0.61818$), which approaches the inverse of what the ancient Greeks called the golden number $\phi$.

---

9) The process parallels that of disappearance of mirror motion in human infants: new-born babies are ambidexterous, the property of which they usually lose later. Jenkins (2000: 165) suggests the possible reason why nature contains asymmetric properties: "$\ldots\text{if the laws of governing a physical system present natural selection with a number of solutions, some of which are asymmetric and stable, and others which are symmetric, but unstable, it is quite possible that it will pick the asymmetric solutions, particularly if stabilizing and maintaining the symmetric solution is too costly.} \text{LCA, a natural language principle governing linear ordering and hierarchical structure, is symmetric, but the actual effects induced by the interaction of principles and parameters in } C_n \text{ is a realization of spontaneous symmetry breaking.}$

Symmetry Breaking and Economy in C_{HL}

(7) \[ \Phi = -\frac{1+\sqrt{5}}{2} \approx 1.61803 \]

"If we multiply the inverse of \( \Phi \) (\( \approx 0.61818 \)) by 360°, we get 222.5°, which, when subtracted from 360°, yields 137.5°, the angle observed by the Bravais brothers" (Jenkins (2000: 157)).

What might be the connection between the Fibonacci numbers and natural language? D'Arcy Thompson directed our attention to a stimulating question of "how tension and pressure can interact with structural anisotropies and asymmetries to determine the shape of biological structures" (D'Arcy Thompson (1992)). Capitalizing on D'Arcy Thompson's insights, Jenkins (2000: 152) speculates that "properties of language, such as syntactic (a)symmetry, could in principle arise, like the Fibonacci numbers, by physical constraints, 'limited by the properties of space and numbers,' in D'Arcy Thompson's word (ibid. 164)." In a sense, the study of linear ordering and structure of natural language is like studying the property of the Fibonacci numbers in phyllotaxis or crystallography. Linear ordering and structure of natural language is an "emergent" phenomenon through a symmetry-breaking mechanism. What is relevant here is the human brain. In this connection, Chomsky (1994b: 83-84) talks about "something about the space of physical possibility," speculating that "it (=the possibility) might be so narrow that under the particular conditions of human evolution, there's one possibility for something with \( 10^{11} \) neurons (or \( 10^{14} \) synapses) packed into something the size of a basketball: namely, a brain that has these computational properties (=natural language)." Jenkins (2000: 165) goes on further: "the origins of (some) properties of language, such as (a)symmetries in syntax perhaps, would be more akin to the physical evolution of molecular chirality."\(^{11}\)

1.4. Chirality in Biological System

In the biological world, all natural amino acids are left-handed (l-type, or laevo-rotatory-type), and natural deoxyribose is right-handed (d-type, or dextro-rotatory-type)\(^{12}\). The l-type molecule has a property of leftward polarization of light, and d-type molecule rightward. Consider the 3-D structures of amino acid and deoxyribose, which are basic building blocks of DNA. An amino acid contains asymmetric carbon atoms. Amino acids have a 3-D structure consisting of a regular tetrahedron that has C at the center, with COOH, H, NH\(_2\), and R (a variable for 20 distinct amino-acids) at the four apices\(^{13}\).

\(^{11}\) It is worth bearing in mind that Chomsky's book *Syntactic Structures* (Chomsky (1957)), which was based on his small lecture notes, a starting point of this whole biolinguistic program back in 1950's, was extensively reviewed by a chemist at MIT (the review was as long as the book, about 100 pages). Chomsky often says that the present state of the biolinguistic program is akin to the state of chemistry in the late 19th century, most of which was integrated without changing the core into the newly born physics in the late 1920's. We believe that the hard problem of reducing the present-state cognitive studies of human brain into the modern physics is impossible in principle, namely, the present-state cognitive science should be integrated into the still-to-be-discovered physics in the future. See Chomsky (1993b).

\(^{12}\) This is a general tendency. There are exceptions.

\(^{13}\) See Hirayama (2000), Yanagisawa (1997), Tominaga (2001), Gardner (1990), and Stewart & Golubitsky (1992) for relevant discussions.
(8a) and (8b) are 3-D-mirror images of each other (a plane-mirror, as that in our world): it is impossible for the two to overlap each other in 3-D. These are said to be chiral. For the sake of discussion, let us confine ourselves to 2-D\(^{14}\). The following are 2-D structures of asymmetric amino acids.

![](image)

1-α-amino acid

d-α-amino acid

The 1-α-amino acid and d-α-amino acid are 2-D mirror images of each other (the mirror is a line in 2-D). Suppose we rotate d-α-amino acid 180° around the central point C at 2-D. The result is the following.

![](image)

180°-2-D-rotation of d-α-amino acid

(9a) and (9a') do not overlap each other. In fact, it is impossible for the two to overlap each other at 2-D. Hence, (9a) and (9b) are chiral. What about deoxyribose? The following are the 2-D structures of asymmetric deoxyribose.

---

14) When we talk about chirality using mirror images, the dimension must be consistent. For example, if we talk about chirality of 2-D structures, we have to confine ourselves to 2-D.
In 2-D, it is impossible for (9°a) and (9°b) to overlap each other. That is, the two structures are chiral. Generally found for the DNA of all creatures on this planet are l-α-amino acid and d-deoxyribose but, yet unknown reasons, not their mirror images d-α-amino acid and l-deoxyribose.

1.5. Chilarity in C_HL.

Merge is a basic operation in C_HL. It concatenates two terms^{15}. Given α and β, Merge applies to α and β, and we obtain a new term γ^{16}.

\[ \gamma = \alpha \cdot \beta \]

The new term γ is not a union set of α and β: γ is identified either by α or by β^{17}. Let us call the structure left-handed (L-phrase structure; L-PS) if γ is identified by β. If γ is identified by α, the structure is right-handed (D-PS)^{18}. In other words, when two elements merge, the outcome structure must be asymmetric in that the newly constructed element must be identified by either of the two elements^{19}. The identification by a terminal is indicated by a bold line.

---

15) For a general framework of the Minimalist (Biolinguistic) Program, see Chomsky (1994a) and Chomsky (1995b).
16) γ is a label of the structure. We will propose a label/projection-free system when we provide the modification of the LCA. See Collins (2001) for arguments for eliminating labels and projections. The two types of Merge, internal Merge and external Merge, are introduced later. See Chomsky (2001: 9) for the distinction.
17) See Chomsky (2001: 6) for the notion of identity in place of projection: "(\(\alpha, \beta\)) is identified either by "α or by "β."
18) The distinction between lower l/d and capital L/D matters. In chemical studies, the lower letters are used to refer to polarization of light (effects induced by the structure of molecule), and capital letters the structure of an individual molecule.
19) See Chomsky (2001: 7, fn. 23) for a view that "α and β are merged with no asymmetry."
(11a) and (11b) are chiral in 2-D: it is impossible that the two overlap each other in 2-D\(^2\). They are asymmetric phrase structures. Notice that the notion of asymmetry is distinct from that of Kayne (1994)'s. However, the Linear Correspondence Axiom (LCA), which connects 1-dimensional linearity and 2-dimensional hierarchical structure, has the symmetric property as a principle, in that the output (linear ordering) is invariant whether we have L-PS or D-PS. Let us assume the following definition of the LCA, which is originally proposed in Kayne (1994)\(^2\), and modified by Uriagereka (1998).

\[(12) \quad \text{Linear Correspondence Axiom (LCA)}
\]
\[\text{A category } \alpha \text{ precedes a category } \beta \text{ iff (a) } \alpha \text{ asymmetrically}
\]
\[\text{c-commands } \beta, \text{ or (b) } \gamma \text{ precedes} \beta \text{ and } \gamma \text{ dominates } \alpha.
\]

(Uriagereka (1998))

Assume the definition of c-command as given in the following (Chomsky 1986, Kayne 1994).

\[(13) \quad \text{Definition of C-Command (domination version)}
\]
\[\alpha \text{ c-commands } \beta = \text{def. } \alpha \text{ excludes } \beta \text{ and the first node dominating } \alpha \text{ dominates } \beta.
\]

\[(14) \quad \text{Definition of Exclusion}
\]
\[\alpha \text{ excludes } \beta = \text{def. no segment of } \alpha \text{ dominates } \beta.
\]

\[(15) \quad \text{Definition of Domination}
\]
\[\alpha \text{ dominates } \beta = \text{def. every segment of } \alpha \text{ dominates } \beta.
\]

Given the above definition of the LCA, both (11a) and (11b) yield the same linear ordering \(\langle \alpha, \beta \rangle\). A complement precedes the head here. The result is incompatible with Kayne (1994)'s, in which a head precedes the complement. Let us pursue how we can make the asymmetric identification compatible with Kayne's solution. Notice that the structures in (11) are representationally indistinguishable from adjunction structures. Kayne (1994) assumes X'-the-

\(^2\) We are dealing with set theoretical notions. Concatenation of a set of categorial features \(\{\alpha\}\) and a set of categorial features \(\{\beta\}\) give rise to either a set \(\{\alpha, \beta\}\), where the label is \(\alpha\), or a set \(\{\beta, \alpha\}\), where the label is \(\beta\). The members of a set of categorial features are \(\{\pm N, \pm V, \pm A, \pm P\}\).

\(^2\) The original definition of the LCA is the following (Kayne (1994: 5-6)). For a given phrase marker \(P\), we consider the maximal set of \(A\) of ordered pairs of nonterminals \(T<X_i, Y_j>\) such that for each \(j\), \(X_i\) asymmetrically c-commands \(Y_j\).

(i) \text{Linear Correspondence Axiom}
\[d(A) \text{ is a linear ordering of } T.\]
ory and the definition of c-command, in which a non-branching node as well as a branching node blocks c-command relation.\(^{22}\) If we assume asymmetric relation to include asymmetric identification (feature-asymmetric relation) in addition to asymmetric c-command (structure-asymmetric relation), we can talk about asymmetric property in the case of external Merge without speculating the level of projection of \(\alpha\). Not only the notion of label, the notion of projection can be eliminated. We argue for the two distinctive elements of asymmetric c-command in Section 1.8.

1.6. Symmetric LCA Forces Asymmetric Relation

LCA is a symmetric principle. LCA is like the equation \(y = x^2\) in that the two distinct inputs (e.g. \(x\) or \(-x\)) yield an invariant output (an invariant linear order, corresponding to invariant \(x^2\)).

\[
\begin{array}{|c|c|c|}
\hline
\text{equation} & \text{input} & \text{output} \\
\hline
y = x^2 & y = x, y = -x & \text{invariant } x^2 \\
\hline
\text{LCA} & \text{L-PS, D-PS} & \text{invariant order} \\
\hline
\end{array}
\]

There is a difference, however. In the former case, the two inputs (\(x\) and \(-x\)) have point-symmetric nature in 1-D, whereas in the latter case, the two inputs (L-PS and D-PS) are line-asymmetric in 2-D. In the case of \(y = x^2\), the symmetric output is assured by the symmetric nature of the input. In order to obtain symmetric property of the two inputs in LCA, we have to raise the dimension from 2-D to 3-D. In 3-D, 2-D structures (L-PS and D-PS) become symmetric: they overlap each other. Notice, however, that it is not usual to talk about chirality in terms of raising the dimension. The level of the dimensions must be kept consistent. It is not clear how we can connect this dimension-expansion to the symmetric property of LCA.\(^{23}\)

Given LCA, only asymmetric relation is permitted. Symmetric (mutual) c-command is prohibited. The meaning of “asymmetric” relation at external Merge (base structure) is asymmetric identification (feature-asymmetric).

\(^{22}\) Kayne (1994) assumes the standard X'-theoretic assumption: “Certain features (categorial features) project from a terminal element to form a head, then on to form higher categories with different bar level” (Chomsky (1994a)). In fact, LCA crucially relies on this assumption. Chomsky (1994a) abandons this assumption in his bare phrase structure (BPS) theory. However, the fundamental notion of asymmetry survives in BPS and the label/projection-free phrase structure, the notion on which we capitalize in order to revise LCA.

\(^{23}\) We argue later that a hierarchical structure is determined, to some extent, for each case of internal Merge (base structure) and of internal Merge (derived structure) by a general axiom.
(17a) and (17b) are chiral at 2-D. Consider the symmetric relations.

(18) both-handed

\[
\begin{align*}
\text{a.} & \quad \text{a} \quad \text{b} \\
\text{a} & \quad \text{a} \\
\text{b} & \quad \text{b} \\
& \quad \text{a = identifier} \\
& \quad \text{(Identified by both terminals)} \\
& \quad \downarrow \\
& \quad \text{too symmetric (impermissible)} \\
\end{align*}
\]

(18a) is excluded by Economy principle: the same identification process is done twice, either one is redundant. (18b) is excluded because it yields an element which is not identified (it is invisible to C_m), i.e., a superfluous element is produced. Thus, the property of natural language in which a structure must have an identifier (a head, X’-theoretically) is deduced from the Economy principle.

1.7. Problems of BPS-based LCA
Let us repeat the definition of the LCA, which is originally proposed in Kayne (1994), and modified by Uriagereka (1998).

(19) **Linear Correspondence Axiom (LCA)**
A category α precedes a category β iff (a) α asymmetrically commands β, or (b) γ precedes β and γ dominates α.

(Uriagereka (1998))

Assume the definition of c-command as given in the following (Chomsky 1986, Kayne 1994).

(20) a. **Definition of C-Command (domination version)**

24) From the point of view of symmetry, both-handed structure is too symmetric to exist, too unstable and costly to maintain. We argue later that the cost is related to unchecked uninterpretable features.

25) The original definition of the LCA is the following (Kayne (1994: 5-6). For a given phrase marker P, we consider the maximal set of A of ordered pairs of nonterminals \( T < X_i, Y_i > \) such that for each j, \( X_j \) asymmetrically c-commands \( Y_j \).

(i) **Linear Correspondence Axiom**
d(A) is a linear ordering of T.
Symmetry Breaking and Economy in Ch. 31

\( \alpha \) c-commands \( \beta = \text{def.} \) \( \alpha \) excludes \( \beta \) and the first node dominating \( \alpha \) dominates \( \beta \).

b. Definition of Exclusion
\( \alpha \) excludes \( \beta = \text{def.} \) no segment of \( \alpha \) dominates \( \beta \).

c. Definition of Domination
\( \alpha \) dominates \( \beta = \text{def.} \) every segment of \( \alpha \) dominates \( \beta \).

Let us consider how we can modify LCA within the bare phrase structure theory (BPS)\(^{26}\). Consider the following BPS.

(21)

V1 identifies the structure\(^{27}\). Let us indicate the identification relation by a bold line. According to the original LCA, the structure in (21) does not determine any linear ordering of two terminals, V and Obj, since they mutually (symmetrically) c-command each other. Suppose either V or Obj raises to a higher position. Assume that the LCA is insensitive to traces (Chomsky (1994a)).

(22)  

In (22a), V adjoins to v1. V excludes the functional category v. The first node dominating V does not exist. This is so because every identical branch becomes a segment of the same category, given BPS. In fact, the branch vs. segment distinction is lost in BPS. In the X'-theory, the notion of branch is used with the notion of projection, whereas the notion of segment is used with the operation of Adjoin, which does not involve projection. Since the notion of projection is dispensed with in BPS, all branches become segments. Crucially, v3 does not dominate V in (22a). The same situation holds even when T and C merge at the later stages.

---

26 See Chomsky (1994a) and Chomsky (2001). The number on the term is used only for expository purpose.

27 More strictly, identity means non-distinctness. See Chomsky (2001) for the notion of identification. See Collins (2001) for label-free phrase structure, which we adopt when we modify the LCA. The notion of identification still assumes the existence of label. In Collins (2001), the label V2 does not exist.
Neither T2 nor C2 c-command V. In (22a), V c-commands nothing, v1 does not exclude V or Obj, therefore v1 does not c-command V or Obj. The first node dominating Obj does not exist. Therefore, Obj c-commands nothing. It follows that there is no c-command relation among V, v1, and Obj in (22a). LCA and BPS predict that (22a) is not a possible phrase structure.

What about (22b)? The first node dominating Obj does not exist. Obj c-commands nothing. v1 does not exclude Obj or V. Therefore, v1 c-commands nothing. The first node dominating V1 does not exist. Therefore, V1 c-commands nothing. It follows that there is no c-command relation among Obj, v1, and V in (22b). LCA and BPS predict that (22b) is not a possible phrase structure, either. If (21), (22a), and (22b) are all excluded, the theory incorrectly predicts that the natural language lacks linear ordering or that the order is not determined, which is contrary to fact.

Let us keep BPS. A possible way out to solve the dilemma is to modify the definition of c-command as in the following. We adopt the less strict version of c-command. 

\[(23)\]
\[a. \text{Definition of C-Command (containment version)}
\]
\[\text{a c-commands } \beta = \text{def. } \alpha \text{ excludes } \beta \text{ and the first node containing } \alpha \text{ contains } \beta.\]

\[b. \text{Definition of Containment}
\]
\[\alpha \text{ contains } \beta = \text{def. some segment of } \alpha \text{ dominates } \beta.\]

Consider (22a). V excludes v1, but v1 does not exclude V. V asymmetrically c-commands v1, but V does not asymmetrically c-command Obj. Obj does not c-command V or v1. The orders between V/Obj and v1/Obj are not determined. It follows that (22a) is not a possible phrase structure. What about (22b)? Obj asymmetrically c-commands v1, and v1 asymmetrically c-commands V1. The order is <Obj, v1, V>. The theory predicts that the all natural languages have OV order, and that V0 order does not exist, which is contrary to fact.

Incidentally, there is a way (22a) becomes a possible option. Assume that the [V + v] amalgam in (22a) is restructured as a terminal element. The amalgam [V + v] excludes Obj, and the first node containing the amalgam (= v3) dominates Obj. Obj does not c-command the amalgam.

---

28) See Chomsky (2001: 116), Seely (2000), and Collins (2001: 7) for the similar proposal: \(\alpha \text{ c-commands } \beta \text{ iff } \beta \text{ is contained in the sister of } \alpha.\) We need the notion of exclusion to deal with the adjunction. We assume self-containment: \(\alpha \text{ contains itself.}\)
Therefore, the amalgam asymmetrically c-commands Obj. It follows that (22a) yields the linear ordering of VO. It follows that V must raise in VO-language, and that O must raise in OV-language, as in (22b). This conclusion contradicts some of the empirical evidence found in these languages: there is evidence that main V remains in situ in VO-language (Lasnik 1981, 1995; See Section 2.4.), whereas V overtly raises in OV-language (Whitman 1997), Koizumi (1995), Nakajima (1999)). If these evidence were correct, it follows that VO-language has OV order after Obj-raising, given that either V or O must raise (V remains according to this evidence), which is a contradiction. The theory also predicts that OV-language ends up having VO order, which is a contradiction. In sum, we end up being in a complicated situation if we assume LCA and BPS in their original forms, however we modify the relevant definitions.

1.8. Feature-Asymmetric vs. Structure-Asymmetric C-Command

Consider the following pair of phrase structures produced by external Merge, in which a lexical item is taken from the Lexicon to form the base structure. We assume the label/projection-free phrase structure (Collins 2001)).

\[
\begin{align*}
(24) & \\
\begin{array}{c}
(a) \\
\begin{array}{c}
\alpha \\
\beta \\
\end{array}
\end{array} & \\
\begin{array}{c}
(b) \\
\begin{array}{c}
\beta \\
\alpha \\
\end{array}
\end{array}
\end{align*}
\]

LCA excludes both structures in (24); they are too symmetric. But is it true that the structures in (24) are symmetric? We argue that the structures in (24) are asymmetric in nature. Collins (2001) claims that the syntactic relations include at least the following.

\[
\begin{align*}
(25) & \\
\begin{array}{ll}
(a) & \text{Theta (X, Y)} \\
(b) & \text{EPP (X, Y)} \\
(c) & \text{Agree (X, Y)} \\
(d) & \text{Subcat (X, Y)}
\end{array} & \\
\begin{array}{ll}
& \text{X assigns a theta-role to Y} \\
& \text{Y satisfies the EPP feature of X} \\
& \text{X matches Y, and Y values X} \\
& \text{X subcategorizes for a feature Y}
\end{array}
\end{align*}
\]

Consider the concrete structure of external Merge involving a transitive V.

\[
\begin{align*}
(26) & \\
\begin{array}{c}
(a) \\
\begin{array}{c}
V \\
\text{Obj}
\end{array}
\end{array} & \\
\begin{array}{c}
(b) \\
\begin{array}{c}
\text{Obj} \\
V
\end{array}
\end{array}
\end{align*}
\]

Merge has a last resort nature (Collins 2001: 4): before V and Obj merge, V and Obj have unsaturated (unvalued) features, and the external Merge must take place in order to saturate (value) some (or all) of the unsaturated (unvalued) features. Thus, V assigns a theta-role to Obj, but not the reverse: Obj does not assign a theta-role to V. The structural-Case feature of

\[29\] Kayne (1994) in fact suggests that obj raises to a higher position in OV-language.
Obj matches that of V, and V values the Case feature of Obj as [ACC], but not the reverse: Obj does not value V. V subcategorizes for a feature of Obj, i.e., a categorial feature [N], but not the reverse: Obj does not subcategorize V. V is the assigner, valuator, and subcategorizer. Let us call V as feature-evaluator. Let us use a bold line to indicate that a terminal is a feature evaluator. More accurately, the structures in (26) have the following structures.

\[(26')\]

\[
\begin{array}{c}
V \\
\text{Obj} \\
\text{Obj} \\
V
\end{array}
\]

Let us see how we can modify LCA without labels and projections. It is crucial to assume that "asymmetric c-command relation" here means two things.

\[(27)\] Elements of Asymmetric C-Command

(i) feature-asymmetric c-command, OR

(ii) structure-asymmetric c-command (containment version).

Crucially, feature-asymmetric c-command allows mutual c-command relation to show asymmetric nature. Recall that the original definition of LCA always excludes mutual c-command relation.

At the first external Merge, which always involves symmetric (mutual) c-command relation between two terminals, feature-asymmetric c-command is the crucial criterion for asymmetric property. In the case of (27i), a feature evaluator precedes a non-feature evaluator. We propose that the LCA should be modified as in the following.

\[(28)\] a. Linear Correspondence Axiom\(^{30}\)

\[\alpha \text{ precedes } \beta \text{ iff (a) } \alpha \text{ asymmetrically c-commands } \beta, \text{ or (b) } \gamma \text{ precedes } \beta \text{ and } \gamma \text{ dominates } \alpha.\]

b. Definition of Asymmetric C-Command (Revised)

\[\alpha \text{ asymmetrically c-commands } \beta \text{ iff }\]

(i) a terminal \(\alpha\) feature-asymmetrically c-commands a terminal \(\beta\), or (ii) \(\alpha\) structure-asymmetrically c-commands \(\beta\).

It is crucial that feature-asymmetric c-command relation holds between terminals only. We keep the modified definition of c-command using the notion of containment. In (29a) and (29b), since a terminal V feature-asymmetrically c-commands Obj, V precedes Obj.

---

\(^{30}\) We basically adopt the simpler version of LCA in Uriagereka (1998).
Symmetry Breaking and Economy in CML

(29) a. \[ \text{Tree with VO order} \] b. \[ \text{Tree with VO order} \]

For the first external Merge, either (29a) or (29b) can be the source of VO order. No complication as we have seen in (22a) exists in yielding VO order. V precedes O without V raising. This is consistent with the Economy principle since it is costless to have no raising. This conforms to evidence that V remains in some VO-language such as English (Lasnik 1995).

For internal Merge, in which an unsaturated lexical item merges with the topmost construct that is also unsaturated, the structure-asymmetric c-command becomes relevant. Consider the concrete structure as in the following.

(30)

The bold lines indicate feature evaluators. V feature-asymmetrically c-commands Obj, therefore V precedes Obj. There is no feature-asymmetric relation between T and (V, Obj). However, T structure-asymmetrically c-command V and Obj, therefore T precedes V and Obj. There is no feature-asymmetric relation between Subj and T. However, Subj structure-asymmetrically c-command T, therefore Subj precedes T. By transitivity, the linear order is <Subj, T, V, Obj>. Notice that Subj does not merge with T. Subj merges with T′, which does not exist any more in our system. Notice also that feature-asymmetric c-command relation holds between two terminals. Since T′ is not a terminal element, the order between Subj and T′ is not determined by feature-asymmetric c-command.

Let us see how the modified LCA without labels and projections explains VO and OV order. We propose that one of the possible structures for VO-language is the following.
Obj remains in situ. The EPP feature of the light verb v is satisfied and delete after Spell-Out. V is a feature evaluator: V assigns a theta-role to Obj and Subj, V checks and deletes the Case feature of Obj, and V subcategorizes for Obj. Thus, V precedes Obj. The light verb v structure-asymmetrically c-commands V and Obj. Therefore, v precedes V and Obj. Subj merges with the topmost node (T', which does not exist) after Merge of T. Subj satisfies the EPP feature of T'. T' is a feature evaluator for Subj: the phi feature of Subj matches that of T', and the Case feature of T' values the Case feature of Subj as [NOM]. Since Subj and T do not involve feature evaluation, the relevant relation is structure-asymmetric c-command. Subj structure-asymmetrically c-commands T. It follows that the order is <Subj, T, v, V, Obj>. We speculate that Subj and V may remain in situ in some VO-languages according to the relevant parameters. If Subj remains and V raises to v, VSO order is obtained\(^{31}\).

We propose that one of the possible phrase structures of OV-language is the following.

---

31) This explains why VO-language, say English, lacks scrambling. Suppose Obj scrambled, and adjoined to the topmost node. Scrambling takes place before Spell-Out since it has the phonetic effect. [EPP] of v is weak in VO-language by definition. After Spell-Out, EPP feature of Obj must lower to v. This lowering causes a violation of LF cyclicity (Watanabe 1995).
What is crucial is that Obj satisfies the EPP feature of the light verb v before Spell-Out in OV-language\(^{32}\). Subj and V may remain in situ according to the relevant parameters. If both Subj and V remain, the order OSV is obtained. If Subj remains and V raises to v, the order OVS is obtained. Thus, the crucial distinction between VO-language and OV-language is that the pied-piping of Obj must take place in the latter, but not in the former\(^{33}\).

What about the order between V and v? Consider the structure in detail\(^{34}\).

Suppose v1 is the original terminal before V-raising. Then, V merges with (adjoins to) v1,

---

32) Roughly, only 3% among 71 VO-languages have overt Case particles on Subj and Obj, but 75% among 55 OV-languages have them (Tsunoda (1991)). We assume that overt structural-Case particles found in OV-language is the reflect of the realization of the pied-piping to Case-checking position. The view is consistent with Chomsky (2001) that phonetic effects arise when uninterpretable features delete at Spell-Out. In (32), EPP features of v and T are satisfied and delete. Incidentally, 73% of 71 VO-languages shows overt wh-fronting, whereas 43% of 55 OV-languages shows overt wh-fronting (ibid.). We conjecture that the lower tendency of existence of overt wh fronting in OV-language is related to the fact that the successive (consecutive) pied-piping is relatively expensive.

33) This explains why OV-language, say Japanese, has scrambling. Suppose EPP feature of v is satisfied before Spell-Out in OV-language. Obj scrambles for some kind of focus-feature checking before Spell-Out. There is no violation of cyclicity condition after Spell-Out since EPP feature of v has already been satisfied. Thus, scrambling is possible.

34) According to BPS (Chomsky 1994a, 1995a: 248), adjunction structures (L) are defined as L= \{<H(K), H(K)>, \{□, K\}\}, in which □ is adjoined to K. <H(K), H(K)> is the label for the structure. If there were no label, the adjunction structures cannot be defined in this way, as Collins (2001: 8) states. See Chomsky (2001; 17–25) for a new way to look at the issue of adjunction.
but not with \((\text{to}) v2\). \(v2\) is the new segment which is created by adjunction. Suppose that adjunction is just another case of Merge. More accurately, (33) should be as in the following without the label \((= v2)\).

\[
(33')
\]

In \((33')\), the light verb \(v1\) is the selector of theta-assigning domain of \(V\). Suppose \(v1\) is the valuator for \(V\).

\[
(33'')
\]

\((33'')\) yields the order \(<v1, V>\). Suppose \(V\) is the valuator for \(v1\).

\[
(33''')
\]

\((33''')\) yields the order \(<V, v1>\). We eliminate the adjunction operation.

Chomsky (2001: 8) argues that one of the complications LCA brings into the theory is those unmotivated movements of lexical items (pied-piping) in order to produce correct linearity. However, the movements in (32) are simple and are motivated by uninterpretable-feature checking. The analysis is consistent with the evidence that some VO-language lacks V-raising, and that some OV-language has V-raising\(^{35}\). We argue against remnant-VP movement into [Spec, IP] for OV-language, in which \(V\) first raises out of VP, and the remnant VP containing Subj and Obj raises to [Spec, IP] (Whitman (1997)). This kind of remnant VP movement is unmotivated or ad-hoc as Chomsky (2001; 8) alludes to.

With the extended notion of asymmetric c-command incorporating the notion of feature-asymmetric c-command in addition to the conventional structure-asymmetric c-command, the LCA is modified in a simple and natural manner without losing their explanatory powers. The notions such as label and projection are also eliminated. We also suggested the possibility of eliminating adjunction.

1.9. Polarization Axiom

Consider external Merge of terminals. The iteration of external Merge forming head-complement relation gives rise to the following four possible phrase structures.

---

(34) a. 

\[ \ldots \varepsilon \] 

[Diagram of L + L + L + L + ...]

b. 

\[ \ldots \varepsilon \] 

[Diagram of D + D + D + D + ...]

c. 

\[ L + D + L + D + \ldots \] 

[Diagram of L + D + L + D + ...]

d. 

\[ D + L + D + L + \ldots \] 

[Diagram of D + L + D + L + ...]

(34a) & (34b) and (34c) & (34d) are chiral in 2-D, respectively. All four cases yield the same ordering: \(< \varepsilon, \delta, \gamma, \beta, \alpha>\), which shows that LCA is a symmetric principle. Kayne (1994) assumes that (34b) is chosen for natural language, but the reason is not clear there. Is there any way we can deduce this from more general principle? We argue that the structures in (34c) and (34d) are excluded by the Economy principle: they are costly. Simply, consistent l-type effect (left branching) with D-PS or consistent d-type effect (right branching) with L-PS is cheaper than mixing l-type and d-type effects\(^{36}\). We propose the following axiom, which is related to the Economy principle.

---

36) This sort of \(D\) vs. \(l\)-mismatch between the structure of an individual molecule and the overall effect with respect to polarization of light is a general aspect of chemical systems. For example, Glycerol-aldehyde exhibits a D-type molecule structure, but the products made out of it have an l-type effect, e.g., glyceric acid, lactic acid, and tartaric acid exhibit l-type effects with respect to polarization of light.
(35) **Polarization Axiom**

The biological system shows either $l$-type effect or $d$-type effect.

Thus, we conclude that the mixed situation of $l$- and $d$-type effects is costly for *external Merge*. External Merge of terminals is costless, and it cannot tolerate this costly situation\(^{37}\). We do not understand at this point what determines (34b) over (34a). Consider internal Merge.

(36) a. $$\begin{array}{c}
\delta \\
\downarrow \\
\alpha \\
\downarrow \\
\beta \\
\downarrow \\
\gamma \\
\downarrow \\
\alpha \\
\end{array}$$

L + L + L + L +...

b. $$\begin{array}{c}
\delta \\
\downarrow \\
\delta \\
\downarrow \\
\beta \\
\downarrow \\
\gamma \\
\downarrow \\
\alpha \\
\end{array}$$

D + D + D + D +...

c. $$\begin{array}{c}
\delta \\
\downarrow \\
\delta \\
\downarrow \\
\beta \\
\downarrow \\
\gamma \\
\downarrow \\
\alpha \\
\downarrow \\
\beta \\
\end{array}$$

L + D + L + D +...

d. $$\begin{array}{c}
\delta \\
\downarrow \\
\delta \\
\downarrow \\
\beta \\
\downarrow \\
\gamma \\
\downarrow \\
\beta \\
\downarrow \\
\alpha \\
\end{array}$$

D + L + D + L +...

All four structures yield either of the following order: $<\alpha, \delta, \gamma, \beta>$ if the copy $\alpha$ is spelled out, or $<\delta, \gamma, \beta, \alpha>$ if the original $\alpha$ is spelled out. From PA, however, (36a) and (36b) are excluded. Either (36c) or (36d) should be the possible structures for natural language as a biological effect. Why do (36c) and (36d) contain the mixed type of $D$ and $L$? Notice that $\gamma$ is a

\(^{37}\) Both external Merge and internal Merge involve the last resort nature (Collins 2001). We argue that there is a distinction: external Merge is relatively inexpensive.
specifier of $\beta$, and that $\alpha$ is forced to raise to eliminate uninterpretable features. We propose that external Merge of specifier and internal Merge are costly, and that the relative high cost is reflected by the mixed Merge of D-PS and L-PS. We argue that these cases involve erasure of uninterpretable features. In other words, the higher degree of energy of uninterpretable features guarantees the mixed employment of D-PS and L-PS in order to obtain consistent d-type effect or l-type effect of overall structural configuration.

1.10. Symmetry as Unstable State
What does it mean specifically to say that the symmetric structure is unstable and costly? We propose the head-complement relation is too symmetric to exist if the relation retains unchecked uninterpretable features. Consider the passive$^{38}$.

\begin{center}
(37)
\end{center}

\begin{center}
\begin{tikzpicture}
  \node[anchor=east] (T) at (0,0) {T};
  \node[anchor=west] (TP) at (1,0) {TP};
  \node[anchor=west] (VP) at (1,-1) {VP};
  \node[anchor=west] (was) at (0,-1) {was};
  \node[anchor=west] (kissed) at (0,-2) {kissed};
  \node[anchor=west] (John-CASE) at (1,-2) {John-CASE};
  \node[anchor=west] (FF) at (1,-3) {FF};

  \draw (T) -- (TP);
  \draw (T) -- (VP);
  \draw (VP) -- (was);
  \draw (VP) -- (kissed);
  \draw (kissed) -- (John-CASE);
  \draw (John-CASE) -- (FF);

  \node at (0,-0.5) {$[-\text{Int } \Phi]$};
  \node at (1,-0.5) {$[\text{Int } \Phi]$};
  \node at (0,-2.5) {$[\text{Int } \Phi]$};
  \node at (1,-2.5) {$[-\text{Int } FF]$};
\end{tikzpicture}
\end{center}

The structure (37) is unstable and costly since it contains two branching structures which retain unchecked uninterpretable features$^{39}$. In other words, the head-complement relation can exist (contra Kayne (1994) and pro Chomsky (1995a/b)) only if there is no unchecked uninterpretable feature is involved at the final stage. Symmetry breaking is forced by one of the Legibility Conditions: *Erase uninterpretable features as soon as possible.* Thus, unchecked uninterpretable features are symmetry breakers. Copy is the operation which breaks structural symmetry. By application of Copy, the complement John-CASE of the verb kissed merges with the uppermost TP in (37), breaking the symmetric relation.

---

38) We adopt the basic line of the analysis of passive in Nunes (2001). We come back to some questions concerning his approach in later section.

39) $[\Phi]$ of T is not interpretable because it does not contribute to the semantic interpretation. Consider the following.

(i) Mary walks.
In (i), the morpheme s is $[\Phi]$ of T: 3rd person, singular, and masculine. Adding the morpheme s does not change the substantial meaning of walk. If a morpheme does not change the meaning, it is not interpretable since it lacks semantic features. Similarly, structural Cases such as the nominative or accusative Case do not contribute to the meaning of a sentence. Consider the following.

(ii) Mary considers that he is a fool

(iii) Mary considers him to be a fool
Both in (ii) and (iii), he and him indicate the same meaning, i.e., the subject of the predicate be a fool. Thus, the difference of the structural Case does not affect the meaning, which means that they lack semantic features.
Φ of John and T match and agree. As a result of Match and Agree, uninterpretable features of CASE (now case being invisible at PF; CASE being visible at PF before Copy) and T are erased. Finally, John-CASE, which is the complement of kissed, and its uninterpretable feature are deleted for convergence at PF. The output ordering is <John-case, was, kissed>.

1.11. Information and Point-Asymmetric Nature

Let us consider the linear order at 1-D. Does a linear order have a mirror image? It does, as in the following.

(38) \(<a, b, c, d> \cdot <d, c, b, a> \rightarrow \text{asymmetric}\)

1-D mirror

In a 1-D world, a mirror is a point. A string \(<a, b, c, d>\) is asymmetric (chiral): the two strings \(<a, b, c, d>\) and \(<d, c, b, a>\) do not overlap each other by moving one to the other at 1-D. In other words, there is no point-axis in the strings as \(<a, b, c, d>\) or \(<d, c, b, a>\) which produces overlapping of the former half and the latter half by folding the string at the point-axis. What kind of string is symmetric? If we want the output of LCA to be symmetric (achiral), we would have to deal with the symmetric string as, say, \(<a, b, a>, <a, b, c, c, b, a>, <b, b, b, b>, \text{or} <a, a, a,a>^{40}\).

(38') \(<a, a, a, a> \cdot <a, a, a, a> \rightarrow \text{symmetric}\)

1-D mirror

40) Jenkins (2000: 162) cites Stewart (1998: 57), which tells us what happens if Morse code were highly symmetric, e.g., assigning "the same letter (S, say) to every sequence of dots and dashes." Stewart says that it is "a totally useless one." Stewart explains what he means by symmetric code as follows. "For symmetries of codes, the relevant transformations are not motions in space; they are operations that swap sequences of code symbols around. A symbol sequence possesses such a symmetry if its meaning is unchanged by the swap."
The highly symmetric `<a, a, a, a>`, however, is useless and non-economical information. For example, the string `<a, a, a, a>` is less economical since there is more economical string, namely, `<a>`. In the same vein, `<a, b, c, c, b, a>` is less economical since there is more economical one `<a, b, c>`. Thus, information as `<a, a, a, a>` is in violation of Economy principle (Minimal Energy). Symmetric strings are costly. Suppose we still deal with a clause with information as `<a, a, a, a>` or `<a, b, b, a>`. Then the relevant examples of clauses would be as follows.

(39) a.  
    John  
    eats  
    eats  
    eats  
    eats  
    bread  

b.  
    John  
    eats  
    bread  
    bread  
    eats  

These structures give rise to the following outputs, which exhibit a less economical way of conveying information.

(40)  
    a. *John eats eats eats eats bread.  
    b. *John eats bread bread eats.  

The working of LCA is symmetric in the sense that the same linear order is obtained out of two mirror-image-structures L-PS and D-PS (2-D-asymmetric and 3-D-symmetric). However, the output linear order is asymmetric: symmetric linear order is excluded by Economy. Suppose `<a>` is some kind of focus information. There are at least two possible ways to place `<a>` in a string. `<a>` is either at the beginning or at the end of a sentence. This is forced by one of the Legibility Conditions imposed to C\textsubscript{HL}: \textit{Place focus on the place easily detected}.

(41)  
    a. `<a, b>` (focus at the beginning)  
    b. `<b, a>` (focus at the end)  

The first type involves the nominative Case particle `ga`, indefinite article `a`, cleft, passive,
and scrambling, among others. The second type involves the topic marker *wa*, definite article *the*, heavy NP shift, and transitive-expletive, among others.

(42)  \(<a, b>\) type
   a. *The book* is here
   b. It is *this book* that I bought
   c. *John* was kissed
   d. *Many men* entered the room
   e. *watasi-ga* hannin da
      *I-nom* criminal *copula*
      ‘It is I who is a criminal.’
   f. *John-ga* Mary-ni ais-are-te-i-ru
      *John-nom* Mary-*by* love-passive-to-be-nonpast
      ‘John is loved by Mary’
   g. *John-o* Mary-ga suki-da
      *John-acc* Mary-nom *like copula*
      ‘(lit.) John, Mary likes him.’

(42')  \(<b, a>\) type
   a. There is a *book*
   b. John gave to Bill *all his old linguistics books* (Kayne (1994: 729))
   c. There entered the room *many men* (Chomsky (2001))
   d. *watasi-wa* hannin da
      *I-top* criminal *copula*
      ‘I am the criminal in question (whom you are looking for).’
   e. *watasi-ga* katta-no-wa *kono-hon da*
      *I-nom bought that-top this book copula*
      ‘It is this book that I bought.’
   f. Mary-*wa* *John-o* ais-ite i-ru
      *Mary-top John-acc* love-to-be-nonpast
      ‘As for Mary, she loves John.’

In (42), a focus is put at the relative beginning of the sentence, whereas in (42'), it is relatively at the end. Let us consider active-passive transformation. The active-passive transformation is symmetric with respect to propositional meaning, while it is asymmetric with respect to

\[41\] We need to assume that the passive morpheme *nare* is a member of Infl set, which is related to modality, and that VP as \(\theta\)-assigning domain in a strict sense corresponds to the proposition.

The situation is more like the fact that “the genetic code is redundant” in that “different triplets often code for the same amino acid,” ... “a definite degree of symmetry - albeit imperfect... For example, GA? is always leucine, and CG? always arginine. In short, the code for these amino acids is symmetric under changes of the third base. If this symmetry were perfect, then the 64 (=4\(^3\)) triplets would break up into 16 (=4\(^2\)) quartet triplets, such as GAC, GAG, GAA, GAT, with each triplet of the quartet coding for the same amino acid (but a different amino acid for
time slot (linear ordering) and focus information.\footnote{We will argue in the next section that active-passive transformation has “imperfect” symmetric property.}

<table>
<thead>
<tr>
<th>active/passive transformation</th>
<th>w. r. t. proposition</th>
<th>w. r. t. focus</th>
<th>w. r. t. time slot</th>
</tr>
</thead>
<tbody>
<tr>
<td>symmetric</td>
<td>asymmetric</td>
<td>asymmetric</td>
<td>asymmetric</td>
</tr>
</tbody>
</table>

In a sense, free word order is more symmetric, in that swapping the order does not affect the meaning, whereas fixed word order is less symmetric, in that changing the order radically affects the interpretation.\footnote{Fukui & Takano (2000) argue that free-word property is related to overt Case particles and no V-raising, while fixed-word property, covert Case particles and V-raising.} For example, if nothing changes in the interpretation from scrambling, the scrambling operation has a symmetric property. In the next section, we argue that active-passive and scrambling transformations possess “imperfect” symmetric property.

1.12. “Imperfect” Symmetry in Passive and Scrambling

Assume Copy theory of movement. The structures before and after passive and scrambling operations are the following.

(44) before

\[
\begin{tikzpicture}
  \node (a) at (0,0) {$\alpha$};
  \node (b) at (1,-1) {$\beta$};
  \draw (a) -- (b);
\end{tikzpicture}
\]

(44') after

\[
\begin{tikzpicture}
  \node (b) at (0,0) {$\beta$};
  \node (a) at (-1,-1) {$\alpha$};
  \node (c) at (-1,-2) {copy};
  \node (d) at (1,-2) {original};
  \draw (a) -- (b);
\end{tikzpicture}
\]

The 1-D string $<\alpha, \beta>$ in (44) is asymmetric at 1-D, whereas the 1-D string $<\beta, \alpha, \beta>$ in (44') is symmetric in 1-D. We propose that the symmetric nature with respect to propositional meaning in active-passive and scrambling is guaranteed by this symmetric property. To obtain
asymmetry, the original must somehow become inactive both phonetically and syntactically. Thus, the crucial property of displacement (Move) in natural language has two aspects: (i) to obtain symmetry which is realized by Copy, and (ii) to obtain asymmetry which is realized by Feature-Checking (& Eliminating).

Jenkins (2000: 163) argues that symmetry breaking is involved in language acquisition: “the initial state (of the cognitive component) of the language faculty allows perfect symmetry with respect to word order. Then as we make transitions from one cognitive state to another: S₀, S₁, S₂, ... Sₙ, the perfect symmetry is broken, resulting in the word order found in the language learned⁴⁴.” We propose further that symmetry breaking is involved not only in the process of language acquisition (the process from the initial state to the final state), but also in the derivational process of structure building in the computational system of natural language at the final state itself.

2. Empirical Discussion

2.1. Passive ~ Interaction of LCA, Economy (Minimize Energy) and Symmetry Breaking ~ Nunes (2001) asks: given copy theory, why is it that Cᵥ prohibits (45a) and (45b), allowing only (45c) as a legitimate sentence?

(45)   a. *John was kissed John.
       c. John was kissed.

In (45a), both the copy John at the beginning and the original John at the end remain. In (45b), the original remains, and the copy is deleted. In (45c), the copy remains and the original is deleted. Suppose the derivation has reached the following stage.

(46)

44) Jenkins emphasizes that “UG (=the initial state of the natural language system) for humans is maximally symmetric with respect to possible word orders, possibly apart from a preferred basic word order, which results from the necessity of linearizing spoken language. All other combinations of word order result from symmetry-breaking (linguistic input being the symmetry breaker) (ibid. 164).” He also says that “word order types would be the (asymmetric) stable solutions of the symmetric still-to-be-discovered "equations "governing word order distribution (ibid., 164).” The still-to-be-discovered “equation” might be some version of LCA.
Symmetry Breaking and Economy in C\textit{nl.}

\([-\text{Int } \Phi]\) include uninterpretable \(\Phi\)-features of \(T\) (person, number, and gender). \([\text{Int } \Phi]\) include interpretable \(\Phi\)-features of the nominal \textit{John}. \([-\text{Int FF}]\) include uninterpretable formal features. Uninterpretable features must be eliminated as soon as possible in the derivation, since these uninterpretable features are not interpreted anywhere in C\textit{nl} in the human brain. Every feature has to be interpreted in the brain. In order to eliminate uninterpretable features, \textit{John-CASE} is copied, and the copy merges with TP.

\[(46')\]

\[
\begin{array}{c}
\text{TP} \\
\text{John-case} \\
[\text{Int } \Phi] \quad [-\text{Int FF}] \\
\text{T'} \\
\text{T} \\
[\text{Int } \Phi] \\
\text{was} \\
\text{VP} \\
\text{kissed} \\
\text{John-CASE} \\
[\text{Int } \Phi] \quad [-\text{Int FF}] \\
\end{array}
\]

The original is \textit{John-CASE}. \textit{John-CASE} is copied, and the copy merges with TP (now \(T'\)). \(\Phi\) of \textit{John} and \(\Phi\) of \(T\) match and agree, and uninterpretable features are deleted. \textit{John-CASE} becomes \textit{John-case}, which means that \([-\text{Int FF}]\) is deleted, and so is now invisible at PF. If \([-\text{Int FF}]\) were visible at PF, the derivation would crash at PF since \([-\text{Int FF}]\) is not interpreted at PF. If Spell-Out sends phonetic information to PF at this point, the example (45a) is produced. But what's wrong with (45a)? From the economy perspective, since the deletion is minimum, the derivation is costless, and it is expected that the derivation should converge. However, (45a) is excluded by LCA. In (45a), for example, \textit{John} asymmetrically c-commands \textit{was}, and \textit{was} asymmetrically c-commands \textit{John}. It follows that \textit{John} must precede and follow \textit{was}. If C\textit{nl} were to allow such a situation, the solution of C\textit{nl} would be very symmetric, in that the asymmetric c-command calculation makes no difference. It is more like the situation in which the equation \(y=x^2\) makes no difference whether we have \(y=x\) or \(y=-x\). But such symmetry is prohibited by LCA in natural language computation. This situation violates the asymmetry condition imposed by LCA. Put differently, keeping both the original and the copy of a chain yields too symmetric a situation, which is unstable and too costly for stabilization and maintenance, so it is not economical. Symmetry breaking is forced by the economy principle. At the same time, such a symmetrical situation violates the irreflexivity condition, which prohibits \(\alpha\) from preceding and following itself. In (45a), \textit{John} precedes and follows itself. This situation is again very symmetric: preceding something and following something make no difference. In sum, LCA tries to exclude symmetric structures, which are unstable and expensive. So either the copy or the original must be deleted. Suppose the copy were deleted as in (45b). The uninterpretable FF of \textit{John-CASE} is still visible at PF. Therefore,
we have to delete \([-\text{Int FF}].\) Delete applies twice. Suppose the original were deleted as in (45c). No uninterpretable feature remains. Therefore, Delete applies once. (45c) is cheaper since there are fewer applications of Delete. Therefore, deleting the original is chosen by the Economy principle, yielding (45c).

2.2. Problems

First of all, Nunes is right in saying that (45a) is excluded by LCA, but he is incorrect in saying that (45a) is economical: (45a) is not economical from the perspective of symmetry breaking. (45a) exhibits a very symmetric state. Consider the relation of John and was. John asymmetrically c-commands was, and at the same time, John does not asymmetrically c-command was. Given LCA, John precedes was, but at the same time, John follows was. The situation is symmetric in that asymmetric c-command does not contribute to linear ordering production. Notice that a string \(<a, b, a>\) is symmetric. A symmetric state is costly. It follows that (45a) is costly\(^45\).

Secondly, Nune’s solution contains a global economy, which we want to get rid of in order to avoid wild computational complexity. A global economy states: Compare a number of derivations, and then choose the cheapest one among them. The number of possible derivations becomes infinite, leading to wild computational complexity. Given that a child can acquire any human language effortlessly in a short period, the basic principles of CIL must be very simple.

2.3. Solutions

For the first problem, there is nothing wrong in saying that (45a) is excluded since it is costly.

For the second problem, we propose that deleting the original is chosen because of the principle of local economy: Erase uninterpretable features as soon as possible whenever you can. Since it is the original, not the copy, that retains the uninterpretable feature, the original must be deleted whenever possible. Immediately after the application of Copy, at least the phonetic features and the uninterpretable feature of the original are deleted promptly. No comparing of the number of applications of Delete is necessary.


There is an interesting contrast with respect to the order of a verb and a negation between French and English. In English, the main verb must not precede the negation.

(47) a. *John likes not Mary
    b. John does not like Mary

In French, on the contrary, the main verb must precede the negation\(^46\).

\(^{45}\) One might be tempted to conclude that LCA is deducible from symmetry breaking. I leave this for future research.

\(^{46}\) Historically, \(ne\) appeared before \(pas\). However, in modern colloquial French, \(ne\) is often deleted.
(47') a. Jean (n')aime pas Marie
   b. *Jean (ne) pas aime Marie

How can we account for the difference? Lasnik claims that an item which is used in human language computation is either “lexicalist (= featural)” or “bare (= affixal).” An item is featural if it is a bundle of abstract features that have to be checked. Featural items are called “lexicalist,” in that “they aren’t constructed syntactically or in any other way; rather, they’re introduced (from the Lexicon) with all their phi-features on them, which they later check” (Lasnik 1995; 188). On the other hand, an item is bare if an affix attaches to the bare item, and the attachment demands linear adjacency. He claims that the attachment is not “a true syntactic operation, but a sort of “interface” operation between syntax and morphology” (ibid; 192).

Let us describe the distinction picturesquely. A featural item consists of more active features (like molecules of gas in higher temperature), which have to be checked and deleted, whereas an affixal item is more solid (like molecules of solid matter in lower temperature), in that the features are inactive. In a sense, unchecked uninterpretable features consisting of featural items have a greater degree of entropy, more energy, more information, and a higher degree of symmetry breaking, which is realized by feature-movement and checking. Therefore, once they become active, the system turns to a more stable and economical state. On the other hand, inactive features consisting of affixal items have a low degree of entropy, less energy, less information (related to the PF side, not LF), and a low degree of symmetry breaking, which is shown by the requirement of adjacency: what are involved are two items which have a very local (near-symmetric) relation.

Takano (1998) and Fukui & Takano (2000) provide the image of the machinery. Active features of featural items are eliminated by checking operations, whereas inactive features of affixal items are eliminated by Spell-Out: the relevant feature “gets eliminated from the derivation going to λ (a linguistic expression at LF) when Spell-Out strips away phonological features from the structure Σ” (Fukui & Takano 2000; 244). Thus, inactive features of an affixal item that is eliminated by Spell-Out must be “linked with phonological features” (ibid; 244), which means that the process of affixal item attachment is taken to the PF side.

The relevant stages of operation for featural and affixal items are schematized in the following. WI stands for Word Interpretation, which is related to the morphological process.

We take this tendency of PF-deletion seriously, and assume that pas, like not in English, is the head of Negation. I am indebted to Raux Yamasaki Annie (St. Andrew’s University) for the judgments of French.
Given this much, let us consider how the hybrid model explains the data. Lasnik argues that verbs in French are featural, requiring checking operation, while in English, main verbs are bare and the auxiliaries are featural. T (=Infl) in both languages is either featural or affixal. Let us repeat the data.

(49)  
a. *John likes not Mary  
b. John does not like Mary  
c. Jean (n’aime pas Marie  
d. *Jean (ne) pas aime Marie

Consider (49a). Lasnik stipulates that the English main verb is bare, i.e., its formal features are inactive and they are eliminated by Spell-Out. A bare verb must amalgamate with the relevant affix, and the amalgamation should observe the linear adjacency. In (49a), the bare verb like tries to attach to the affix s in WI. The relevant derivation is the following.

(50)  
\[
\text{TP} \quad \text{s} \quad \text{NegP} \quad \text{not} \quad \text{VP} \quad \text{like} \quad \ldots
\]

The linear adjacency between the affix s and the main verb like is not obtained due to the existence of the intervening negation not. As the consequence, the affix and the verb cannot be amalgamated. In (49a), the verb attaches to the affix in violation of adjacency requirement. The derivation crashes. Similarly, we obtain the unacceptable *John not likes Mary, in which the affix lowers for amalgamation, again in violation of adjacency. Consider (49b).

(49)  
b. John does not like Mary
Symmetry Breaking and Economy in Cal.

The bare verb must remain in situ due to the lack of adjacency between the affix and the verb. Because the affix s cannot survive in that form if it’s stranded, the affix must be supported by a light verb do, eventually realized as does. The do-support is the last resort.

\[(50')\]

\[
\begin{array}{c}
\text{TP} \\
\text{s} \\
\text{NegP} \\
\text{do} \\
\text{not} \\
\text{VP} \\
\text{like} \\
\end{array}
\]

Consider (49c).

(49)  c. Jean (n’)aime pas Marie

Lasnik stipulates that the French main V is featural, i.e., its formal features are active, in that the checking operation before Spell-Out must eliminate them. First, the relevant feature [F] of the main V adjoins to T, causing the elimination of [F] by checking followed by matching and agreement, and then the bundle of features that consists of the main V is attracted to T (Copy or Pied-Piping).

\[(50'')\]

\[
\begin{array}{c}
\text{TP} \\
\text{T} \\
\text{NegP} \\
\text{aime} \\
\text{T} \\
\text{[F]} \\
\text{pas} \\
\text{T} \\
\text{[F]} \\
\text{[F]} \\
\text{(aime)} \\
\end{array}
\]

The linear adjacency is irrelevant to a movement which is driven by feature-movement. In fact, linear adjacency is crucial in PF interface. In the syntactic process leading to LF interface, what is crucial is that Attract F always seeks the closest relevant feature. What matters is the closest relevant feature. The intervening negation pas does not bear the relevant feature. First, T attracts [F] (the step \(\text{①}\)), and then the pied-piping of the verb follows (the step \(\text{②}\)). Pied-piping is not concerned with adjacency. The derivation converges. Consider (49d).

(49)  d. *Jean (ne) pas aime Marie
In this case, featural main V stays in situ. Since there is no need for the active formal feature [F] to adjoin to T for pied-piping, [F] remains within VP.

(51)

\[
\text{TP} \\
\text{T} [F] \\
\text{NegP} \\
\text{pas} \\
\text{VP} \\
\text{aime [F]} \\
\ldots
\]

As a result, the uninterpretable feature [F] of T and that of the verb are not eliminated, causing the derivation to crash.

2.5. Hybrid Model vs. Two-Chain Hypothesis

The hybrid model is compatible with the two-chain hypothesis (Chomsky 1995b, Ochi 1999). Consider the lexicalist case first (French (49c)).

(49) c. Jean (n')aime pas Marie

*Attract* forms CHF, in which the target (T in this case) attracts the closest relevant formal feature FF before Spell-Out. The relevant derivation is the following*’).

(52)

\[
\text{T} \\
\ldots \{\text{FF}\} \\
\text{Overt Attraction} \\
\text{V} \\
\ldots
\]

Assume that V = {pf, sf} is an illegitimate object at PF, i.e., PF operations still need information about formal features (p.c. by Željko Bošković in Ochi (ibid., 27)). V = {pf, sf} undergoes generalized pied-piping (GP) to avoid a crash at PF (*Greed*): V and T get together, and V recovers [FF]. GP forms CHCAT. The target-based MLC is silent for GP.

47) pf= phonetic features, sf= semantic features, FF= formal features.
V = \{\text{pf}, \text{sf}\} \text{ and } \{\text{FF}\}\) are close enough to repair V. This accounts for the French main verb paradigm. Notice that \{\text{FF}\} is not completely erased, i.e., if there is no FF at all when GP applies, the repair strategy would not work. It follows that FF is only \textit{deleted}, not totally \textit{erased}; it somehow remains. Let us introduce the definition of the Law of Conservation in natural language (Uriagereka (1998: 536)).

(54) \textbf{Law of Conservation}

(i) No operation can eliminate derivational terms.

(ii) All interpretable features that are present in the lexical array are present at LF.

The following is the definition of \textit{term} (ibid., 536).

(55) \textbf{Definition of term}

Where (a), (b), and (c) are syntactic objects — (a) an FFM [Formal feature matrix]; (b) a labeled lexical item; (c) a phrase marker \(K = \{L, \{\alpha, \beta\}\}\), where \(\alpha, \beta\) are syntactic objects and \(L\) is the label of \(K\) and where \(K\) is a root phrase marker — then (i) \(K\) is a term, or (ii) if \(P\) is a \textit{term}, the members of the members of \(P\) are \textit{terms}.

We suspect that the definition of Law of Conservation for natural language in (54) may be incomplete: not only interpretable features but also uninterpretable features somehow remain. The Law of Conservation in (54) should be modified as follows.

(56) \textbf{Law of Conservation} (modified)

(i) No operation can eliminate derivational terms.

(ii) \textit{All features} that are present in the lexical array are present at LF.

In fact, the \textit{First Law of Thermodynamics}, namely, the \textit{Law of Conservation of Energy} (LCE, or \textit{Minimal Energy}) in physics should not allow total erasure. LCE is a perfectly symmetric principle. LCE is associated with the invariance of time (Jenkins (2000: 154). Consider next the
bare case (English (49b)).

(49) b. John does not like Mary

*Attract* forms CH\textsubscript{FF} after Spell-Out. GP does not operate, since there is no need to repair the remnant element for the purpose of PF convergence: PF convergence is irrelevant at LF.

(57)

Since there is no need for \( V \) to raise to \( T \), it must not raise. This is forced by Economy. Thus, under the Two-Chain Hypothesis, \( V \) remains in situ not because \( \text{NEG} \) blocks adjacency relation of \( T \) and \( V \), but because there is no need for \( V \) to undergo GP. Thus, the *Two-Chain Hypothesis* is simpler than the *Hybrid Model*, in that the former does not stipulate WI component in which the bare \( V \) and \( T \) are amalgamated, respecting the adjacency requirement. The final step is the PF realization of \{FF\} using a light verb *do*.

(58)

2.6. English Imperative

Lasnik (1995) provides further evidence for his hypothesis. The evidence comes from the English imperative.

(59) Do not eat
English main verb is bare, so it must remain in situ. Since the imperative affix is stranded, it is phonologically realized as *do.*

![Diagram](image)

The *Two-Chain Hypothesis* offers simpler account. At LF, it is not necessary that the main verb *eat* raises. Therefore, V does not raise. To avoid *[FF]* stranding, the light verb *do* is inserted as a last resort.

### 2.7. Imperative in Old Japanese and Modern Japanese

In Old Japanese, there were two types of negative imperative.

(61)  

a. na *tabe* so  
   *not eat do*  
   ‘Do not eat’  

b. *tabe-ru* na  
   *eat-nonpast not*  
   ‘Do not eat’

Only the form (61b) has survived in Modern Japanese. Consider (61a). Assume LCA. Since we get the adjacency effect, i.e., *tabe na so,* the verb must be bare remaining in situ, as in English. We argue that the sentence particle *so,* which is thought to be derived from the verb *su ‘do,’* is a phonological realization of the affix generated as the light verb v, which takes VP as the complement. The phonological realization of the affix *so* is expected if we assume that the elimination of inactive features of affixal item is executed by Spell-Out, forcing the derivation to flow into the PF side. The bare verb *tabe* and the affix *so* are amalgamated since there is no intervening negation, observing the linear adjacency.
(62)

Let us consider (61b), the form which has survived up to Modern Japanese. In this case, we propose that the main verb is featural, as in French. The active formal feature [F] of the main verb first adjoins to T, and [F] is eliminated. The entire bundle of features then raises to T. The linear adjacency is irrelevant to a movement that is forced by feature checking.

(63)

The uninterpretable formal features [F] are eliminated. The derivation converges. It follows that the main verb in Modern Japanese is featural, and that the overt V-raising is forced by the checking operation, by which uninterpretable active features are eliminated\(^{48}\). In Old Japanese, both bare and featural properties of V coexisted. In Modern Japanese, the bare property has disappeared.

2.8. Past and Perfective Morphemes in Japanese

In English, the featural/bare distinction is observed within the same language. The main verb is bare, whereas the auxiliary verb is featural. Another piece of evidence for the featural/

\(^{48}\) The conclusion supports the hypothesis proposed in Koizumi (1995) and Nakajima (1999) that the main verb in Modern Japanese undergoes overt raising. If V raises to T in Japanese, Object must raise higher than T. Otherwise, LCA cannot yield the SOV order for Japanese. In this connection, Whitman argues for remnant-clausal-phrase movement into Specs of IP and CP. After V-raising to I, the remnant VP raises to [Spec, IP]. After the [V+I] amalgam raising to C, the remnant IP raises to [Spec, CP]. This explains why Japanese lacks overt wh-movement: [Spec, CP] is overtly occupied. We have argued against this view in Section 1.8.
bare distinction comes from Modern Japanese.

\begin{itemize}
\item[(64)] a. A: moo gohan tabe-ta?
\begin{itemize}
\item\textit{already meal eat-past}
\item\textquoteleft Have you finished your meal?\textquoteright \\
\item B: iya, mada tabe-te-i-na-i
\begin{itemize}
\item\textit{no yet eat-perfective-be-not-nonpast}
\item\textquoteleft No, I haven't eaten yet.\textquoteright \\
\end{itemize}
\item C: *iya, mada tabe-na-katta
\begin{itemize}
\item\textit{no eat-not-past}
\item\textquoteleft (lit.) No, I didn't yet.\textquoteright \\
\end{itemize}
\end{itemize}
\item b. A: kinoo gohan tabe-ta?
\begin{itemize}
\item\textit{yesterday meal eat-past}
\item\textquoteleft Did you eat yesterday?\textquoteright \\
\item B: ? iya, tabe-te-i-na-i\textsuperscript{49}
\begin{itemize}
\item\textit{no eat-perfective-be-not-nonpast}
\item\textquoteleft No, I have not.\textquoteright \\
\end{itemize}
\item C: iya, tabe-na-katta
\begin{itemize}
\item\textit{no eat-not-past}
\item\textquoteleft No, I didn't\textquoteright \\
\end{itemize}
\end{itemize}
\end{itemize}

In (64a), when the main V is followed by the perfective morpheme \textit{ta}, the negative answer is \textit{perfective-not}. In (64b), when the main V is followed by the past morpheme \textit{ta}, the negative answer is \textit{not-past}. How can we account for the contrast? We argue that perfective \textit{ta} is featural, whereas past \textit{ta} is affixal. The derivation of (64a-B) is the following. (64a-B) has a biclausal structure.

\textsuperscript{49} In (64b-A), \textit{ta} is a past-tense morpheme. This question asks whether there is a point in the past at which eating took place. The tense of (64b-B) is present tense. Why is this mismatch permitted? Notice that this kind of temporal mismatch is barred in (64a-C). We propose that present tense can include a point in the past, but past tense cannot contain the speech time (cf. Enç (1987)). The past-tense \textit{katta} in (64a-C) cannot contain the speech time represented by perfective \textit{ta} in (64a-A), thereby excluding the former as a mismatched answer. On the other hand, the present tense \textit{i} in (64b-B) can contain the past time represented by past \textit{ta} in (64b-A), thereby avoiding the mismatch.
When the upper TP is created, the closer active formal feature [F1] of the perfective morpheme *te* is attracted to the upper T, eliminating [F1]. The bundle of features of *te* then is attracted to the upper T by GP. The next closer active formal feature [F2] of the main V *tabe* is attracted to the upper T, eliminating [F2]. Then, the bundle of features of *tabe* is attracted to the upper T by GP. All uninterpretable features are eliminated. LCA calculates that the output order is *tabe-te-i-na-i*. The derivation converges. Consider (64b-C). In this case, past *ta* is bare, and it obeys the linear adjacency.

50) The phonological realizations of [ta] and [te] are allomorphs of the perfective phoneme /ta/. 
Symmetry Breaking and Economy in C

When the upper TP is created, the past morpheme *katta* adjoins to the lower T, observing the linear adjacency by not crossing the negation \(^{51}\). Notice that formal features of an affixal item are inactive, so they do not have to be (and therefore they are not) attracted to the lower T to be erased immediately after the lower TP is created. Procrastination matters to inactive features. The active formal feature of the main V is attracted to the upper T, eliminating the uninterpretable feature [F]. The whole category then raises to the upper T by GP. LCA calculates that the output order is *tabe-na-katta*. The derivation converges. We have shown that featural/affixal distinction exists with respect to the perfective/past morpheme in Japanese, as it does with respect to the main verb/auxiliary verb in English.

2.9. Island Effects ~ Interaction of MLC, PBC, UCA, Economy and Symmetry Breaking ~

Let us adopt the following definitions.

(67) **Minimize Chain Links (MCL)** (Move-based Approach)
MCL = def. Minimize chain links \(^{52}\). (Rizzi 1990, Chomsky & Lasnik 1993)

(68) **Minimal Link Condition (MLC)** (Attract-based Approach)
MLC = def. H(K) attracts \(\alpha\) only if there is no \(\beta\), \(\beta\) closer to H (K) than \(\alpha\), such that H (K) attracts \(\beta\).  
(Chomsky (1995b; 311))
\(\beta\) is closer to H (K) than \(\alpha\) iff \(\beta\) c-commands \(\alpha\), and \(\beta\) is not in the minimal domain of CH, where CH is the chain headed by \(\gamma\), and \(\gamma\) is adjoined to H (K).  
(ibid., 311))

Let us consider how Economy Framework explains the basic paradigm.

(69) **Superiority Effect**
   a. Who saw what?
   b. *What did who see?

(70) **Subject Condition Effect**
   ?* Who did [a picture of t\_wh] please John?

(71) **Complex NP Constraint (CNPC Effect)**
   a. *I wonder what John met [NP someone [CP who read t\_wh]]
   b. **I wonder why John met [NP someone [CP who read the book t\_why]]

(72) **Adjunct Condition Effect**
   a. *I wonder what John left New York [PP before [CP he read t\_wh]]
   b. **I wonder why John left New York [PP before [CP he read the book t\_why]]

---

\(^{51}\) The phonological realization of [ta] and [katta] are allomorphs of the past phoneme /ta/.[katta] appears when adjective-stem is involved.

\(^{52}\) This is also equivalent to the *Shortest Movement Condition* (SMC), which states: *Make the shortest move.*
(73)  *Wh-island Effect, HMC, Superraising
   a. *Guess what John wondered  [CP why we fixed t_{what}]
   b. *How fix [IP John will t_{fix} the car]
   c. *John seems [CP that [IP t_{John} to fix the car]]

2.9.1. Superiority Effect
2.9.1.1. English
Consider the Superiority Effect in (69).

(69) Superiority Effect
   a. Who saw what?
   b. *What did who see?

Suppose the derivation has reached the following stage.

(74)  C T  [VP who [VP saw what]]

C has a strong uninterpretable [WH]-feature. Since who and what are not in the same minimal domain\(^{53}\), who is closer to C than what is. Therefore, C attracts who.

(75)  who C T  [VP t_{who} [VP saw what]]

The phonetic information in (75) is sent to PF. The [WH]-feature of what is checked covertly. Consider (69b). Suppose the derivation has reached the following stage.

(76)  C T  [VP who [VP saw what]]

To derive (69b), C attracts what overtly. This step violates MLC: a non-closer what is attracted. The derivation crashes due to an Economy violation.

2.9.1.2. Japanese
Consider the following examples.

---

53) We adopt the following definitions of minimal domain, domain, and Max (α) (Chomsky (1993a)).
   (i) The minimal domain Min(D(CH)) of CH is the smallest subset K of D(CH) such that for any x belonging to D(CH), some γ belonging to K dominates x.
   (ii) The domain D(CH) of CH is the set of categories/features dominated by Max(α) that are distinct from and do not contain α or t.
   (iii) Max (α) is the smallest maximal projection dominating α.

Given that a main verb remains in situ in English, the relevant chain CH we are dealing with here is a trivial chain, which has a single member. What is inside the Min (D(V)), i.e., VP, and who is inside the Min(D (v)), i.e., vP.
Symmetry Breaking and Economy in Cnl

(77)  a. John-ga nani-o naze katta no?
      John-nom what-acc why bought Q
      (lit.) What did John buy why?

b. *John-ga naze nani-o katta no? (unstressed)
      John-nom why what-acc bought Q
      (lit.) Why did John buy what?

Consider (77a). Suppose the derivation has reached the following stage.

(78)  C [IP John-nom [VP what-acc [VP why [VP twhat bought Q]]]]
      [WH] [WH] [WH]

Assume that what-acc is in [Spec, vp] for Case-checking, and why is adjoined to VP. What-acc is closer to C than why is: they are not in the same minimal domain. At LF, C attracts what-acc.

(79)  What-acc C [IP John-nom [VP twhat [VP why [VP twhat bought Q]]]]
      [WH] [WH] [WH]

[WH] of what-acc is checked by Spec-head agreement, and [WH] of why is checked by unselective binding by C. The trace of the argument is easily recovered because arguments are what a verb obligatorily selects. Consider (77b). Suppose the derivation has reached the following stage.

(80)  C [IP John-nom [VP why [VP what-acc bought Q]]]
      [WH] [WH] [WH]

Why and what-acc are not in the same minimal domain: what-acc is inside Min(D(V)), which is the lower VP projection, whereas why is outside Min(D(V)). Why is closer to C than what-acc is. C attracts why at LF.

(81)  Why C [IP John-nom [VP twhy [VP what-acc bought Q]]]

The trace of an adjunct why is not easily recovered, since a verb does not obligatorily select adjuncts. Thus, the contrast in Japanese is accounted for by the principles of Economy and Recoverability.54)

---

54) Consider (77b) in Osaka dialect.
(i) ?? John-wa nande nani koo-ta-n?
      John-top why what-acc buy-past-Q
      (Lit.) Why did John buy?

In (i), an informant observes that the adjunct nande ‘why’ is interpreted as if it is an interjection that affects the entire sentence ‘Why the hell, … what did John buy?’ To avoid the LF crash, Cnl somehow solves the problem by interpreting the adjunct wh as a non-wh interjection, which is forced by the Last Resort Principle.
There is evidence that supports this analysis. If we stress what-acc in (77b), the sentence becomes acceptable.

(81) John-ga naze NANI-O katta no? (nani-o stressed)
John-nom why what-acc bought Q
‘(lit.) What did John buy why?’

We propose that the stressed object argument undergoes focus-movement before its wh-feature is checked. Assume that a focus-feature is checked by adjoining to IP at LF. Thus, we have the following representation before wh-movement.

(82) C [IP WHAT-ACC [IP John-nom [VP why [VP tWHAT-ACC bought Q]]]]
[WH] [WH] [WH]

WHAT-ACC adjoins to IP for checking focus-feature. C attracts WHAT-ACC since it contains the closest wh-feature. The final LF representation is as follows.

(83) WHAT-ACC C [IP tWHAT-ACC [IP John-nom [VP why [VP tWHAT-ACC bought Q]]]]
[WH] [WH] [WH]

[WH] of WHAT-ACC is checked by Spec-head agreement, and [WH] of why is checked by unselective-binding by C. Notice that what is left is the trace of an argument, therefore it is easily recovered. (81) conforms to the principles of Economy and Recoverability.

2.9.2. Wh-Island Effect, HMC Effect, and Superraising

2.9.2.1. English

Consider the wh-island effect, HMC effect, and Superraising in (73).

(73) Wh-Island Effect, HMC, Superraising
a. *Guess what John wondered [CP why we fixed twhat.]
b. *How fix [IP John will tfix the car]
c. *John seems [CP that [IP it is certain [IP tJohn to fix the car]]]

In (73a), why is closer to C than what is, but the non-closer what is attracted. In (73b), will is closer to C than fix is, but the non-closer fix is attracted. In (73c), the expletive it is closer to T than John is, but the non-closer John is attracted. All examples in (73) violate MLC. The derivations crash.

2.9.2.2. Japanese

The wh-island effect in Japanese is shown in the following.

(84) *Mary-wa [John-ga nani-o katta kadooka] siritagatteru no?
Symmetry Breaking and Economy in Chu.

Mary-top John-nom what-acc bought whether wonders Q
(lit.) Mary wonders whether John bought what?

The sentence in (84) is excluded in the meaning in (85a) but allowed in the meaning of (85b).

(85) a. What is x such that Mary wonders whether John bought x?
    b. Is it true that Mary wants to know what John bought?

We assume the following structure for the embedded clause.
The following is the schematic LF representation of (84) in the meaning of (85a).

(86) \[
\begin{array}{c}
\text{CP} \\
\mid \text{IP} \quad \text{whether} \\
\quad \text{John-nom} \quad \Gamma' \\
\quad \text{VP} \\
\quad \quad \text{\(t_{\text{John-nom}}\)} \\
\quad \quad V' \\
\quad \quad \text{what-acc} \quad [WH] \\
\quad V \\
\end{array}
\]

(87) \[
\text{What-acc C [Mary-top [[John-nom t_{\text{what-acc}} bought] whether] wonders Q]} \\
\quad [WH] \quad [WH] \quad [WH]
\]

C attracts a distant wh-phrase what-acc, thereby violating the Economy principle. Notice that (84) is acceptable as a yes-no question (85b), in which case, C attracts the closer wh-element whether.

(88) \[
\text{Whether C [Mary-top [[John-nom what-acc bought] \text{t}_{\text{whether}}] wonders Q]} \\
\quad [WH] \quad [WH] \quad [WH]
\]

[WH] of whether is checked by Spec-head agreement, while [WH] of what-acc is checked by unselective binding by C\(^{55}\).

\[55\text{ It is not clear what kind of examples correspond to HMC and super-raising effects in Japanese.}\]
2.9.3. Eliminating UCA

Consider the Subject Condition Effect, CNPC Effect, and Adjunct Condition Effect.

(89) **Subject Condition Effect**

"Who did [a picture of t\_who] please John?"

(90) **Complex NP Constraint (CNPC) Effect**

a. "I wonder what John met [NP someone [CP who read t\_what]]"

b. "I wonder why John met [NP someone [CP who read the book t\_why]]"

(91) **Adjunct Condition Effect**

a. "I wonder what John left New York [VP before [CP he read t\_what]]"

b. "I wonder why John left New York [VP before [CP he read the book t\_why]]"

These effects should be explained in the same way, since all these cases involve a non-complement (=adjunct). Takahashi (1994) proposes the following principle to explain these data.

(92) **Uniformity Corollary on Adjunction (UCA) (Takahashi (1994: 25))**

Adjunction is impossible to a proper subpart of a uniform group, where a uniform group is a non-trivial chain or a coordination\(^{56}\).

Let us briefly review the background of UCA. Suppose we have the following non-trivial chain: \(\alpha\) is copied to a displaced position.

(93) \((\alpha, \alpha)\)

Suppose \(\beta\), which is inside the original \(\alpha\), adjoins only to the copy (i.e., a proper subpart of a uniform group). (94) is obtained with the structure (95).

(94) \([\ldots \beta \ldots]\)

---

\(^{56}\) When a non-trivial chain is involved, UCA is a legibility condition imposed on the A-P interface. When a coordination is involved, UCA is a legibility condition imposed on the C-I interface (Ochi (1999: 13)).
The step (1) copies $\alpha$ in [Spec, XP] to [Spec, YP], and the step (2) adjoins $\beta$ within the original $\alpha$ to the copied $\alpha$. The two elements of the chain (the circled elements) are not uniform, thereby violating the principle that chains must be uniform (Chomsky (1991, 1994a), Browning (1987)).

A question arises as to what the actual example of this movement would be, wherein something is extracted out of the original (trace). Typically, we have situations in which something is extracted from a trivial-chain (1-membered chain) or out of the copy (= antecedent), but not out of the original (trace).

Why does a non-trivial chain matter in UCA anyway? A relevant case is the subject condition effect: an element cannot be extracted out of a subject. In this case, something is extracted out of the copy (subject), given that the subject raises from [Spec, vP] to [Spec, TP]. Assume the VP-internal subject hypothesis and the copy theory of movement. When a wh-phrase within the subject is extracted, the relevant structure would be the following.

The step (1) is forced by $[-\text{Int FF}]$-checking, and the step (2) is forced by MCL$^{57}$). After wh-

---

57) (i) **Minimize Chain Links** (MCL) (Move-based Approach)
MCL = def. Minimize Chain Links.
(Rizzi (1990), Chomsky & Lasnik (1993))
(ii) **Minimal Link Condition** (MLC) (Attract-based Approach)
extraction out of DPsbj, the chain is the following.

(97) \([\text{DP wh} \ [\text{DP...wh...}]], \ [\text{DP...}]]\)

The members of the chain are not uniform, thereby violating UCA. To obtain a uniform chain, *wh* within the subject must first adjoin to the subject DP in \([\text{Spec, vP}]]\), and then the entire subject phrase raises to \([\text{Spec, TP}]]\). Then, a uniform chain (\([\text{DP wh} \ [\text{DP...wh...}]], \ [\text{DP wh} [\text{DP...wh...}]]\)) is obtained. The relevant structure would be following.

(98)

The theory incorrectly predicts that there should be no subject condition effect in English, contrary to fact. What’s wrong with (98)? We argue that the step \(^1\) violates the Local Economy principle, i.e., \([-\text{Int FF}]\) must be eliminated as soon as possible. When T merges with vP, DPsbj must raise and merge with TP (now T'). This is forced because the uninterpretable features [CASE] of DPsbj and [Φ] of T must be eliminated immediately. Thus, the following derivation is forced.

---

*MLC*=def. H (K) attracts α only if there is no β, β closer to H (K) than α, such that H(K) attracts β.  
(Chomsky (1995b; 311))

β is closer to H (K) than α iff β c-commands α, and β is not in the minimal domain of CH, where CH is the chain headed by γ, and γ is adjoined to H (K).  
(ibid., 311)

(i) is equivalent to the Shortest Movement Condition (SMC), which states: Make the shortest move. \([-\text{Int FF}]\)-checking must take place first, since uninterpretable-feature elimination must take place as soon as possible whenever/wherever possible. This is forced by the Local Economy principle.
The step ① is forced for eliminating uninterpretable formal features. The steps ② and ③ take place so as to obtain a uniform chain. The step ② violates the Proper Binding Condition (PBC): wh-copy (=wh, which is adjoined to DPsubj-original in [Spec, vP]) does not c-command wh-original (=wh, which is contained in DPsubj-copy in [Spec, TP]). PBC was originally defined as follows.

(100) *Proper Binding Condition (PBC)*
Traces must be bound. (Fiengo 1977, May 1977)

In our terms, a trace is the original element, and the antecedent is the copy. According to the minimalist reading of PBC, the copy must c-command the original.

(101) *PBC (revised)*
The copy must bind the original.

Let us adopt the version (101) of PBC. In fact, the step ② is a lowering operation, which does not exist in natural language. We obtain the same PBC violation if *wh* first adjoins to the subject DP in [Spec, TP], and then to the subject DP in [Spec, vP].

Whenever we want a uniform non-trivial chain with extraction out of the copy, the chain is excluded by PBC. If we want to avoid PBC violation, MCL violation is induced. Notice that extraction from the original (trace) does not exist. PBC and MCL take mechanical care of Economy-violation effects. UCA is redundant, and that it should be dispensed with.

Suppose we have to extract and adjoin something out of the copy. Whenever we want a uniform chain with a non-trivial chain, it ends up in PBC violation. In order to avoid PBC violation, the adjunction must take place only to the copy, not to the original. But then, it always ends up in MCL violation. The following is the general structure of PBC violation.
(102) \[
\begin{array}{c}
\text{XP} \\
\alpha \\
\beta \\
\vdots \\
\beta \\
\alpha \\
\ldots \vdots \\
\alpha \\
\ldots \text{XP} \\
\beta \\
\text{Y} \\
\vdots \\
\beta \\
\alpha \\
\ldots \beta \\
\end{array}
\]

1 is forced by MCL. But then it always violates PBC. The copy does not c-command the original, and the original does not c-command the copy. Let us call this situation mutual anti-c-command. We propose that mutual anti-c-command is symmetric in nature, as is mutual c-command. The symmetric relation is unstable, and therefore it is costly to maintain the relation. LCA also excludes mutual c-command and mutual anti-c-command since they are too symmetric to yield an order in 1-D. Thus, the PBC violation is considered as computational failure due to symmetric structure.

Let us consider the coordination part of UCA. The crucial assumption is that adjuncts involve coordination (Davidson 1967, Higginbotham 1985, Takahashi 1994, Ochi 1999:10-13). An adjunct is one of the conjuncts within the coordination structure. Consider the adjunct slowly in the following example.

(103) John walks slowly

The following is the semantic representation.

(104) \(\exists e \ [\text{walk}(\text{John}, e) \& \text{slow}(e)]\)

This is read as “there is an event such that it was a walking by John and it is slow.” The following is the LF representation.

(105) \([_{\text{IP}} \text{John} \ [_{\text{TP}} \text{INFL}(e) \ [_{\text{VP2}} \text{t walks} \text{slowly}]]]\]

VP2 and slowly are coordinated. Given this much, let us consider the CNPC effect in (90a).

(90a) *I wonder what John met \([_{\text{NP}} \text{someone} \ [_{\text{CP}} \text{who read t\_wh}]])\]
Assume that NP1 and CP are coordinated.

(106)

By MCL, what adjoins to every possible landing site. What first adjoins to CP, and then to NP1.

(107)

The copy what2 does not c-command the original what1, which is a PBC violation. Therefore, PBC forces the following structure.

(108)

However, (108) violates MCL (=SMC). If we respect MCL, then what must adjoin to NP1, violating PBC. There is no way we can extract the wh-phrase out of the complex NP: it would always violate either PBC or MCL. We lost the argument-adjunct contrast between (90a) and (90b), which ECP could account for, however\(^8\). Consider the adjunction condition effect (91a).

(91a) *I wonder what John left New York [\[pp \ before \ [cp he read t\_what]]\]

\(^8\) Chomsky (1993b) suggests that the argument-adjunct contrast may be an artifact.
Suppose we have the following structure.

(109)

```
  . . VP1

  VP2     PP

. . left N.Y.    after . . what . .
```

*What* first adjoins to PP. In the next step, MCL forces *what* to adjoin to VP2; if it did, the copy would not c-command the original, thereby violating PBC. Therefore, *what* must adjoin to PP, and then to VP1. This is forced by PBC. But this derivation is barred by MCL since it skips a possible landing site, namely, VP2. There is no way we can extract *wh* out of adjunct clause. Again, we lost the argument-adjunct contrast.

In fact, whenever we want a uniform chain with a coordination structure, PBC violation takes place. Consider the general structure. Let us assume that a coordination structure has a head *And*, the phonetic feature of which is either absent or present.

(110)

```
    And

    α

    γ   α   And

    And  β

    γ   β

②  ①

... γ ...
```

The step ② always violates PBC. That is to say, the copy does not c-command the original: γ, which is raised by the step ② excludes the original γ, which is raised by the step ①, but the first node containing the former γ (= the upper node α in [Spec, AndP]) does not contain the latter γ. In fact, the original does not c-command the copy, either. This is the relation of mutual anti-c-command, which is too symmetric a structure. The symmetric structure is excluded by Economy since it is unstable and costly to maintain.

2.9.4. Subject Condition Effect, CNPC Effect, and Adjunct Condition Effect in Japanese

Consider the subject condition effect, CNPC effect, and adjunct condition effect in Japanese.
Symmetry Breaking and Economy in C_HL

(111) a. [[John-ga nani-o kattta] koto]-ga Mary-o komaraseteru no?
   John-nom what-acc bought fact-nom Mary-acc cause-trouble Q
   (lit.) The fact that John bought what is causing trouble to Mary?

b. Mary-wa [[nani-o katta] hito]-o mita no?
   Mary-top what-acc bought person-acc saw Q
   (lit.) Mary saw a person who bought what?

c. Mary-wa [[pro nani-o katta] ato]-de dekaketa no?
   Mary-top what-acc bought after-at left Q
   (lit.) Mary left after buying what?

In all these cases, wh-phrases are interpreted at the matrix level.

(112) a. What is x such that the fact that John bought x is causing trouble to Mary?
   b. What is x such that Mary saw a person who bought x?
   c. What is x such that Mary left after she bought x?

How can we account for these examples? The simplest account at this point is as follows. Assume the following.

(113) a. C bearing [WH] can check [WH] either by Spec-head agreement or by unselective binding.
   b. Each checking takes place once.
   c. [WH] of C in Japanese is weak: it is checked covertly.

(113a) and (113b) are deducible from more general principles. The only parameter we have to speculate for Japanese is (113c). In these examples, [WH] is unselectively bound by C covertly.

3. Summary

We summarize our main points in the following.

(114) a. We have argued for the possibility of connecting Economy principles of C_HL with more general spontaneous symmetry breaking in natural science. Hence, the possibility of integration of natural language studies with the still-to-be-discovered physics, which might contain the issues of spontaneous symmetry breaking.
   b. We have argued for the existence of 1/d and L/D distinctions in natural language, which interact with the extremely general Polarization Axiom.
   c. We have proposed the way in which LCA, BPS and label-free phrase structure are unified. The modified LCA incorporating BPS and label-free phrase structure is more general, more natural, and simpler.
   d. We argued that the notions such as label and projection must be eliminated.
   e. We have pointed out some problems found in the analysis of the passive by Nunes
(2001), and proposed a simpler explanation.

f. We have pointed out some problems found in the Hybrid Model by Lasnik (1995), and proposed a simpler explanation which incorporates the Two-Chain Hypothesis (Ochi 1999).

g. We have argued that the conventional Law of Conservation of Cth (cf. Uriagereka (1998)) must be strengthened. The modification conforms to the Law of Conservation of Energy generally assumed in natural sciences.

h. We have argued that the UCA should be eliminated. The PBC takes care of those examples the UCA tries to deal with.

References


Symmetry Breaking and Economy in CL


Symmetry Breaking and Economy in $C_{HL}$

Koji ARIKAWA

The computational system of the human natural language ($C_{HL}$) is a realization of the emergent and self-organizing properties of the human brain. $C_{HL}$ is the typical case of a complex system: (i) it shows the butterfly effect, i.e., the final state of $C_{HL}$ shows remarkable variation in the phonetic and formal aspects, (ii) it is sensitive to the initial-state conditions, i.e., the language system of a human infant (the initial state of $C_{HL}$) is crucially affected by the internal/external environments which s/he first encounters, (iii) iteration of a simple operation plays an important role, i.e., a simple operation, such as Merge, which concatenates two elements, plays a crucial role in forming the phrase structure, and (iv) it has boundary conditions, within which the system allows variations, i.e., the legibility conditions imposed on $C_{HL}$ by various external performance systems determine what possible form the human natural language should have. Our study deals with the general problem of optimal stuffing: What is "the possibility for something with $10^{11}$ neurons (or $10^{14}$ synapses) packed into something the size of a basketball that has these computational properties" (Chomsky 1994)? The linear orderings and structures of the human natural language are emergent phenomena through a symmetry-breaking mechanism, which is governed by the Law of Entropy. This paper tries to support the view given in Jenkins (2000) that "the origins of (some) properties of language, such as (a)symmetries in syntax perhaps, would be more akin to the physical evolution of molecular chirality."