

More Evidence for Geometrical Cost Approach To Basic Word Order Asymmetry in Human Language

ARIKAWA Koji

Keywords : unmarked word order; third factor; symmetry;
Galois group; algebraic structure of equations

0. Introduction

I would like show some evidence for the hypothesis that human mathematical capacity is derived from human language (Chomsky 2005: 16; 2007: 7, 20; 2010: 53). The paper is organized as follows. In Section 1, I claim that a set of syntactic relations constitutes a group (G) under a syntactic operation Merge. In Section 2, I review Arikawa (2012 b) that proposes that geometrical cost asymmetry is the fundamental cause of the word order asymmetry among S, O and V. I indicate a correspondence between transformational cost and geometrical cost. Section 3 suggests that a type of conservation law is working in C_{HL} and that word-order cost, agreement cost, and scrambling cost interact. Section 4, which employs the 24 isometries of a regular tetrahedron, applies the geometrical cost approach to DP-internal unmarked word order. Section 5 summarizes the paper. Appendix uses elementary algebra in a more radical attempt to speculate, at least roughly, about what it would be like if something such as a language equation truly existed in C_{HL} . Despite a possible lack of promise from a purely mathematical viewpoint, I hope that my approach will lead to possible future research from the combined perspective of applied

mathematics and biolinguistics.¹

1. Syntactic Relation as Group under Merge

I first argue that a set of syntactic relations constitutes a group (G) under a syntactic operation Merge.² Merge takes a pair (unordered set) of syntactic objects (SO_i, SO_j) and replaces them by a new combined syntactic object SO_{ij} (Chomsky 1995: 226). A group G , unlike a set, is a good mathematical tool for characterizing dynamic phenomena such as syntactic relations under Merge.³ G must satisfy the following four requirements (G axioms).

(1) *A group G Axioms*

- a. G is closed under a relevant operation:
If $a \in G$ and $b \in G$, then $a \bullet b \in G$.⁴
- b. G has an identity element:
 $a \bullet x = a$ and $x \bullet a = a$, where x is a member of G . x is the identity element (I).
- c. G has an inverse element:
 $a \bullet y = I$ and $y \bullet a = I$, where y is a member of G . y is the inverse element of a .
- d. G obeys the associative law:
 $a \bullet (b \bullet c) = (a \bullet b) \bullet c$, where a , b , and c are arbitrary members of G .

Consider axiom (1 a). In the structure of a sentence, the terms stand in constituent-command (c -command) relations (syntactic relations). The c -command relation is defined as follows:⁵

(2) *C-command*

α *c-commands* β if and only if

- (i) α does not dominate β , and
- (ii) all nodes that dominate α also dominate β .

The *c-command* relation expresses an equilibrium between connection and disconnection among the terms in a tree.⁶ Condition (2i) expresses the disconnection; no dominance, i.e., no direct descent, and (2ii) expresses the connection; α and β share the maternal nodes. Suppose that X *c-commands* Y, and Y *c-commands* Z, expressed as $X \gg Y \gg Z$ (i.e., X is higher than Y, which is higher than Z), as in Figure 1:

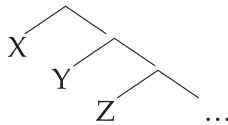


Figure 1 : $X \gg Y \gg Z$

If a copy of Z remerges (internally merges) with the node that dominates X, Y, and Z, Merge transforms $X \gg Y \gg Z$ to $Z \gg X \gg Y \gg Z$, as shown in Figure 2:⁷

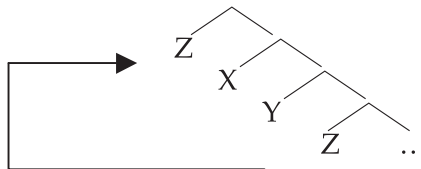


Figure 2 : $Z \gg X \gg Y \gg Z$

All these terms stand in the *c-command* relation. A merge of any two syntac-

tic objects realizes a syntactic relation. The C_{HL} syntactic relation is closed under the merge operation, thus obeying axiom (1a).⁸

Consider axiom (1b). I propose that C_{HL} creates the base vP, which is the identity element under the Merge operation.⁹ The base vP has the c-command relation $S \gg O \gg V$. The base vP is formed with the least effort, that is, only an external merge (the simplest possible structure-building operation) builds it. Every sentence structure starts with the base vP.

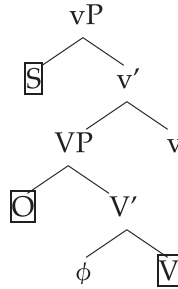


Figure 3 : Base vP that is Mapped to $S \gg O \gg V$

Why is this structure the base?¹⁰ First, it is the most cost-effective structure: the base vP is built by external merges only. If the cost is zero, the base vP corresponds to the identity (do-nothing) operation, which is the most cost-effective transformation. It is like the identity operation $+0$ under addition, which does not affect a number (for example, $3 + 0 = 3$). Second, it is the most fundamental structure: every sentence structure contains the base vP at its deepest structure. Third, it gives us semantic universality: the base vP is the minimal domain where the V's inherent semantic information is assigned to O and S, and this holds universally. Fourth, there is V's affinity for O: universally, V has an affinity for O rather than S.¹¹ Thus, C_{HL} disallows other possibilities.

Let us demonstrate how the base vP is constructed. Given that each set

includes the empty set by definition and that a syntactic object is a set, each syntactic object includes the empty set \emptyset . V externally merges with ϕ .¹² V' and O merge, and V assigns Patient θ (a semantic role) to O .¹³ The light verb v merges with VP . The v' merges with S and v assigns Agent θ to S . Thus, the base vP is the most inexpensive base for building the structure of $\{S, O, V\}$ because it is formed by external merges only, given the Merge-over-Move hypothesis, and so every sentence starts with the base vP . Every final structure contains the base vP as a subset, and the base vP does not affect the usable c -command relations in the final structure. As noted above, the base vP is like the identity element 0 in addition. Probe uninterpretable feature in v agrees with the goal interpretable feature in O , the relevant structural feature is valued and deleted (Chomsky 2000).¹⁴ The structural Case variable is deleted within the C_{HL} language system because such a variable is unknown to the performance systems (the sensorimotor system and the thought system).¹⁵

The base vP is the most economical structure that satisfies the Linear Correspondence Axiom (LCA; originally proposed by Kayne 1994). LCA is a principle at the sound interface that maps two-dimensional structures to one-dimensional linear orders. A structurally higher term should be pronounced earlier. Assume the following definition of LCA (Uriagereka 2012: 56).¹⁶

(3) *LCA*: When x asymmetrically c -commands y , x precedes y .

The base vP does not influence later structures. Suppose we arrived at $V \gg S \gg O$ as the final structure. LCA sees only the boxed terms in Figure 4 (T = Tense).

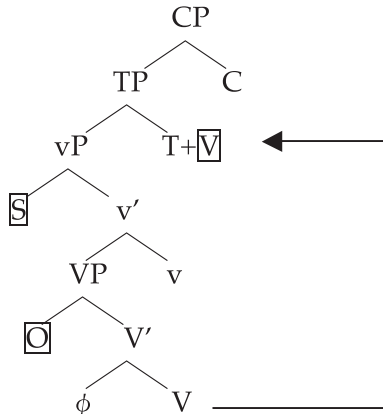


Figure 4 : $V \gg S \gg O$

Spell-out sends the final CP structure to the PF (semantic interface), and LCA maps this structure to the linear order $\langle VSO \rangle$ or $[VSO]$. Although the final CP structure contains the base vP whose syntactic relation is $S \gg O \gg V$, the final structure is not affected by the base vP (recall that the base vP is like the identity element 0 for addition). The C_{HL} syntactic relation thus obeys axiom (1b).

Consider axiom (1c). Suppose that we reached the structure shown in Figure 4. The inverse of $V \gg S \gg O$ corresponds to movements of O and S, where O moves to the lower edge of TP and S to the higher edge, thus yielding the c-command relation of the base vP, that is, $S \gg O \gg V$, as in Figure 5.

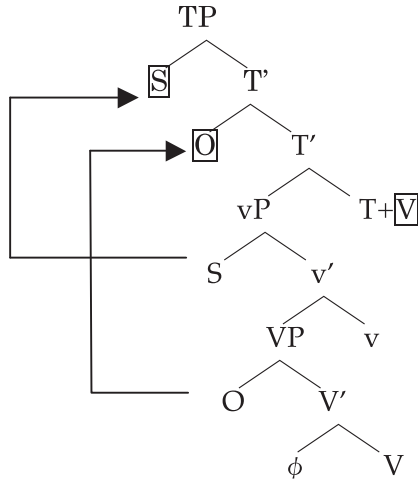


Figure 5 : Application of the Inverse Operation to $V \gg S \gg O$
 Produces the Identity Relation $S \gg O \gg V$

A set of internal merges can transform any relation, $V \gg S \gg O$ in this case, to the identity relation $S \gg O \gg V$. This relation-changing operation, $V \gg S \gg O \rightarrow S \gg O \gg V$, is the inverse element. The C_{HL} syntactic transformation thus has an inverse element and obeys axiom (1c).

Consider axiom (1d). Let us assume that a set-merge structure $\{\alpha, \beta\}$ is asymmetrical in that either α or β projects. Suppose that α and β merge and α projects, forming α . Does the following equation hold in C_{HL} ?

$$(4) \quad (x \bullet y) \bullet z = x \bullet (y \bullet z)$$

On the left side of the equation, in the first step, x and y merge and form x (x projects). In the second step, x and z merge and form x (x projects). On the right side of the equation, in the first step, y and z merge and form y (y pro-

jects). In the second step, x and y merge and form x (x projects). The equation holds. The following trees show the associativity.



Figure 6 : Head Initial: $(x \bullet y) \bullet z = x \bullet (y \bullet z) = x$

The final output is the same: x is the maximal dominator.

Suppose next that α and β merge and β projects, forming β . Does the equation hold? On the left side of the equation, in the first step, x and y merge and form y (y projects). In the second step, y and z merge and form z (z projects). On the right side of the equation, in the first step, y and z merge and form z (z projects). In the second step, x and z merge and form z (z projects). The equation holds. The following trees show the associativity.



Figure 7 : Head Final: $(x \bullet y) \bullet z = x \bullet (y \bullet z) = z$

The final output is the same: z is the maximal dominator. Therefore, the C_{HL} syntactic relation obeys the associative law in axiom (1d).¹⁷ The C_{HL} syntactic relation constitutes a group G under Merge.¹⁸

2. Transformational Cost as Geometrical Cost

2.1. Equilateral Triangle and Basic Word Order Asymmetry

In Arikawa (2012b), I argued that the symmetry structure of an equilateral triangle, expressing the group-theoretical structure of cubic equation, accounts for the asymmetry of basic word orders. I used an equilateral triangle that is the Identity Element (the basic word order <SOV>) as in the following.

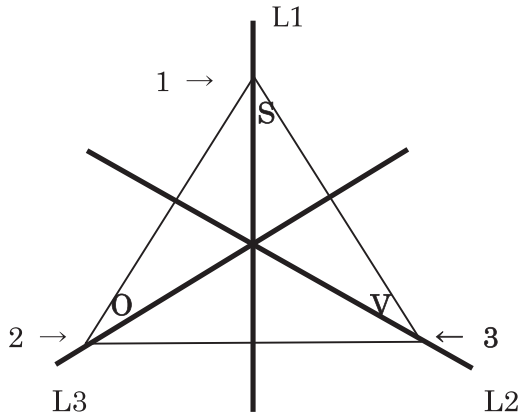


Figure 8 : Identity Element $I = \langle \text{SOV} \rangle$

The six symmetrical transformations are as follows.

- (5) a. $r0 = 0^\circ = I$ (do-nothing rotation)
- b. $r1 = 120^\circ$ (counterclock) rotation
- c. $r2 = 240^\circ$ rotation
- d. $f1 = \text{Flip around axis L 1}$
- e. $f2 = \text{Flip around axis L 2}$
- f. $f3 = \text{Flip around axis L 3}$

The six operations are expressed by $r0$, $r1$, and $f1$. These three operations are “atoms” of transformation in that they are more basic (Armstrong 1988).

- (6) a. $r0$
- b. $r1$
- c. $r2 = r1 \times r1 = r1^2$
- d. $f1$
- e. $f2 = f1 \times r1$
- f. $f3 = r1 \times f1$

The following table summarizes the transformations and costs.

Transformation	Cost	Input	Output	Ratio
$r0$	0	<SOV>	<SOV>	48.5%
$r1$	2	<SOV>	<VSO>	9.2%
$r2$	4	<SOV>	<OVS>	0.7%
$f1$	1	<SOV>	<SVO>	38.7%
$f2$	3	<SOV>	<OSV>	0.5%
$f3$	3	<SOV>	<VOS>	2.4%

Table 1 : Transformations and Costs for {S, O, V}

I assume the ratio observed in Yamamoto (2002), which considers the largest number (2,932) of languages for typological analysis to date (gross \doteq 6000). The Galois theory and the Economy Principle can explain the current ratio of languages with the top three unmarked word orders:

- (7) a. $r0$ (cost 0) produces <SOV> with a ratio of 48.5%.
- b. $f1$ (cost 1) produces <SVO> with a ratio of 38.7%.
- c. $r1$ (cost 2) produces <VSO> with a ratio of 9.2%.

Although the geometrical cost approach fails to predict the internal ranking among $f2$, $f3$, and $r2$, it does predict their relatively low probability:

- (8) a. $f2$ (cost 3) produces $\langle OSV \rangle$ with a ratio of 0.5%.
 b. $f3$ (cost 3) produces $\langle VOS \rangle$ with a ratio of 2.4%.
 c. $r2$ (cost 4) produces $\langle OVS \rangle$ with a ratio of 0.7%.

The geometrical cost approach accounts for the fact that C_{HL} shows the following asymmetry with respect to basic word order frequency.

- (9) $SOV > SVO > VSO > VOS > OVS >? OSV$

2.2. C_{HL} Selects Cheaper Subgroups

The steps in the top two transformations, unlike the other four, constitute a subgroup of G . Let us consider the multiplication table that consists of the single steps $r0$ and $f1$. The intersection of each column and row expresses the multiplication operation for that column and row.

	$r0$	$f1$
$r0$	$r0$	$f1$
$f1$	$f1$	$r0$

Table 2 : Multiplication Table for $r0$ and $f1$: Closed

The table entries are $r0$ and $f1$. By axiom (1a) of the definition of groups, $\{r0, f1\}$ is closed. It constitutes a subgroup of G . Incidentally, $\{r0, f2\}$ and $\{r0, f3\}$ are also closed and constitute a subgroup of G . The cyclic permutations (rotations), namely $r0$, $r1$, and $r2$, are closed and constitute a subgroup of G . On the other

hand, the set of noncyclic permutations $\{f1, f2, f3\}$ is not closed and does not constitute a subgroup of G (Stewart 2007: 112). C_{HL} seems to employ a subgroup that consists of the cheapest costs among both rotations and of flips, avoiding exclusive use of either type of transformation. Consider the multiplication table that consists of the steps in $r1 (= r1), r2 (= r1 \times r1), f2 (= f1 \times r1)$, and $f3 (= r1 \times f1)$:

	$r1$	$r2$	$f2$	$f3$
$r1$	$r2$	$r0$	$f1$	$f2$
$r2$	$r0$	$r1$	$f3$	$f1$
$f2$	$f3$	$f1$	$r0$	$r1$
$f3$	$f1$	$f2$	$r2$	$r0$

Table 3 : Multiplication Table for $r1, r2, f2$, and $f3$: Not Closed

The table entries include other operations, namely $r1, r2, f2$, and $f3$. According to axiom (1a) of the definition of G , the set $\{r1, r2, f2, f3\}$ is not closed under the multiplication operation and therefore does not constitute a subgroup of G . I believe it is significant that the transformational steps involved in the two basic word orders with relatively high probabilities, $\langle SOV \rangle$ and $\langle SVO \rangle$, constitute a subgroup of G , while those involved in producing remaining word orders, which have relatively low probabilities, do not. The set $\{r0, r1, r2, f1, f2, f3\}$ has six subgroups: $\{r0, r1, r2, f1, f2, f3\}, \{r0, r1, r2\}, \{r0, f1\}, \{r0, f2\}, \{r0, f3\}, \{r0\}$ (Stewart 2007: 112-113). C_{HL} selects the two cheapest subgroups, $\{r0\}$ and $\{r0, f1\}$, which produces $\langle SOV \rangle$ and $\langle SVO \rangle$ as the two most common basic word orders.

2.3. Geometrical Transformation Corresponds To Syntactic Transformation

The geometrical cost of a syntactic structure corresponds to the transformational cost. The spell-out structure of <SOV> is the identity vP. LCA demands that the boxed terms be pronounced. The identity vP corresponds to Θ -Domain (Thematic relation) proposed in Grohmann (2000: 55; 2011: 274-275).

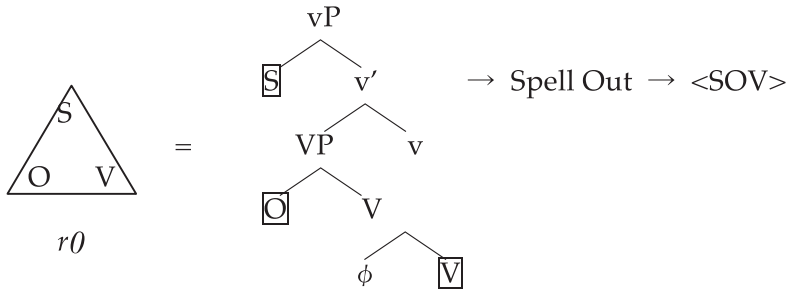


Figure 9 : Spell-Out Structure Corresponding $r0$ and <SOV>

Only external merges are involved in forming the base vP, and its structure building is the most cost-effective. This is the reason why C_{HL} demonstrates that <SOV> has the highest probability (48.5%) of the six possible unmarked word orders. The base vP is the domain in which v initiates *n-agreement* (Elouazizi and Wiltschko 2006). The spell-out transfers the base vP structure to the semantic interface, where LCA computes it as <SOV>. Let us next consider the spell-out structure of <SVO>, namely, TP. The TP corresponds to Φ -Domain (Agreement properties) proposed in Grohmann (2000: 55; 2011: 274-275).

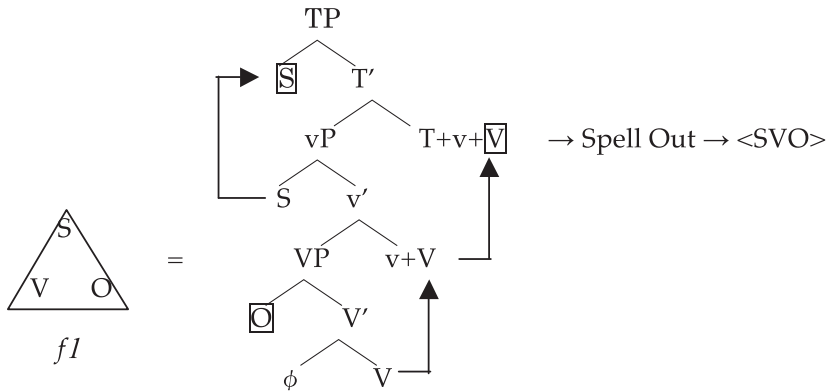


Figure 10 : Spell-Out Structure Corresponding to *fI* and <SVO>

The new head T merges with vP, and the heads and the subject undergo movement (internal merge).¹⁹ Since internal merge = external merge + copy + remerge and there are three internal merges, this structure is more costly to build than the base vP. This is the reason why <SVO> has a lower probability (38.7%). TP is the domain in which T initiates what Elouazizi and Wiltchko call *Φ-agreement*. Let us consider the spell-out structure of <VSO>, namely CP. The CP corresponds to *Ω*-Domain (Discourse information) proposed in Grohmann (2000: 55; 2011: 274-275).

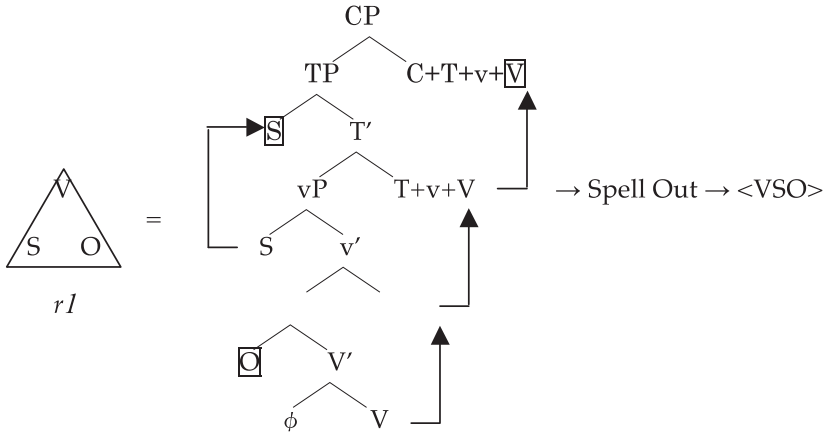


Figure 11 : Spell-Out Structure Corresponding to *r* 1 and <VSO>

C merges with TP, and V moves to C. Building the structure for such a CP requires more energy. The CP structure is the third most cost-effective, and this is the reason why the unmarked <VSO> order has a probability of 9.2%. CP is the domain in which C initiates what Elouazizi and Wiltschko call *D-agreement*. In modern standard Arabic (<VSO>), for example, *D-agreement* occurs only when V moves to C (Elouazizi and Wiltschko: 156).

I propose that MLCA applies to the base vP and derives <VOS> in languages such as Malagasy (Austronesian family). In Malagasy, MLCA applies to phrases and LCA applies to heads. If this approach is on the right track, we have a partial explanation of why C_{III} produces the unmarked word order asymmetry. Table 4 summarizes the four major word orders.

	Ordering principle	Input structure	Output order	Geometrical transformation	Cost	Probability
Type 1	LCA	Base vP	<SOV>	$r0$	0	48.5%
Type 2	LCA	TP (S+V mmt)	<SVO>	$f1$	1	38.7%
Type 3	LCA	CP (S+V mmt)	<VSO>	$r1$	2	9.2%
Type 4	MLCA	Base vP	<VOS>	$f3$	3	2.4%

Table 4 : Deriving the Major Unmarked Word Orders

Why is MLCA so costly when it applies to the base vP in Type 4? Note that MLCA generally applies to heads, but in Type 4, it applies to a phrase. This unusual application of MLCA to a phrase may be responsible for the relative low probability.²⁰

A mathematical fact is that the six symmetric transformations of an equilateral triangle with the three vertexes {a, b, c} express the six permutations of the three roots {a, b, c} of a cubic equation. A linguistic fact is that the cost difference in the syntactic tree formation with three terms {S, O, V} matches the probability difference in the unmarked word orders of {S, O, V}. C_{HL} must be solving a cubic equation with the roots {S, O, V}. Appendix provides a baby algebra of {S, O, V}. Let us summarize the discussion up to this point.

- (10) a. The C_{HL} syntactic relations constitute a group.
 b. The cost hierarchy among the six geometrical operations that correspond to the six unmarked word orders in C_{HL} is:

$$r0 < f1 < r1 < f2 = f3 < r2,$$

where $r0$ produces <SOV>, $f1$ <SVO>, $r1$ <VSO>, $f2$ <OSV>, $f3$ <VOS>, and $r2$ <OVS>. The geometrical cost approach predicts the current percentages of languages. The top three word orders: <SOV> (48.5%), <SVO> (38.7%), and <VSO> (9.2%).

- c. Although this approach fails to predict the internal relative ranking of the lower three basic word orders, it nevertheless predicts a division between the higher three orders (<SOV>, <SVO>, and <VSO>) and the lower three orders (<VOS>, <OSV>, and <OVS>).
- d. The fourth major unmarked word order <VOS> (2.4%) can be derived by applying MLCA to the base vP. LCA applies to it generally.
- e. The steps in the transformations corresponding to the top two unmarked orders, that is, the operation $r\theta$ (I) that produces <SOV> and the operation fI (OV) that produces <SVO>, are closed and constitute a subgroup of G . The transformation for the remaining four orders are not closed and do not constitute a subgroup of G .
- f. The sound interface of C_{HL} employs LCA and MLCA for phrases and heads. LCA and MLCA eliminates all technologies related to head movement, simplifying the model of structure-order mapping.

3. Scrambling and the Conservation Law

The conservation law answers question about asymmetry of operations.

(11) *Conservation law*

The gross cost is fixed.

Suppose that the maximum cost for C_{HL} computation is 1. <SVO> languages have an overt (phonetically realized, hence costly) agreement morphology. <SOV> has cost 0 for deriving the basic order, whereas <SVO> has cost 0.5 for the same purpose and 0.5 for pronouncing agreement. If the gross cost is 1, more energy is left for other work (scrambling) in <SOV> languages, whereas no energy is left in <SVO> languages. A language such as Hindi shows a

mixed order: head-final <SOV> for the main clause and head-initial <SVO> for the subordinate clause.²¹ Duch and German demonstrate the opposite: the main clause shows <SVO>, and the subordinate clause shows <SOV>. It is reported that <VSO> languages have relatively rigid word order. A hypothetical cost calculation is shown in Table 5.²²

	<SOV>	SOV/SVO Mixed ²³		<SVO>	SVO/ VSO Mixed	<VSO>	
		Main: SOV Sub: SVO	Main: SVO Sub: SOV				
Language family (selected)	Afro-Asiatic, Altaic, Chibchan, Dravidian, Indo-Aryan, Uto-Aztecan	Indo-Aryan	Indo-European	Arawakan, Anstronesian, Indo-European, Niger-Congo, Sinae	Anstronesian, Indo-European, Semitic, Totozoque-An	Austronesian, Celtic, Otomanguean, Niger-Congo, Semitic	
Examples (selected) ²⁴	Amharic, Korean, Bengali, Hopi, Quechua, Tamil	Hindi	Dutch, German	English, Indonesian, Mandarin, Swahili, Wayuu	Batak (karo), Modern Greek, Syrian Arabic, Topehua	Modern Standard Arabic, So, Tagalog, Welsh ²⁵ , Zapotec	
C o s t	Basic order	0	About 1/3 or less	About 1/3 or more	About 0.5	About 0.5 or more	About 1
	Overt agreement	About 0	About 1/3	About 1/3	About 0.5	About 0.5 or less	About 0
	Scrambling	About 1	About 1/3 or more	About 1/3 or less	About 0	About 0	About 0
	Gross	1	1	1	1	1	1

Table 5 : Cost of {Basic Order, Agreement, Scramblability}, and the Conservation Law

The gross cost is the same in all languages. C_{HL} has to manage on the same cost. <SOV> shows relatively higher symmetry of derived orders (scrambling) because C_{HL} can use (and uses) more energy in scrambling. On the other hand, <SVO> and <VSO> show relatively lower symmetry of derived orders (permutation is more restricted) because C_{HL} needs more energy in producing the unmarked orders and not much energy is left for scrambling. C_{HL} obeys the conservation law. The unmarked word order, the overtness of agreement, and the possibility of scrambling interact. They are epiphenomena resulting from the dynamic cost equilibrium of C_{HL} .

4. DP-Internal Order and Geometrical Cost

How does the geometrical-cost approach explain the word-order asymmetry within nominal expressions? Extending Greenberg (1963, 1966), Cinque (2005: 319-320) reports the four major orders of the four elements in DP as follows:

(12) *Top four word orders in DP*

- a. <Dem, Num, A, N> (Very many)²⁶
- b. <N, A, Num, Dem> (Very many)²⁷
- c. <Dem, Num, N, A> (Many)²⁸
- d. <Dem, N, A, Num> (Many)²⁹

I propose that C_{HL} selects these four patterns from the 24 (4!) possible patterns simply because they are cheaper. Cinque assumes that order (12a) is the structure that is produced only by external merges and that the other three orders are derived from the movement of the terms in (12a). Assuming that external merge is cheaper than internal merge, (12a) is the most cost-effective order:

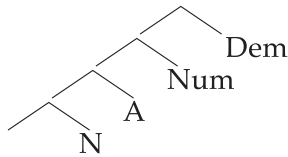


Figure 12 : Base DP: Nominal Structure by External Merge Only

Following Cinque, I propose that the above structure is the identity element I , which is realized with the minimum cost 0. Let us consider this structure to be the base DP. The base DP for a nominal expression corresponds to the base vP for a verbal expression. LCA produces $\langle \text{Dem}, \text{Num}, \text{A}, \text{N} \rangle$.³⁰ A syntactic structure of four terms corresponds to the geometrical image of a regular tetrahedron. The linear order of the four terms corresponds to the following vertexes in their relative positions. Dotted lines indicate see-through edges.³¹

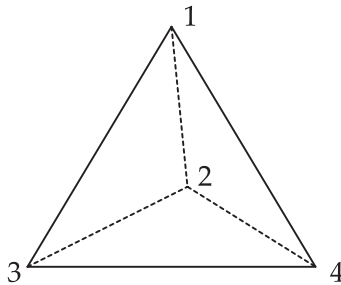


Figure 13 : Mapping Between Tetrahedral Vertexes and Linear Order

A regular tetrahedron has 24 isometries, forming the symmetry group T_d , which is isomorphic to S_4 . Three kinds of symmetry axis exist. The first kind is rotational axis L , which goes through a vertex, perpendicular to the oppo-

site plane. Rotations of 0° , 120° , and 240° around L preserve the symmetry. Four Ls exist.

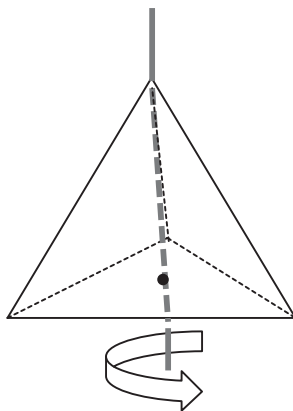


Figure 14 : Rotation around Axis L

The second kind is another rotational axis M, which penetrates through the tetrahedron between the middle point *mp* of an edge and the *mp* of the edge on the opposite side. Rotations by 0° and 180° around M preserve the symmetry. Three Ms exist. An origami tetrahedron is necessary at this point.

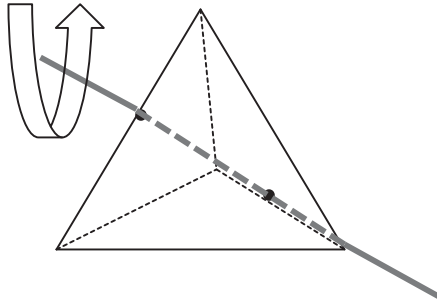


Figure 15 : Rotation around Axis M

The third kind is reflectional axis R (mirror), which is perpendicular to an edge. Reflections in the plane R replace two positions in the base, which is an equilateral triangle. Three Rs exist.

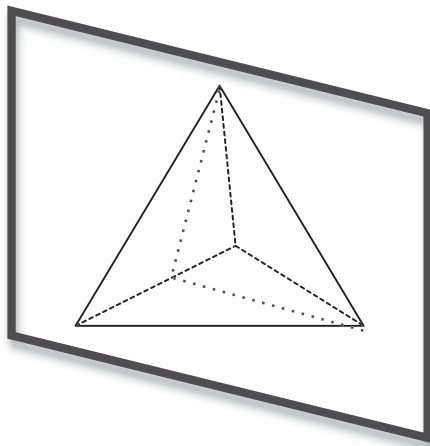


Figure 16 : Reflections in R

Let us assume that LCA computes the base-DP tetrahedron as

$\langle \text{Dem}, \text{Num}, A, N \rangle$. The base DP, which involves no internal movement, is the most cost-effective base (with a cost of 0). The base DP geometrically corresponds to a regular tetrahedron as follows. In the base DP, the vertex (top) *Dem* protrudes over the base equilateral triangle with the apexes *Num* (front), *A* (left), and *N* (right).

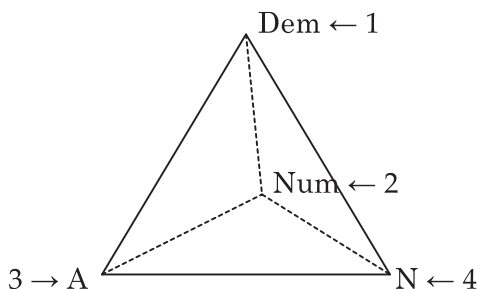


Figure 17 : Base-DP Tetrahedron

Let us simplify the base-DP tetrahedron as follows:

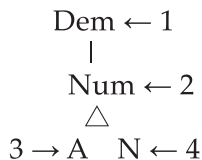


Figure 18 : Simplified Structure for Base-DP Tetrahedron

Let us consider the 24 isometries. First, consider rotations around LI , which goes through the *Dem* vertex of the base DP. LI is the default axis. Start from the base DP, which involves a 0° rotation around LI . Keeping *Dem* at the top, we have three isometries: 0° , 120° , and 240° rotations around LI . In Figure 19-23, the frequency expressions and the letters inside the brackets [] corre-

spond to those in the table in Cinque (2005: 319-320). The structures in the boxes are the attested four major DP-internal unmarked word orders.

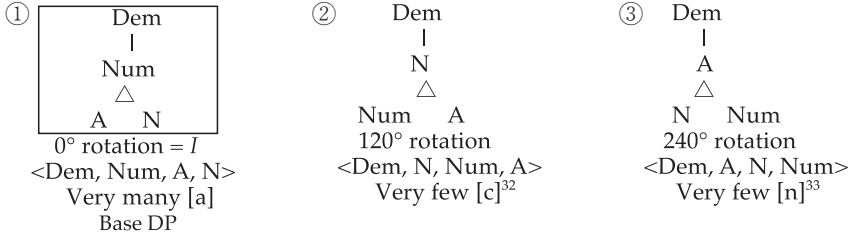


Figure 19 : Rotations of Base DP around $L1$

The isometry ① (the most cost-effective) corresponds to the base DP, which is built by external merges only. Let us next consider the rotations around M . There are three isometries. We begin with $M1$. (The term “mp (x, y)” denotes the midpoint between x and y .)

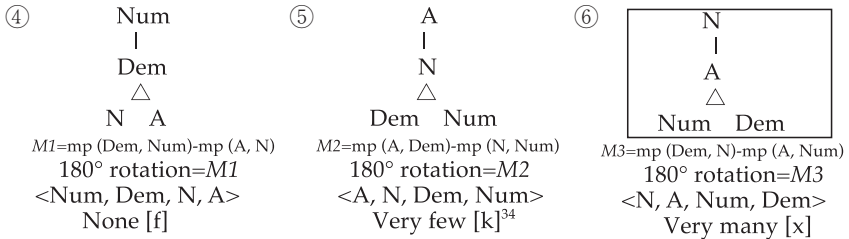


Figure 20 : Rotations of Base DP around M

Why does C_{HL} select the axis $M3$? What is the difference between $M1$, $M2$, and $M3$? $M1$ is obtained by bending $L1$ 45° toward Num , $M2$ by bending $L1$ 45° toward A , and $M3$ by bending $L1$ 45° toward N . Since the base DP is fundamentally a nominal (N), I propose that $M3$, which bends $L1$ 45° toward

N , is the default M axis (the most cost-effective M axis). That is, C_{HL} shows the affinity of $M3$ for N , which is the fundamental lexical property of the base DP. This is the reason why ⑥ is selected over ④ and ⑤. I propose that ⑥ $\langle N, A, Num, Dem \rangle$ is derived from applying MLCA to the base DP structure. Therefore, both LCA and MLCA apply to the base DP. If LCA applies to ①, $\langle Dem, Num, A, N \rangle$ arises. If MLCA applies to ①, $\langle N, A, Num, Dem \rangle$ arises. These word orders are the most cost-effective because the linearing correspondence principle applies to the most cost-effective structure. The DP structure has both the phrasal property and the head property. (Recall that the probability that MLCA applies to the base vP is very small (2.4%)). However small, the fact that the unmarked word order $\langle VOS \rangle$ emerges (the ranking is fourth) indicates that S and O —with V being a head universally—have a strong head property in the languages that have this unmarked order.³⁵ Let us now consider reflections, start from the base DP.

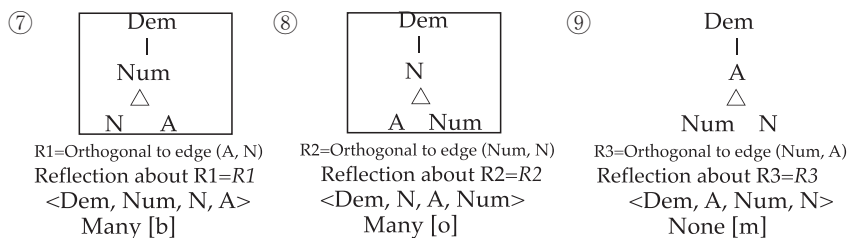


Figure 21 : Applying R to Base DP

The reflections ⑦ and ⑧ are more cost-effective than ② and ③ because the former replace just two positions in the base equilateral triangle, whereas the latter replace all three positions. Why does C_{HL} choose reflection axes $R1$ and $R2$, but not $R3$? What is the difference between $R1$ and $R2$, on the one hand, and $R3$, on the other? I propose that C_{HL} chooses a reflection R that is perpendicular to an edge whose end points constitute a natural class. That is, $\{A, N\}$

(both have the semantic feature *lexical*) and {Num, N} (both are connected to the structural feature $[number] \in \Phi$) constitute natural classes, whereas {Num, A} (Φ and *lexical*) does not. Therefore, C_{HL} selects *R1* and *R2*. C_{HL} allows *A* and *N* to interchange and allows *Num* and *N* to interchange, but it does not allow *Num* and *A* to interchange. It is more cost-effective to select two switching elements within the same natural class, rather than selecting two elements from two distinct classes. Therefore, *R1* and *R2* are more cost-effective axes.

So far, we have considered a simple application of transformations. It is significant that the four major DP-internal unmarked word orders appear in the subgroup of simple applications of the entire group G. Simple applications are more cost-effective, and C_{HL} therefore chooses them. Let us next consider multiple applications of transformations. Multiple applications of non-zero transformations are more costly. C_{HL} avoids them as much as possible. When ② and ③ are followed by an application of *M*, we have the following results.

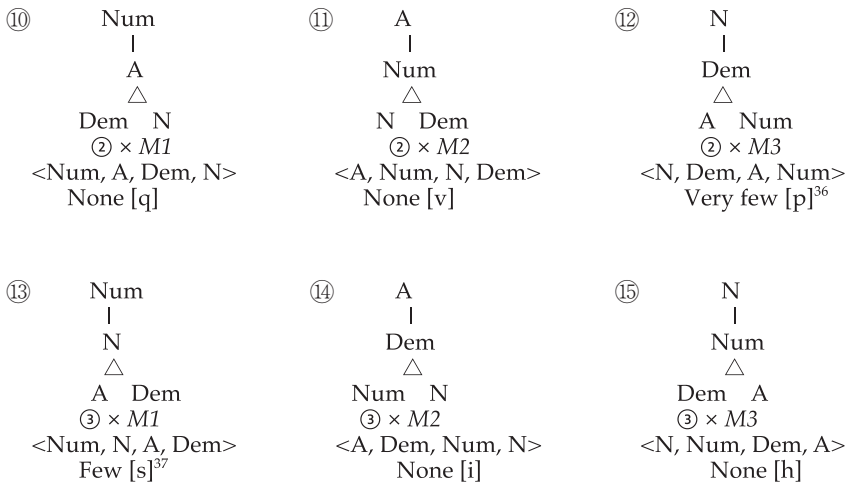


Figure 22 : $L1 (120^\circ) \times M$ and $L1 (240^\circ) \times M$

When simple Rs (⑦, ⑧, and ⑨) are followed by an application of M, we have the following results.

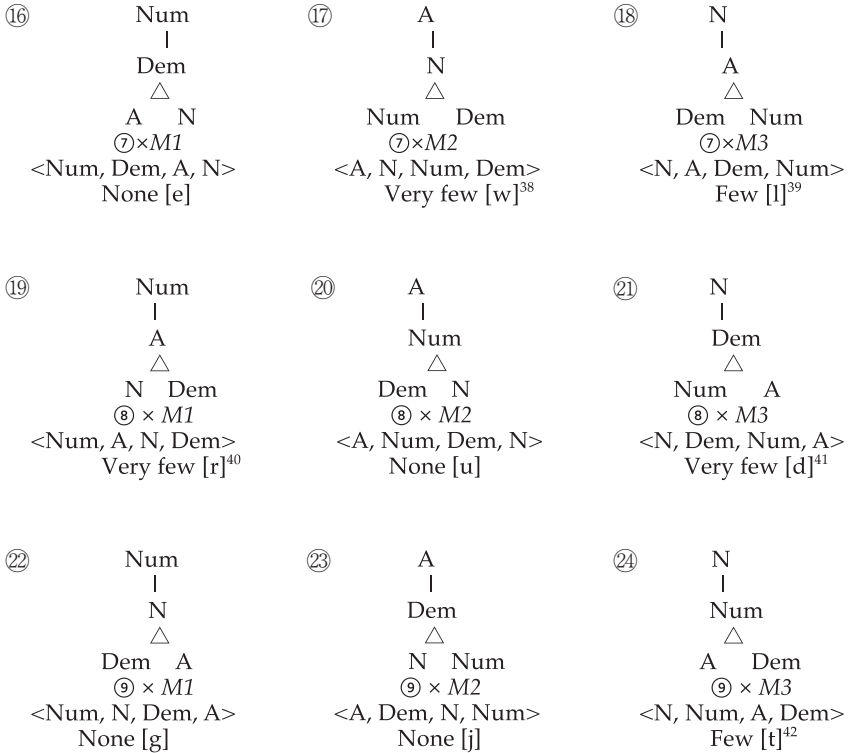


Figure 23 : $R \times M$

It is significant that the four major DP-internal unmarked word orders correspond to the more cost-effective symmetric transformations of a regular tetrahedron with a simple application and the default axes of symmetry. I distinguish necessary (loose and broad) and sufficient (strict and narrow) conditions for the optimal selection of DP-internal word order in C_{HL} as follows:⁴³

(13) *Necessary condition for optimal DP -internal word order*

C_{HL} must choose a single operation.

(14) *Sufficient condition for optimal DP -internal word order*

C_{HL} must choose a cost-effective axis.

The necessary condition comes from the Galois-theory (mathematics) taking cost into account. Within 24 possible DP-internal linear orders, the observed top four come from zero or one, rather than two, applications of an operation and this teaches us something. The sufficient condition comes from linguistic facts that are governed by the Economy Principle (physics). C_{HL} selects more cost-effective axes that are compatible with linguistic facts: the most fundamental nucleus of DP is N , and $\{A, N\}$ and $\{Num, N\}$ form natural classes, while $\{Num, A\}$ does not. Table 6 contains the summary. Assume the cost difference as follows. $I = LI$ (0°) has cost 0, $M3$ cost 1, and $R1/R2$ cost 2. As stated above, these operations involve more cost-effective axes of symmetry. Less cost-effective axes, LI (120°), LI (240°), $M1$, $M2$, and $R3$ have cost 3. Assume addition for cost accumulation.

Word Order and Galois Theory

Transformation	Cost	Principle	Input	Output	Probability
① $I = LI(0^\circ)$	0	LCA	Base DP	<Dem, Num, A, N>	Very many
⑥ $M3$	1	MLCA	Base DP	<N, A, Num, Dem>	Very many
⑦ $R1$	2	LCA?MLCA?	Base DP	<Dem, Num, N, A>	Many
⑧ $R2$	2	LCA?MLCA?	Base DP	<Dem, N, A, Num>	Many
⑬ $LI(240^\circ) \times M1$	6	LCA?MLCA?	Base DP	<Num, N, A, Dem>	Few
⑱ $R1 \times M3$	3	LCA?MLCA?	Base DP	<N, A, Dem, Num>	Few
⑳ $R3 \times M3$	4	LCA?MLCA?	Base DP	<N, Num, A, Dem>	Few
② $LI(120^\circ)$	3	LCA?MLCA?	Base DP	<Dem, N, Num, A>	Very few
③ $LI(240^\circ)$	3	LCA?MLCA?	Base DP	<Dem, A, N, Num>	Very few
⑤ $M2$	3	LCA?MLCA?	Base DP	<A, N, Dem, Num>	Very few
⑫ $LI(120^\circ) \times M3$	6	LCA?MLCA?	Base DP	<N, Dem, A, Num>	Very few
⑰ $R1 \times M2$	5	LCA?MLCA?	Base DP	<A, N, Num, Dem>	Very few
⑲ $R2 \times M1$	5	LCA?MLCA?	Base DP	<Num, A, N, Dem>	Very few
㉑ $R2 \times M3$	3	LCA?MLCA?	Base DP	<N, Dem, Num, A>	Very few
④ $M1$	3		Base DP	<Num, Dem, N, A>	None
⑨ $R3$	3		Base DP	<Dem, A, Num, N>	None
⑩ $LI(120^\circ) \times M1$	6		Base DP	<Num, A, Dem, N>	None
⑪ $LI(120^\circ) \times M2$	6		Base DP	<A, Num, N, Dem>	None
⑭ $LI(240^\circ) \times M2$	6		Base DP	<A, Dem, Num, N>	None
⑮ $LI(240^\circ) \times M3$	4		Base DP	<N, Num, Dem, A>	None
⑯ $R1 \times M1$	5		Base DP	<Num, Dem, A, N>	None
㉒ $R2 \times M2$	5		Base DP	<A, Num, Dem, N>	None
㉔ $R3 \times M1$	6		Base DP	<Num, N, Dem, A>	None
㉕ $R3 \times M2$	6		Base DP	<A, Dem, N, Num>	None

Table 6 : Transformations and Costs for DP-Internal Word Order

The four major unmarked DP-internal word order have cost 2 or lower. There is a tendency that simple operations show higher probability. The ten minorities (few and very few) have the average cost of 4.1 ($41 \div 10$), whereas the unattested ten permutations (none) have the average cost of 5.0 ($50 \div 10$). Hence,

the distinction between possible and impossible permutations has mathematical ground. A question remains: Does C_{HL} employ LCA, MLCA, or both?⁴⁴

5. Conclusion

Unlike Charles Robert Darwin (a British naturalist and biologist; 1809-1882), Alfred Russel Wallace (a British naturalist, explorer, geographer, anthropologist and biologist; 1823-1913), the coauthor of the evolutionary theory of natural selection, was puzzled: The “gigantic development of the mathematical capacity is wholly unexplained by the theory of natural selection, and must be due to some altogether distinct cause,” if only because it remained unused.⁴⁵ Capitalizing on the idea of Leopold Kronecker (a German mathematician; 1823-1891), who said that God (the human DNA, environment and the 3rd factor producing C_{HL}) made integers; all else is the work of man (*Die ganzen Zahlen hat der liebe Gott gemacht, alles andere ist Menschenwerk*), Chomsky states that the theory of natural numbers may have derived from a successor function arising from Merge and that “speculations about the origin of the mathematical capacity as an abstraction from linguistic operations are not unfamiliar.”⁴⁶ Chomsky (2007: 7) proposed the following hypothesis:

(15) Mathematical capacity is derived from language.

If so, Wallace’s puzzle is partially answered: “Some altogether distinct cause” is an operation in C_{HL} . I speculate the following hypothesis.

(16) Equations and sentences share an elementary algebraic structure.

If this is true, we can study C_{HL} with Galois-theoretic tools.⁴⁷ As a Galois group

characterizes the algebraic (or symmetric) structure of an equation, it can also characterize the algebraic (or symmetric) structure of a sentence.

Appendix: Unmarked Word Order as a Galois Group for the Language Equation? – A Speculative Introduction to the Elementary (High-School Level) Algebra of the Human Language Equation – A Toy (Baby) Model

A.1. Human Language Equation?

The algebraic structure of an equation E is equivalent to the Galois group G^g that consists of the roots:

$$(1) \quad E \Leftrightarrow G^g.$$

A radical conjecture follows: A sentence is an expression of a human language equation E_{HL} that C_{HL} solves (Jenkins 2000: 164, 2003), and the algebraic structure of E_{HL} is equivalent to the Galois group G^g_{HL} (of unmarked word orders in C_{HL}):

$$(2) \quad E_{HL} \Leftrightarrow G^g_{HL}.$$

A.2. Sentence as an Equation?

When C_{HL} computes a sentence with S , O , and V , it solves a cubic equation that has three solutions: s , o , and v . The word order patterns are the permutation patterns G^g_{HL} of the roots. A simple transitive sentence must, therefore, have an algebraic structure similar to the following cubic equation.⁴⁸

$$(3) \quad ax^3 + bx^2 + cx + d = 0.$$

If factorization is possible, that is, if E_{HL} is reducible and the reducibility varies according to the field used for factorization, we have

$$(4) \quad (x-s)(x-o)(x-v) = 0.$$

Because

$$(5) \quad x-s = 0, x-o = 0, x-v = 0,$$

we have three solutions, s , o , and v :

$$(6) \quad x = s, o, v.$$

Let us imagine that these are rational numbers, that is, the relevant field consists of rational numbers (putting aside a possible puzzle about what this means).⁴⁹ Expanding the factored cubic equation, we get

$$\begin{aligned} (7) \quad & (x-s)(x-o)(x-v) \\ &= (x^2 - (s+o)x + so)(x-v) \\ &= x^2(x-v) - (s+o)x(x-v) + so(x-v) \\ &= x^3 - vx^2 - (s+o)x^2 - (s+o)x(-v) + sox - sov \\ &= x^3 - (s+o+v)x^2 + (so+ov+vs)x - sov = 0. \end{aligned}$$

The coefficients and constant consist of elementary symmetric polynomials with s , o , and v :

- (8) a. Second order coefficient: $-(s + o + v) = b$
 b. First order coefficient: $(so + ov + vs) = c$
 c. Constant: $-sov$.

Then, the equation in (3) with the roots $\{s, o, v\}$ is equivalent to:

$$\begin{aligned} (9) \quad & ax^3 + bx^2 + cx + d \\ &= (x - s)(x - o)(x - v) \\ &= x^3 - (s + o + v)x^2 + (so + ov + vs)x - sov = 0. \end{aligned}$$

This equation indicates the relationship between the solutions and coefficients. The G^s of an equation is a permutation set of solutions that satisfies the following conditions (Nakamura 2010: 91):

- (10) *Definition of the Galois group G^s of an equation*
- G^s is closed under the multiplication of permutations, and
 - For any rational expression R (with rational coefficients) formed by the solutions, the following holds: the value of R remains the same under all permutations of solutions in $G^s \Leftrightarrow$ the value of R is a rational number.⁵⁰

Condition (10b) maintains that to determine all Rs that have rational values, all one needs to know is the G^s of the equation (ibid.: 91). What is the G^s of an equation in (11)?

$$(11) (x - s)(x - o)(x - v) = 0$$

Elementary algebra tells us the following. Because the value of R is a rational

number, G^g must preserve the same value. In fact, there are Rs whose values remain the same. Such Rs consist of a single root. Assume that there are no multiple roots: that is, $s \neq o \neq v$.

$$(12) R 1 = s, R 2 = o, R 3 = v$$

By definition, G^g should not change the value of R. Thus, G^g should be I alone, in which s changes to s , o changes to o , and v changes to v , i.e., everything remains the same. The other five permutations in which $\langle sov \rangle$ is altered to $\langle svo \rangle$, $\langle osv \rangle$, $\langle ovs \rangle$, $\langle vso \rangle$, or $\langle vos \rangle$ change the values. If E_{HL} were of this type, C_{HL} would produce $\langle SOV \rangle$ only, which is diachronically correct. The ancient C_{HL} may have been solving an E_{HL} that is similar to this equation, in which factorization is possible (reducible), given the rational number field. However, synchronically, this result contradicts the facts about C_{HL} . The current C_{HL} does not solve this type of equation.⁵¹ It follows that the present C_{HL} is solving an E_{HL} that is not reducible if the field consists of rational numbers.⁵²

However, what are $(s + o + v)$, $(so + ov + vs)$ and (sov) ? What do these elementary symmetric polynomials mean for C_{HL} ?⁵³ A polynomial is symmetric when a permutation does not affect it. Let us stipulate that the cubic E_{HL} in (9) has an algebraic structure $G^g = \langle sov \rangle$ with the field of rational numbers. If C_{HL} is solving this equation, it should produce $\langle SOV \rangle$ as the sole possible unmarked word order. This was true for the ancient C_{HL} but not for the current C_{HL} , which solves an E_{HL} with the G^g that includes all six permutations as unmarked word orders.⁵⁴

Let us ask another fundamental question before tackling these questions. What is solving an equation? Solving an equation is the following (Ueno 2011: 50). One starts from symmetric polynomials that consist of coefficients and constant as $(s + o + v)$, $(so + ov + vs)$ and (sov) . These polynomials are sym-

metric in that any permutation does not alter the formulae and the values. One breaks the symmetry little by little.⁵⁵ Finally, the symmetry completely breaks and one obtains the roots, s , o , and v , which are completely asymmetrical; one cannot permute the roots because any permutation will change the values (and the formulae, that is, the roots themselves). This was the starting point of Joseph-Louis Lagrange (a French mathematician, physicist and astronomer born in Italy; 1736-1813) when he took a crucial step forward in solving a conundrum as to why equations of the 5th degree or more resist solutions by a formula. That is, given the general form of $f(x) = x^n + a_{n-1}x^{n-1} + \dots + a_0 = 0$, a formula is a radical expression that is built up from the coefficient a_j by the four basic operations of arithmetic (addition, subtraction, multiplication, division) and n th roots, $n = 2, 3, 4, \dots$ (Stewart 2004: 86). A metaphor of Rubik's Cube works.⁵⁶ A Rubik's Cube with completely random colors (symmetrical state) parallels symmetric polynomials as $(s + o + v)$, $(so + ov + vs)$ and (sov) : any permutation will cause the cube to look like the same as before. In this case, we have an equation of the 6th degree. One breaks the symmetry little by little.⁵⁷ When one obtains consistent colors for each of the 6 planes of the cube, the symmetry is completely broken. That is to say, the consistent-colored 6 planes are the 6 roots of the sextic equation. A random-colored cube is a sextic equation (input) and the final consistent-colored 6 planes express the six roots (output).⁵⁸ On the other hand, we have the opposite situation in solving E_{HL} . We know the roots (output) s , o , and v at the beginning, and are looking for the cubic equation (input). This is an ill-posed (inverse) problem: output is given, but input is unknown.⁵⁹ I hypothesize that E_{HL} shares essentially the same algebraic property as a mathematical equation E . The problem regarding E_{HL} is just ill-posed.⁶⁰

More specifically, we could say that C_{HL} (both ancient and current; at its final state) of native speakers of <SOV>-type languages solves the following

cubic equation, as in (9), which I repeat.

$$\begin{aligned}
 (9) \quad & ax^3 + bx^2 + cx + d \\
 & = (x - s)(x - o)(x - v) \\
 & = x^3 - (s + o + v)x^2 + (so + ov + vs)x - sov = 0.
 \end{aligned}$$

We know that the Galois group G^g of this equation is $\langle sov \rangle$ only. In a sense, the coefficients and constant as symmetric polynomials express the scrambling property (higher level of symmetry) of $\langle SOV \rangle$ -type languages.⁶¹

It is worth noting that the number of argument (the minimum information that is necessary for the event denoted by the predicate to hold) is at most 3, namely, the subject (s), the indirect object (io), and the direct object (do). If we include the verb v in the equation, C_{HL} is solving equations of the 2nd, 3rd, or 4th degree. There is no equation of the 5th degree or more for C_{HL} . This is reminiscent of a mathematical fact that there is no formula for equations of 5th degree or more, the explanation of which Galois finalized about 200 years ago.⁶²

A.3. Is E_{HL} Linear (1st Degree Polynomial)?

Suppose that E_{HL} has an algebraic structure similar to that of a linear equation such as

$$(13) \quad x - s = 0.$$

Then, the only root is s :

$$(14) \quad x = s.$$

Elementary algebra indicates the following. To permute s , one must permute it by itself (I). The G_{HL}^g of E_{HL} must consist of I alone and C_{HL} would produce only $\langle S \rangle$. However, this contradicts the facts about C_{HL} as would an algebraic structure based on other linear equations such as $x - o = 0$ and $x - v = 0$. There is no natural human language with an unmarked word order such as $\langle S \rangle$ alone, $\langle O \rangle$ alone, or $\langle V \rangle$ alone.⁶³ Therefore, E_{HL} cannot be a linear equation.

A.4. Is E_{HL} Quadratic (2nd Degree Polynomial)?

A.4.1 If s and v are in $\mathbb{Q} \dots$ ⁶⁴

Suppose that the algebraic structure of E_{HL} with the roots $\{s, v\}$ is similar to that of the following quadratic equation:

$$(15) \quad x^2 + 3x - 4 = 0.$$

Given the set of two solutions $\{a, b\}$, the G^g of (15) would consist of the identity permutation I alone.⁶⁵ Elementary algebra indicates the following. Suppose that the permutation $K = (sv)$ were in the relevant G_{HL}^g .⁶⁶ If we perform K on $(s - v)$, $(s - v)$ changes to $(v - s) = -(s - v)$. That is, K changes the value of R . Therefore, G_{HL}^g must not contain K . On the other hand, I does not change the value of $R = (s - v)$. If the structure of E_{HL} were similar to this type of quadratic equation, G_{HL}^g with the two solutions $\{s, v\}$ would contain I alone. If we start from the base VP in which S c-commands V and stipulate that the base VP is the identity element, it would follow that C_{HL} exhibits only $\langle SV \rangle$, since I changes $\langle SV \rangle$ to $\langle SV \rangle$. This conclusion is not empirically correct, however. When V is intransitive, the present C_{HL} shows both unmarked orders.⁶⁷

(16) a. <SV> (79.7%)

An example: *English*

The child ran.

b. <VS> (13.0%)

An example: *Tagalog*

Tumakbo ang bata.

ran *ANG child*

‘The child ran.’

The present C_{HL} does not solve a quadratic equation in which factorization is possible and the roots are like rational numbers. There is a remaining puzzle that why do <SV> languages outnumber <VS> languages? It might be economical to apply LCA, rather than MLCA, to the base VP, as we saw in the case of the base vP for {S, O, V}.

A.4.2 If s and v are not in @ ...

Suppose that E_{HL} has an algebraic structure similar to the following quadratic equation:

$$(17) \quad x^2 + 3x + 1 = 0.$$

Factorization is not possible. The roots are irrational numbers. Given the two solutions $\{a, b\}$, the G^g is $\langle ab \rangle$ and $\langle ba \rangle$.⁶⁸ Elementary algebra indicates the following. Because we have two roots, s and v , there are two possible candidates for G^g_{HL} : I and $K = (sv)$. Suppose that G^g_{HL} contained I and K . Would G^g_{HL} satisfy condition (10a) (that is, would G^g_{HL} closed under the multiplication

operation)? As

$$(18) I \times K = K, K \times I = K, K \times K = I, I \times I = I,$$

G_{HL}^g would be closed under multiplication. Would G_{HL}^g satisfy condition (10b)? An example of R is the difference product $\Delta = (s - v)$. Given that $D = \Delta^2$ and $D = 5$, we have

$$(19) \Delta = (s - v) = \pm\sqrt{5}.$$

However, the positive and negative square roots of 5 are not rational numbers. If we are in the rational number field, the value of $R = \Delta$ does not exist in this field.⁶⁹ Therefore, G_{HL}^g would contain a permutation that changes the value of $R = \Delta$. If G_{HL}^g contained only I , G_{HL}^g would not change the value of Δ . G_{HL}^g must contain a permutation other than I ; that is, G_{HL}^g must additionally contain K . With K , $R = \Delta = + (s - v)$ changes to $R' = \Delta' = (v - s) = -(s - v)$. The plus sign of $R = \Delta$ has changed to a minus sign. By I , $R = \Delta = + (s - v)$ remains the same. If G_{HL}^g contained I and K , G_{HL}^g would contain $\langle sv \rangle$ and $\langle vs \rangle$. This is empirically true, as we saw in (16). Hence, the present C_{HL} solves a quadratic equation that has the same type of algebraic structure as (17) with two irrational number roots.

A.5. Is E_{HL} Cubic (3rd Degree Polynomial)?

A.5.1 If s , o , and v are in \mathbb{Q} ...

Suppose that C_{HL} is solving an E_{HL} with a structure that is similar to the following equation:

$$(20) x^3 - x = 0.$$

By factorization,

$$(21) x^3 - x = x(x^2 - 1) = x(x - 1)(x + 1) = 0.$$

This is not a genuine cubic equation because it consists of first degree parts. The calculation cost must be cheap. The three roots are three distinct rational numbers:

$$(22) x = 0, 1, -1.$$

Elementary algebra tells us the following. Let the three roots be s , o , and v . Thus,

$$(23) x = 0 (= s), 1 (= o), -1 (= v).$$

Consider the following difference product Δ as an example of R :

$$(24) R = (s - o)(s - v)(o - v) = -1 \cdot 1 \cdot 2 = -2.$$

Provided that the value of R is a rational number, $G_{\text{HL}}^{\mathbb{R}}$ must exclude the permutation that changes the value of R . Thus, $G_{\text{HL}}^{\mathbb{R}}$ contains I . What about $fI = (ov)$, which exchanges o and v ? fI transforms R as follows:

$$(25) fI: (s - o)(s - v)(o - v) \rightarrow (s - v)(s - o)(v - o) = 1 \cdot -1 \cdot -2 = 2.$$

fI changes the value of R . $G_{\text{HL}}^{\mathbb{R}}$ does not contain fI . What about $rI = (svo)$,

which changes s to v , v to o , and o to s ? rI transforms R as follows:

$$(26) \ rI: (s - o) (s - v) (o - v) \rightarrow (v - s) (v - o) (s - o) = -1 \cdot -2 \cdot -1 = -2.$$

rI does not change the value of R . G_{HL}^g might contain rI . However, we must consider all possible R s. If there is an R whose value is altered by rI , then G_{HL}^g does not contain rI . In fact, rI alters the values of the following R s:

$$(27) \ s = 0, \ o = 1, \ v = -1.$$

Therefore, G_{HL}^g does not contain rI . Only I preserves symmetry (the values remain the same). If C_{HL} were solving this type of a cubic equation, it would produce only <SOV> languages. This might be true diachronically, but not synchronically. The ancient C_{HL} might have been solving a cubic equation where factorization is possible and the roots are like rational numbers.

A.5.2 If s is in \mathbb{Q} , and $\{o, v\}$ are not in \mathbb{Q} ...

Suppose that C_{HL} is solving an E_{HL} with a structure that is similar to that of the following equation:

$$(28) \ x^3 + 3x^2 + x = 0.$$

By factorization, we obtain

$$(29) \ x(x^2 + 3x + 1) = 0.$$

The three roots are one rational number and two irrational numbers:

$$(30) \ x = 0, \frac{-3 + \sqrt{5}}{2}, \frac{-3 - \sqrt{5}}{2}.$$

The calculated cost of (28) must be higher than that of (20). Given the three roots $\{a, b, c\}$, where $a = 0$, the G^g of (28) contains the identity permutation I and the permutation (bc) , which switches the two irrational roots, b and c .⁷⁰ Elementary algebra tells us the following. Let us stipulate that the three roots are s , o , and v :

$$(31) \ s = 0, \ o = \frac{-3 + \sqrt{5}}{2}, \ v = \frac{-3 - \sqrt{5}}{2}.$$

Since s is rational, the G_{HL}^g contains the identity operation $I (= r0)$ and $fI = (ov)$, which switches o and v . I produces Δ as follows:

$$(32) \ \Delta = (s - o) (s - v) (o - v) = R.$$

fI produces Δ as follows:

$$(33) \ \Delta = (s - v) (s - o) (v - o) = -(s - o) (s - v) (o - v) = -R.$$

I and fI produce difference products that have distinct values (the plus symbol in $+R$ changes to a minus in $-R$). The Δ s being irrational numbers, G_{HL}^g contains a permutation that changes the value of R .⁷¹ Therefore, the G_{HL}^g must contain fI . I corresponds to LCA mapping the c -command relation $S \gg O \gg V$ to the linear order $\langle SOV \rangle$ (48.5% of languages), and fI corresponds to LCA mapping $S \gg V \gg O$ to $\langle SVO \rangle$ (38.7% of languages). The present C_{HL} is very close to solving this type of cubic equation.⁷²

**A.5.3 If E_m were a Genuine Cubic Equation with the Three Roots
not in $\mathbb{Q} \dots$**

Suppose that C_{HL} is similar to the following:

$$(34) \quad x^3 - 3x + 1 = 0.$$

As this is a real cubic equation, factorization is not possible. The calculation cost must be the highest. Elementary algebra tells us the following. Let us postulate that the three roots are s , o , and v . The difference product Δ is

$$(35) \quad \Delta = \pm (s - o)(s - v)(o - v).$$

Given a cubic equation in the more general form: $x^3 + px + q = 0$, the coefficients are $p = -3$ and $q = 1$. The formula yields $\Delta = \pm 9$.⁷³ The values of Δ are rational numbers. By (10b), G_{HL}^g must not change the value of Δ . Among the six permutations, three flips, namely $f1 = (ov)$, $f2 = (so)$, and $f3 = (sv)$, change the value of Δ . Therefore, G_{HL}^g excludes $f1$, $f2$, and $f3$.⁷⁴ This leaves us with I , $r1 = (svo)$, and $r2 = (sov)$. I , $r1$, and $r2$ are recurring permutations (rotations) in which all three members are affected. Calculating, we see that I , $r1$, and $r2$ do not change the value of R .⁷⁵ Therefore, G_{HL}^g contains these rotations. However, we do not yet know whether all or only some of them constitute G_{HL}^g . Let us use one root, s , as one of the simplest possible examples of R . We consider the fact that the value of s is an irrational number. G_{HL}^g must contain a permutation that changes the value of s . First, consider I . By applying I to s , the value of R remains the same (s). By applying $r1 = (svo)$ to s , the value of R changes from s to v . By applying $r2 = (sov)$ to s , the value of R changes from s to o . Therefore, G_{HL}^g contains $r1$ and $r2$. Provided that

G_{HL}^g is closed under multiplication, G_{HL}^g must contain I as well because $rI^2 = r^2$ and $rI^3 = I$ and hence, the G_{HL} should select $\langle SOV \rangle$, $\langle VSO \rangle$, and $\langle OVS \rangle$ as the major unmarked word orders. However, this contradicts the facts about C_{HL} . Therefore, the present C_{HL} does not solve a real cubic equation, which resists factorization and has three irrational number roots.

A.5.4 If E_{HL} is a Cubic Equation with G^g that Includes All Six Permutations ...

The G^g of the following equation includes all six permutations (Lieber 1932, Nakamura 2011) :

$$(36) \quad x^3 - 2 = 0.$$

The G^g corresponding to (36) is as follows:⁷⁶

$$(37) \quad G^g = \{ \langle SOV \rangle, \langle SVO \rangle, \langle VSO \rangle, \langle VOS \rangle, \langle OVS \rangle, \langle OSV \rangle \}$$

This G^g is the maximum of all G^g s of cubic equations. Given that the present C_{HL} permits all six unmarked word orders, the present C_{HL} is solving an E_{HL} of this type. The initial state of the current C_{HL} solves an E_{HL} that is similar to (36). When the parameter setting takes place under the $\langle SOV \rangle$ environment, the final state of C_{HL} would be specialized to solve an E_{HL} that is similar to (20), $x^3 - x = 0$, for which the G^g includes only $\langle sov \rangle$. However, a puzzle then remains as to why $\langle SOV \rangle$ and $\langle SVO \rangle$ emerge in almost the same percentage of languages and make up more than 80% of all the six possible unmarked word orders.

A.6. Summary

Let us summarize the typology of the possible algebraic structures ($G^?$) of E_{HL} . I include a possible E_{HL} for DP-internal word order (Section 4). Notes for abbreviations in the table are as follows. The identity operation I : Do nothing to $\langle a \rangle$, $\langle ab \rangle$, $\langle abc \rangle$, $\langle abcd \rangle$, $\langle s \rangle$, $\langle o \rangle$, $\langle v \rangle$, $\langle sv \rangle$ (the base VP), $\langle sov \rangle$ (the base vP), and $\langle \text{Dem, Num, A, N} \rangle$ (the base DP) = $\langle abcd \rangle$, where $a = \text{Dem}$, $b = \text{Num}$, $c = \text{A}$, $d = \text{N}$. #: order of equation. $C_{\text{HL.1}}$: Ancient C_{HL} , $C_{\text{HL.2}}$: Current C_{HL} . *: unattested, \checkmark : attested. Rational field: a field that consists of rational numbers; real field: a field that consists of real numbers.

Mathematical Algebraic Structure					Linguistic Algebraic Structure						
#	E	Reduction	Root	G^s	E_{HL}	Reduction	Root	G^s_{HL}	C_{HL1}	C_{HL2}	
1	$x-p=0$		p	$\langle p \rangle$	$x-s=0$		s	$\langle s \rangle$	*	*	
					$x-o=0$		o	$\langle o \rangle$	*	*	
					$x-v=0$		v	$\langle v \rangle$	*	*	
2	$x^2+3x+1=0$		a	$\langle ab \rangle$	$x^2+px+q=0$		s	$\langle sv \rangle$?	✓	
			b	$\langle ba \rangle$			v	$\langle vs \rangle$			
				rational field				rational field			
				$\langle ab \rangle$				$\langle sv \rangle$?	*	
				real field				real field			
2	$x^2+3x-4=0$	$(x+4)(x-1)=0$	a=-4	$\langle ab \rangle$	$x^2-(s+v)x+sv=0$	$(x-s)(x-v)=0$	s	$\langle sv \rangle$?	*	
			b=1				v				
3	$x^3-x=0$	$x(x-1)(x+1)=0$	a=0	$\langle abc \rangle$	$x^3-(o+v)x^2+ovx=0$	$x(x-o)(x-v)=0$	s=0	$\langle sov \rangle$	✓	*	
			b=1				o				
		c=-1		v							
3	$x^3+3x^2+x=0$	$x(x^2+3x+1)=0$	a=0	$\langle abc \rangle$	$x^3+px^2+qx=0$	$x(x^2+px+q)=0$	s=0	$\langle sov \rangle$	*	✓	
			b	$\langle acb \rangle$				o			$\langle svo \rangle$
			c				v		86%		
									?		
3	$x^3-3x+1=0$		a	$\langle abc \rangle$	$x^3+px+q=0$		s	$\langle sov \rangle$	*	*	
			b	$\langle cab \rangle$				o			$\langle vso \rangle$
			c	$\langle bca \rangle$				v			$\langle ovs \rangle$
3	$x^3-2=0$		a	$\langle abc \rangle$	$x^3+p=0$		s	$\langle sov \rangle$	*	✓	
			b	$\langle acb \rangle$				o			$\langle svo \rangle$
			c	$\langle cab \rangle$				v			$\langle vso \rangle$
				$\langle cba \rangle$							$\langle vos \rangle$
				$\langle bca \rangle$							$\langle ovs \rangle$
				$\langle bac \rangle$							$\langle osv \rangle$
4	$x^4-4x^2-5=0$	$(x^2+1)(x^2-5)=0$	a=i	$\langle abcd \rangle$	$x^4+px^2+q=0$	$(x^2+r)(x^2+t)=0$	a	$\langle abcd \rangle$?	*	
			b=-i	$\langle bacd \rangle$				b			$\langle bacd \rangle$
			$c=\sqrt{5}$	$\langle abdc \rangle$			c	$\langle abdc \rangle$			
			$d=-\sqrt{5}$	$\langle badc \rangle$			d	$\langle badc \rangle$			
4	?		a	$\langle abcd \rangle$?		a	$\langle abcd \rangle$?	✓	
			b	$\langle dcba \rangle$				b			$\langle dcba \rangle$
			c	$\langle abdc \rangle$				c			$\langle abdc \rangle$
			d	$\langle adcb \rangle$				d			$\langle adcb \rangle$

Table 1 : Typology of E and E_{HL}

The $(x^3 + 3x^2 + x = 0)$ -as-similar-to- E_{HL} hypothesis can explain the fundamental asymmetry of unmarked word orders ($\langle SOV \rangle$ (an average of about 45%) and $\langle SVO \rangle$ (an average of about 37%)), whereas it fails to explain all six permutations are available for C_{HL} . On the other hand, the $(x^3 - 2 = 0)$ -as-similar-to- E_{HL} hypothesis can explain the fact that all six permutations are available for C_{HL} , whereas it fails to explain the fundamental asymmetry. The current C_{HL} must be solving E_{HL} that is similar both to $(x^3 + 3x^2 + x = 0)$ and to $(x^3 - 2 = 0)$. What is it? Appendix is summarized as follows.

- (38) a. Unlike the current C_{HL} , the ancient C_{HL} might have been solving an E_{HL} such as $(x - s)(x - o)(x - v) = x^3 - (s + o + v)x^2 + (so + ov + vs)x - sov = 0$ (in which factorization is possible and the roots are rational), which produces $\langle sov \rangle$ only.
- b. E_{HL} cannot be a linear equation.
- c. The present C_{HL} does not solve a quadratic equation in which factorization is possible and the roots are rational numbers.
- d. The present C_{HL} solves a quadratic equation in which factorization is impossible and the roots are irrational numbers.
- e. The present C_{HL} uses the real number field for s and v , whereas rational number field for o and v . This is the reason why $\langle SV \rangle$ (79.7%) outnumbers $\langle VS \rangle$ (13%), while the ratios for $\langle OV \rangle$ (46.9%) and $\langle VO \rangle$ (46.4%) are almost the same.
- f. Unlike the current C_{HL} , the ancient C_{HL} might have been solving a cubic equation where factorization is possible and the roots are rational numbers.
- g. The present C_{HL} is close to solving a cubic equation that consists of linear and quadratic equations (factorization is impossible and there are two irrational roots). The G_{HL}^g of such an E_{HL} includes I and fI

= (vo) , which explains the facts that 48.5% of languages are <SOV> and 38.7% of languages are <SVO>.

- h. The factorized E_{HL} structure consisting of simple and quadratic parts expresses the algebraic structure of the base vP: $E_{HL} = [_{VP} x [_{VP} (x^2 + px + q)] = 0$. The vP edge (the rational root) constitutes s , and the quadratic equation part of the VP (the two irrational roots) constitutes v and o .
- i. The present C_{HL} does not solve a real cubic equation that resists factorization and has three irrational number roots.
- j. The initial state of the current C_{HL} solves an E_{HL} that is similar to $x^3 - 2 = 0$, whose G^g includes all six permutations. If the parameter setting occurs under an <SOV> environment, the final state of C_{HL} would be specialized to solve an E_{HL} that is similar to $x^3 - x = 0$, for which the G^g includes only <SOV>. However, the puzzle then remains as to why <SOV> and <SVO> emerge in almost the same percentage of languages and make up more than 80% of all the six possible unmarked word orders.

References

- Arikawa, Koji. 2011. Flexible Command: A Solution to the Symmetry Problem of Adjunction, Scrambling, and Dislocation. *Human Sciences Review* 40: 65-200. Osaka: St. Andrew's University.
- Arikawa, Koji. 2012a. Some Problems of the Island Explanation Based on the Minimize Chain Link Principle (Shortest Movement Condition) and the Uniformity Corollary on Adjunction. *Human Sciences Review* 42. Osaka: St. Andrew's University.
- Arikawa, Koji. 2012b. Does Word Order Asymmetry Have Mathematical

Grounds? ms. St. Andrew's University.

- Armstrong, Mark A. 1988. *Groups and Symmetry*. New York: Springer-Verlag.
- Barrie, Michael, J. M. 2006. Dynamic Antisymmetry and the Syntax of Noun Incorporation. University of Toronto dissertation.
- Boeckx, Cedric. 2009. The Nature of Merge: Consequences for Language, Mind, and Biology. In Piattelli-Palmarini, M., Pello Salaburu, and Juan Uriagereka (eds.) *Of Minds and Language: A Dialogue with Noam Chomsky in the Basque Country*. 44-57. Oxford University Press.
- Boeckx, C., J. D. Fodor, L. Gleitman, & L. Rizzi. 2009. Round Table: Language Universals: Yesterday, Today, and Tomorrow. In Piattelli-Palmarini, M., Pello Salaburu, and Juan Uriagereka (eds.) *Of Minds and Language: A Dialogue with Noam Chomsky in the Basque Country*. 195-222. Oxford University Press.
- Butterworth, Brian. 2000. *The Mathematical Brain*. London: Macmillan.
- Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, Mass: MIT Press.
- Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 2000. Minimalist Inquiries: The Framework. In Roger Martin, David Michaels & Juan Uriagereka (eds.) *Step by Step – Essays on Minimalist Syntax in Honor of Howard Lasnik*, 89-155. Cambridge, MA: MIT Press.
- Chomsky, Noam. 2005. Three Factors in Language Design. *Linguistic Inquiry* 36: 1-22.
- Chomsky, Noam. 2007. Approaching UG from Below. In Uli Sauerland and Hans-Martin Gärtner (eds.) *Interfaces + Recursion = Language?* 1-29. Berlin and New York: Mouton de Gruyter.
- Chomsky, Noam. 2009. Opening Remarks. In Massimo Piattelli-Palmarini, Juan Uriagereka, and Pello Salaburu (eds.) *Of Minds & Language: A Dialogue*

- with Noam Chomsky in the Basque Country*. Oxford University Press.
- Chomsky, Noam. 2010. Some Simple Evo-Devo Theses: How True Might They Be For Language? In R. K. Larson, V. Déprez, and H. Yamakido (eds.) *The Evolution of Human Language: Bilingual Perspectives*. 45-62. Cambridge: Cambridge University Press.
- Chomsky, Noam & James McGilvray. 2012. *The Science of Language: Interviews with James McGilvray*. Cambridge University Press.
- Cinque, Guglielmo. 2005. Deriving Greenberg's Universal 20 and its Exceptions. *Linguistic Inquiry* 36: 315-332.
- Dryer, Matthew S. & Haspelmath Martin (eds.). 2011. The World Atlas of Language Structures Online. Munich: Max Planck Digital Library. <http://wals.info/> (8 October 2011).
- Elouazizi, Noureddine & Martina Wiltschko. 2006. The Categorical Status of (Anti-) (Anti-) Agreement. In Donald Baumer, David Montero & Michael Scanlon (eds.), *Proceedings of the 25th West Coast Conference on Formal Linguistics*: 150-158. Somerville, MA: Cascadilla Proceedings Project.
- Fukui, Naoki & Yuji Takano. 1998. Symmetry in Syntax: Merge and Demerge. *Journal of East Asian Linguistics* 7: 27-86.
- Gell-Mann, Murray & Merritt Ruhlen. 2011. The Origin and Evolution of Word Order. *Proceedings of the National Academy of Sciences of the United States of America*, 108 (42) : 17290-17295.
- Greenberg, Joseph. 1963. Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements. In Joseph Greenberg (ed) *Universals of Language*, 73-113. Cambridge, Mass.: MIT Press.
- Greenberg, Joseph. 1966. Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements. In Joseph Greenberg (ed) *Universals of Language*, 2nd edition. 73-113. Cambridge, Mass. : MIT Press.
- Grohmann, Kleanthes, K. 2000. Prolific Peripheries: A Radical View from the

- Left. Ph.D. thesis, University of Maryland, College Park.
- Grohmann, Kleanthes, K. 2011. Anti-Locality: Too-Close Relations in Grammar. In Cedric Boeckx (ed.) *The Oxford Handbook of Linguistic Minimalism*, 260-290. Oxford University Press.
- Jenkins, Lyle. 2000. *Biolinguistics: Exploring the Biology of Language*. Cambridge, UK: Cambridge University Press.
- Jenkins, Lyle. 2003. Language and Group Theory. A lecture for a graduate-school class on biolinguistics by Piattelli-Palmarini Massimo at MIT.
- Jurka, Johannes. 2010. The Importance of Being a Complement: CED Effect Revisited. University of Maryland dissertation.
- Kayne, Richard S. 1994. *The Antisymmetry of Syntax*. Cambridge, MA: MIT Press.
- Kamei, Takashi, Rokuro Kono & Eiichi Chino (eds.). 1989. *Sanseido's Linguistic Encyclopedia*, Vol. 2. Tokyo: Sanseido.
- Kim, Jung Myeong. 2011. *13-Sai-no Musume-ni Kataru Garoa-no Suugaku* [Galois Mathematics that I Teach to My 13-Year-Old Daughter]. Tokyo: Iwanami Syoten.
- Larson, Richard. 1988. On the Double Object Construction. *Linguistic Inquiry* 19: 335-91.
- Lebeaux, David. 1988. Language Acquisition and the Form of the Grammar. Amherst, MA: University of Massachusetts dissertation.
- Lieber, Lillian R. 1932. *Galois and the Theory of Groups: A Bright Star in Mathematics*. New York: The Galois Institute of Mathematics and Art.
- Longa, Víctor, M., Guillermo Lorenzo, and Juan Uriagereka. 2011. Minimizing Language Evolution: The Minimalist Program and the Evolutionary Shaping of Language. 595-616. In Cedric Boeckx (ed.) *The Oxford Handbook of Linguistic Minimalism*. Oxford University Press.
- Nakamura, Akira. 2010. *Garoa no Riron: Hooteeshiki wa Naze Tokenakattanoka* 'Galois Theory: Why the Equation Could Not Be Solved.' Tokyo: Kodan-

sya.

Pereltsvaig, Asya. 2011. Was Proto-Human an SOV Language?. *Language of the World*.

<http://languagesoftheworld.info/historical-linguistics/was-proto-human-an-sov-language.html> (25 February 2012)

Richards, Norvin. 1997. *What Moves Where When in Which Language?* Cambridge, MA: MIT dissertation.

Rizzi, Luigi. 1990. *Relativized minimality*. Cambridge, MA: MIT Press.

Ross, John Robert. 1967. *Constraint on Variables in Syntax*. Massachusetts Institute of Technology dissertation. [Published in 1986 as *Infinite Syntax!*. Norwood, NJ: ALEX.]

Stewart, Ian. 1975. *Concepts of Modern Mathematics*. New York: Dover Publications.

Stewart, Ian. 2004. *Galois Theory 3rd edition*. Chapman & Hall/CRC.

Stewart, Ian. 2007. *Why Beauty is Truth: The Story of Symmetry*. New York: Basic Books.

Strang, Gilbert. 2003. *Introduction to Linear Algebra*. Wellesley, MA: Wellesley-Cambridge Press.

Suzuki, Wataru. 2002. Gengo Ruikee-no Junkan-teki Kaisoossee [The Cyclic Hierarchy of Language Types]. *Journal of Nagoya Women's University (Humanities and Social Science)* 48: 249-262. Nagoya.

Ueno, Kenji. 2011. Garoa-no Kangaeta Koto [Things that Galois Thought]. *Gendai Shisoo (revue de la pensée d'aujourd'hui)* Vol. 39-5, 38-58). Tokyo: Seedosya.

Uriagereka, Juan. 2012. *Spell-Out and the Minimalist Program*. New York: Oxford University Press.

Wallace, Alfred Russel. 1889. *Darwinism*. London: Macmillan.

Yamamoto, Hideki. 2002. *Sekai Shogengo no Gojun Ruikee Kenkyuu niokeru*

Syomondai: Chiriteki, Keetooteki Gojun Bunpu nitaisuru Kenkyuu Josetsu 'Issues in Word Order Typology on the World's Languages: An Introduction to the Study of Areal and Genealogical Distribution of Word Order.' *Studies in the Humanities, Cultural Sciences*, 7: 79-101. Aomori, Japan: Hirosaki University.

Yang, Charles, D. 2002. *Knowledge and Learning in Natural Language*. Oxford University Press.

Yuki, Hiroshi. 2012. *Suugaku Gaaru – Garoa Riron* [Mathematics Girl – Galois Theory]. Tokyo: Softbank Creative.

Notes

I would like to thank Makoto Toma (mathematics, St. Andrew's University) for his valuable comments and suggestions. Without his constructive criticism regarding my amateurish mathematics, I could not have realized this article. Further, I thank Piattelli-Palmarini Massimo (physics and biology of language, University of Arizona) for allowing me to join his class on biolinguistics at MIT in 2003, which marked the beginning of this project. I would also like to thank Lyle Jenkins (Biolinguistics Institute, Boston, USA) for the insightful lecture on human language and Galois theory in Massimo's class and for taking the time to listen to my idea in an on-campus café. I submitted a paper on related topics to *Biolinguistics* in November 2011, and after one revision and a year-long reviewing process, the paper was finally rejected (as of October 31, 2012). I would like to thank the editor Kleanthes K. Grohmann and the two anonymous reviewers (a computer scientist and a group theorist) for their constructive criticism and suggestions. I am grateful to Lyle who still encourages me to continue this project. All remaining errors are my own.

- 1 I thank an anonymous reviewer (a group theorist) who expressed concern that my approach may be too simple, immature, groundless, and without promise, and that my research has a long way to go even if it should turn out to be tenable. The reviewer pointed out several fatal faults. First, S_3 and S_4 are too simple to say anything about general patterns. Second, since one can superficially analyze any permutation phenomenon by means of the group theory, there is no substance to the argument that C_{III} works group theoretically. The reviewer advised me to write this speculative paper without claiming to present any scientific findings, at least raise a set of good questions. I hope that this version manages to do that. Given this honest criticism by a pure mathematician, as a biolinguist and a mathematical amateur, I might be offering a groundless metaphor even from the viewpoint of applied mathematics.
- 2 Merge has at least four characteristics: binarity, asymmetric labeling, structural preservation ('extension' and 'no-tampering' conditions), unboundedness, and flexibility (long-distance dependency) (Longa et al. 2011: 599). I propose that G axioms derive these properties. The closure axiom derives the unboundedness. Suppose that terms are like \mathbb{N} (natural numbers) and merge is like addition. Addition is closed over \mathbb{N} , which is discrete and infinite. Merge is closed over terms, which

is discrete and infinite. The associativity axiom derives binarity (two at a time), asymmetric labeling (no double-headedness), and structural preservation (no tampering over two elements that are already computed). The inverse (reversibility) axiom derives flexibility. Copy a term, and remerge the copy. The moved copy is free to lower (reconstruct) to the long-distant original position. Boeckx (2009: 48) proposes $\text{Merge} = \{X, Y\} + \text{Copy}$, where $\{X, Y\}$ is set formation and Copy brings labeling, which gives endocentricity. Chomsky, on the other hand, proposes $\text{Merge} = \{X, Y\}$. Merge cannot be decomposed. For Chomsky, what matters is unbounded Merge (ibid. 52). Note that basic operations of linear algebra include Copy. Suppose v and w are vectors. In vector addition, you place the start of the copy of w at the end of v . In scalar multiplication, you copy v in $2v = v \times v$. See Strang (2003: 1-3).

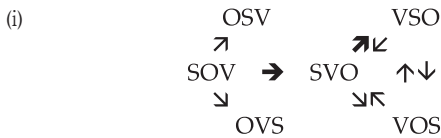
- 3 I thank an anonymous reviewer (a group theorist) for urging me to distinguish two separate issues: (a) the explanation of the permutation group of $\{S, V, O\}$ and (b) a proof that a set of structural relations inherently has the property of mathematical groups. This section concerns issue (b). I examine issue (a) in Sections 2 and Appendix.
- 4 Yuki (2012: 76). Or, $a \bullet b = c$, where \bullet is an operation and a, b , and c are arbitrary members of G .
- 5 The c -command relation plays an important role in C_{III} (Uriagereka 2012: 121).
- 6 See Chomsky (1995: 339) for the intuition that c -command expresses a balance between connection and disconnection among the terms in a tree.
- 7 This is internal Merge $\{X, Y\}$, where X is part of Y . More basically, in external Merge, X is external to Y . External Merge selects building blocks from the lexicon (outside of the structure-building space), whereas internal Merge selects blocks from inside the existing structure. Since $\text{internal Merge (Move)} = \text{external Merge} + \text{Copy} + \text{Rermerge}$, external Merge is more cost-effective than internal Merge (Move) (this is the Merge-over-Move hypothesis).
- 8 An anonymous reviewer raised a question about what happens with VP-fronting as in
 - (i) $[_{\text{VP}} \text{ Love Mary}]_i$, John did t_i .

The reviewer pointed out that in (i), not all terms stand in c -command relation and hence this common example lies outside the bounds of my system. The fact is that the VP as a whole merges with the CP. I propose that C_{III} treats the fronted

VP as O and the example exhibits the derived order [OSV]. Merge operation copies the original VP (= O) and the copy remerges with the CP. C_{III} does not see the internal structure of the VP after the VP-fronting. Alternatively, assuming Uriagereka’s Multiple-Spell-Out (MSO) hypothesis, we can say that the VP, being an adjunct, undergoes spell-out before the entire CP does, and that everything inside the first-spelled-out VP precedes the second-spelled-out CP. In that every resulting term is related by the c-command relation. I thank the reviewer for pointing out this possible problem.

- 9 C_{III} motivates the base vP. I thank an anonymous reviewer for clarification.
- 10 I thank an anonymous reviewer for pointing out this crucial question. In an earlier draft, I adopted the view that O moves to vP Spec for feature checking. The reviewer pointed out that such a vP competes in cost with the one in which V moves to v, that is, both structures have one internal merge. The reviewer’s observation has improved the structure of the base vP; it is constructed by an external merge alone, which yields the simplest possible architecture for S, O, and V.

Gell-Mann and Ruhlen’s (2011) proposal supports the analysis. Scrutinizing known word order change, Gell-Mann and Ruhlen proposed “natural drift,” as in (i), which explains evolution of word order. The heavy lines indicate “the most frequent changes caused by natural drift without diffusion” and the other lines “other possible changes.”



No language retruns to SOV. The nonreversibility obeys the entropy law. Pereltsvaig (2011) introduces Talmy Givon’s study that a human infant acquires his or her mother tongue as SOV at the initial state of C_{III}. (the ontology of C_{III}.) For phylogeny, the third factor (geometrically lowest cost) determines the unmarked word order <SOV>. But for ontogeny, capitalizing on Yang (2002: 72), the learner can reliably associate an irregular order (OSV, VOS, OVS) with its matching irregular rule, and reliably apply the rule over the default <SOV>. The existence of irregular unmarked word orders parallels that of irregular verbs. These studies support

the view that SOV is the base pattern. Pereltsvaig (2011) also contains a critical review of Gell-Mann and Ruhlen (2011).

- 11 There is much evidence which indicates that V merges with O. V selects O (e. g., the V *say* selects a *that* clause as O but the V *kill* does not), V forms idioms with O (e.g., *Kick the bucket*), a transitive verbal noun N^V produces a compound word with O (e.g., *manslaughter*), and sequential voicing occurs between V and O (e.g., compound words in Japanese). The head parameter is defined with respect to the complement (O) but not the specifier (S) (Uriagereka 2012: 13-14).
- 12 See Barrie (2006: 99-100) for this solution, which avoids the initial-merge problem (or the “bottom of the phrase-marker” linearization problem; Uriagereka 2012: 141, fn.23; citing Chomsky 1995, Chap. 4). In fact, the structure-building space is empty (\emptyset) before V enters. “...take only one thing, call it “zero,” and you merge it; you get the set containing zero. You do it again, and you get the set containing the set containing zero; that’s the seccessor function.” (Chomsky and McGilvray 2012: 15) “The empty set is not ‘nothing’; nor does it fail to exist. It is just as much in existence as any other set. It is its *members* that do not exist. It must not be confused with the number 0: for 0 is a number, whereas \emptyset is a set.” (Stewart 1975: 48) “...the empty set \emptyset is a subset of any set you care to name – by another piece of vacuous reasoning. If it were not a subset of a given set S, then there would have to be some element of \emptyset which was not an element of S. In particular there would have to be an element of \emptyset . Since \emptyset has no elements this is impossible.” (ibid. 49)
- 13 An intermediate projection such as V’ is used for expository purposes.
- 14 The base vP is consistent with the MSO hypothesis (which states that there is more than one point when a structure with sound features attached is sent to the PF (sound interface) (Uriagereka 2012: 113, fn.33). According to MSO, a domain, such as S, that is moved to TP Spec and spelled out independently becomes opaque to subextraction. O in the base vP remains in situ and is not spelled out independently, and hence, no island effect is detected for O. Uriagereka cites Jurka (2010), who maintains that Kayne’s (1994) hypothesis that <SVO> derives <SOV> is dubious because it incorrectly predicts that the moved O should exhibit the island effect. My base-vP (base-SOV) hypothesis rejects Kayne’s (1994) base-SVO hypothesis. See also Fukui and Takano (1998) for the base-SOV hypothesis. See Arikawa (2012 a) for calculation of complexity level of island.

- 15 C_{HL} is a virus-checking system (Piattelli-Palmarini and Uriagereka 2004).
- 16 The original definition of LCA is as follows (Kayne 1994: 6). Given that $d(X)$ = the set of terminals T that X dominates and A = the set of ordered pairs $\langle X_j, Y_j \rangle$ such that for each j , X_j asymmetrically c -commands Y_j , where X asymmetrically c -commands Y iff X c -commands Y and Y does not c -command X , LCA is defined as follows: LCA = def. $d(A)$ is a linear ordering of T .

The mirror image of LCA is Mirror LCA (MLCA) (Uriagereka 2012: 56). MLCA states that “when x asymmetrically c -commands y , x follows y .” I propose that LCA generally applies to phrases (VP, vP , TP, and CP), while MLCA generally applies to heads (C, T, v , and V; head movement is dispensed with). A typical exception is Malagasy (Austronesian family of languages), in which LCA applies to heads (e.g., $T \gg v \gg V$ maps to $\langle TvV \rangle$), and MLCA applies to phrases ($S \gg O \gg V$ maps to $\langle VOS \rangle$, MLCA applying to the base vP). See Section 4 and Appendix A for concrete applications of MLCA. C_{HL} employs both LCA and MLCA.

- 17 These are external merges. Note that the initial merge problem does not arise here because merging of x and ϕ always yields x . Focusing on the final label, internal merge and tucking in (Lebeaux 1988, Richards 1997) also satisfy the associative law. Take the left-hand structure in Figure 7. Remerging x with z followed by tucking y in under the remerged x results in z . Remerging y with z followed by tucking x in under the remerged y results in z . The equation $(x \bullet z) \bullet y = x \bullet (z \bullet y)$ holds. In an earlier draft, I assumed the tucking-in operation and that O moves to vP . Take the following structure.

$$(i) \quad [_{vP} v [_{VP} V O]]$$

Remerging O with vP followed by externally merging S ($S \bullet (vP \bullet O)$) results in the syntactic relation $S \gg O \gg V$. Remerging S with vP followed by tucking O in under S ($(S \bullet vP) \bullet O$) results in the relation $S \gg O \gg V$. The equation $S \bullet (vP \bullet O) = (S \bullet vP) \bullet O$ holds. However, I now reject my previous analysis because it wrongly predicts that the object NP would become an island, given the MSO hypothesis. I thank an anonymous reviewer for pointing out a potential problem with my previous analysis.

A puzzle remains, however. C_{HL} obeys the associative law with respect to the syntactic property of asymmetric projection or labeling, as in (i-b). C_{HL} seems to disobey the associative law with respect to the semantic property (meaning), as in (i-a).

- (i) a. $(\text{old} \bullet \text{English}) \bullet \text{teachers} \neq \text{old} \bullet (\text{English} \bullet \text{teachers})$
- b. $(A \bullet A) \bullet N = A \bullet (A \bullet N) = N$

This study focuses on the syntactic linguistic group.

This issue relates to a problem that was raised by an anonymous reviewer, namely, external merge violates the associative law because the resulting structures are simply distinct (what merges to what is different). However, provided that a merge operation is inherently asymmetrical in that double headedness is avoided and that what matters is the label of the entire structure, the resulting structures are identical. A puzzle remains about why the associative law is violated in the semantics. I thank the reviewer for helping to clarify the issue.

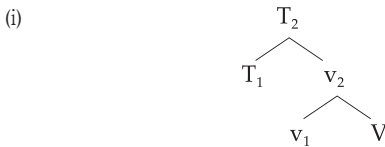
- 18 Besides syntactic relation, there are at least four other candidates that constitute G and G-like structures (Abelian G and Monoid). Abelian G is named after Niels Henrik Abel, a Norwegian mathematician (1802-1829). Abelian G (commutative G; stronger G) is a G-like structure with closure, identity, inverse, associativity, and commutativity ($a \bullet b = b \bullet a$). Monoid (weaker G) is a G-like structure with closure, identity, and associativity, but without inverse.

- (i) Syntactic operations: G
 - a. $\text{Merge}_1 \bullet \text{Merge}_2 = \text{Merge}_{1+2}$. (Merge is a closed operation.)
 - b. Copy is the identity (do-nothing) operation. It does not build a tree.
 - c. Reconstruction and Delete are the inverse operations.
 - d. Given three Merges, any two Merges can precede and the final output is a merged structure.
- (ii) Categories (CAT) under Merge (Boeckx 2009: 47) :
 - a. $\text{CAT}_1 \bullet \text{CAT}_2 = \text{CAT}_{1+2}$.
 - b. N is the identity element. ($N + X = X$, where X is a head (V, v, T, C).
 - c. Nominalizers (*-er/-tion/-ing*) are the inverse element of V: $V + \text{Nominalizer} = N$.
Complementizer Comp *that* is the inverse element of TP: $that + \text{TP} = N$.
 - d. $(\text{CAT}_1 \bullet \text{CAT}_2) \bullet \text{CAT}_3 = \text{CAT}_1 \bullet (\text{CAT}_2 \bullet \text{CAT}_3)$.
- (iii) Terms under Merge: Monoid
 - a. $\text{Term}_1 \bullet \text{Term}_2 = \text{Term}_{1+2}$. (Terms are closed under Merge.)
 - b. ϕ is the identity term.

- c. No inverse term. There is no term which merges with a term to form ϕ .
- d. Given three terms and no double-headedness, a merge of any two terms yields the same term.
- (iv) Formal features (FF) under Agree: Abelian G
 - a. FF_1 (probe) \bullet FF_2 (goal) = ϕ . (FFs are closed under Agree.)
 - b. ϕ is the identity FF.
 - c. FF_1 (probe) agrees with FF_2 (goal) and they become ϕ . FF_1 is the inverse element for FF_2 and vice versa.
 - d. Given three FFs (one probe FF_1 and two goals, FF_2 and FF_3) and multiple Agree, Agree ($FF_1 \bullet FF_2$) and Agree ($FF_1 \bullet FF_3$) yield the same result, namely, ϕ . (It satisfies Associative law.)
 - e. Given two FFs (a probe FF_1 and the goal FF_2), Agree ($FF_1 \bullet FF_2$) = Agree ($FF_2 \bullet FF_1$). That is, FF_1 agreeing with FF_2 and FF_2 agreeing with FF_1 yield the same result, namely, ϕ . (It satisfies Commutative law.)

I concentrate on syntactic relation as G under Merge because it is more consistent with the discussion on the symmetrical transformation of equilateral triangle. Chomsky (2009: 32-33) contains discussion (Chomsky, Piattelli-Palmarini, Higginbotham) about relationship among C_{III} , ϕ , number system, and semigroup. Semigroup satisfies closure and associativity, and it lacks identity, inverse element, and ϕ .

- 19 In the highest T position, V adjoins to v and [_v v [_v V]] adjoins to T, as in the following:



The active terms are T_1 , v_1 , V , and the two-segment categories [_T T₂, T₁] and [_v v₂, v₁]. T_2 and v_2 are inactive and have no role. Assuming the exclusion version of c-command (Kayne 1994) and that a two-segment category dominates the lower segment (Chomsky 1995: 339) (putting aside slight misgivings about self-dominance)

tion), V asymmetrically c -commands the two-segment category $[_v v_2, v_1]$, which asymmetrically c -commands the two segment-category $[_T T_2, T_1]$. LCA produces the universal $\langle VvT \rangle$ linear order. See Chomsky (1995: 338-340) for the relevant discussion, and Arikawa (2011) for the flexible c -command, which subsumes Chomsky's insight regarding the various levels of disconnection.

Alternatively, and in fact, I adopt this line of thought that, if we assume MLCA (Uriagereka 2012), we can eliminate all the technologies related to head movement adjunction, two-segment category, exclusion version of c -command, etc. All we need is c -command. That is, quite simply, without head movement, T asymmetrically c -commands v and v asymmetrically c -commands V and MLCA produces the universal $\langle VvT \rangle$ linear order. MLCA simplifies the model of the structure – order mapping. The geometrical cost approach is consistent with MLCA: the base vP is cheaper than TP , which has an additional external merge of T and subject movement, and TP is cheaper than CP , which has an additional external merge of C .

- 20 Regarding the rare unmarked orders $\langle OVS \rangle$ (0.7%) and $\langle OSV \rangle$ (0.5%), they might be outrageous enough to violate one of the most rigorous laws such as Relativised Minimality (RM) (Rizzi 1990). Without applying MLCA, $\langle OVS \rangle$ and $\langle OSV \rangle$ demonstrate O skipping S , which is usually prohibited by RM.
- 21 $\langle SOV \rangle$ with overt agreement corresponds to a 360° rotation (an expensive I). $\langle SOV/SVO \rangle$ mixed order rates 43.3% (29 out of 67). Concerning relative clauses (RC) and their heads (H), the order $\langle H, RC \rangle$ occurs in 70.3% of languages, $\langle RC, H \rangle$ in 17.1%, and mixed order in 7.8% (Dryer and Martin 2011).
- 22 See Dryer and Martin (2011) for extensive data.
- 23 Other language-family examples of SOV/SVO mix include Chibchan, Niger-Congo, Papuan, and Uralic (Dryer and Martin 2011).
- 24 If Russian is $\langle SVO \rangle$, it is a counterexample since the language has scrambling. If Russian is $\langle SOV \rangle$, as Pereltsvaig (2011) reports so in speech, it is consistent with the hypothesis.
- 25 Welsh word order is relatively fixed (Suzuki 2002).
- 26 Dem = demonstrative, Num = number, A = adjective, N = noun. See Cinque (2005: 319-320) for a detailed literature study and typological sources. This first DP word order is the most common and is found in the Afro-Asiatic, Altaic, Caucasian, Indo-European, and Uralic language families (see Cinque 2005: 319; fn.7 for sources).

- Cinque states that the statistics (very many, many, few, very few, and *) indicates “whether the order is attested or unattested” and it is based on various “typological (or other) sources available in the literature on the order of N, demonstrative, numeral, and adjective (that I have been able to find)” (ibid. 318). I assume that Cinque deals with unmarked word orders in nominal expressions.
- 27 This order is found in Cambodian, Javanese, Karen, Khmu, Palaung, Shan, Thai, Enga, Dagaare, Ewe, Gungbe, Labu and Ponapean, Mao Naga, Selepet, Yoruba, West Greenlandic, Amele, Igbo, Kusaeian, Manam, Fa d’Ambu, Nubi, Kugu Nghancara, Cabécar, Kunama, and Māori (see Cinque 2005: 320; fn.19 for sources).
- 28 Cinque notes that this order seems common in Europe, citing Rijkhoff (1998: 357). Example languages outside Europe include Yao, Burushaski, Guarani, Abkhaz, Farsi, Kiowa, Mam, Cape Verdean, Mauritian, Seychelles Creoles, Kristang, Kriyol, and Tok Pisin (Cinque 2005: 319; fn.8 for sources).
- 29 This order is found in Kabardia, Warao, Burmese, Lolo, Maru, Rawang, Ladakhi, Gambian Mandinka, Cuna, Kaki Ae, and Pech (see Cinque 2005: 320; fn.14 for sources).
- 30 Cinque (2005: 318) derives the four major unmarked nominal word orders by internal merge. The solution has problems. First, it is not clear what motivates the internal merge. Second, the advocated internal merge produces “few” and “very few” as well as “many” and “very many” orders at the same time. See Boeckx et al (2009: 218-220) and Chomsky (2009: 402) for relevant discussion.
- 31 In an earlier draft, I used the polygon square, which has only eight isometries (four rotations and four flips) – far fewer symmetrical points than are needed for 24 linear orders. I ended up dreaming about acrobatic transformations that are far outside the Galois group theory. I thank an anonymous reviewer who showed me a detailed (14 × 14) multiplication table for symmetry and other transformations to persuade me that what I was doing was pure nonsense and that I was certainly on the wrong track using the regular square for the problem in the first place.
- 32 This order is found in languages such as Sampur, Camus, and Masai (see Cinque 2005: 318; fn.9 for sources). The fact that this order is not assigned * (unattested; none) indicates that C_{III} , that is set for these languages has a way to avoid cost break down. The low probability parallels with those as 0.7% for <OVS> and 0.5% for <OSV>.
- 33 This order is found in languages such as Newari, Dulong, Tamang, Sinhala, and Shipibo-Konibo (see Cinque 2005: 320; fn.13 for sources).

- 34 This order is found in Koiari, Bai, and Zande (see Cinque 2005: 319; fn.11 for sources).
- 35 The unmarked word order <VOS> is found in Austronesian languages such as Malagasy, Batak, and Seediq, Native American, and Chibchan languages. See Dryer and Martin (2011) for more samples.
- 36 This order is found in Pitjantjatjara (see Cinque 2005: 323; fn.27 for the source).
- 37 This order is found in languages such as Mon-Khmer languages, Basque, Celtic, Easter Island, Hebrew, Hmong, Indonesian, Jacaltec, Rapanui, Wolof, creoles, and Watjarri (Australian language) (see Cinque 2005: 320; fn.16 for sources).
- 38 This order is found in languages such as Sango, Gude (?), and Zande (?) (see Cinque 2005: 320; fn.18 for sources).
- 39 This order is found in languages such as Lalo, Lisu, Akha, Qiang, Aghem, Port Sandwich, Koiari, Lingala, Babungo, and Woleaian (?) (see Cinque 2005: 319; fn.12 for sources).
- 40 This order is found in Berbice Dutch Creole, Sranan (creole), Bislama (creole), Xârâcùù, Laai, Puluwatese, Polish, and Russian (see Cinque 2005: 320; fn.15 for sources).
- 41 This order is found in Kikuyu, Turkana, Rendille, Noni, Nkore-Kiga, Abu', Arbore, Bai, Moro, and Romanian (see Cinque 2005: 319; fn.10 for sources).
- 42 This order is found in Gabra, Logoli, Luo, Lango, Kele, Buma, and Manam (see Cinque 2005: 320; fn.17 for sources).
- 43 A necessary condition (q) and a sufficient condition (p) stand in the relation: if p , then q ($p \rightarrow q$). In a sense, q looks for a premise given some conclusion, and p proposes a stronger restriction in order to obtain the desirable result. The premise q here is Galois-theoretic cost-effectiveness and the stronger restriction p concerns linguistic facts. In other words, if a transformation T is linguistically cost-effective, then T is also Galois-theoretically cost-effective, but the reverse does not hold.
- 44 Let Dem , Num , A , and N be α , β , γ , and δ , respectively. Then C_{III} must be solving a quartic E_{III} (human language equation) that has a Galois group G^g as follows. See Appendix for relevant discussions.

(i) a. $\alpha, \beta, \gamma, \delta$ ↓ $\alpha, \beta, \gamma, \delta$ I	b. $\alpha, \beta, \gamma, \delta$ ↓ $\delta, \gamma, \beta, \alpha$ M3	c. $\alpha, \beta, \gamma, \delta$ ↓ $\alpha, \beta, \delta, \gamma$ R1	d. $\alpha, \beta, \gamma, \delta$ ↓ $\alpha, \delta, \gamma, \beta$ R2
---	--	--	--

What is the E_{III} with the G^s that has these permutations? Clearly, the E_{III} cannot be similar to the following biquadratic equation that I borrow from Stewart (2004):

(ii) $f(x) = x^4 - 4x^2 - 5 = 0$,

which is factored as

(iii) $(x^2 + 1)(x^2 - 5) = 0$,

whose roots are

(iv) $\alpha = i, \beta = -i, \gamma = \sqrt{5}, \delta = -\sqrt{5}$.

The G^s of this biquadratic equation consists of the following, given the facts of C_{III} :

(v) a. $\alpha, \beta, \gamma, \delta$ ↓ $\alpha, \beta, \gamma, \delta$ I	b. $\alpha, \beta, \gamma, \delta$ ↓ $\beta, \alpha, \gamma, \delta$ R1×M1	c. $\alpha, \beta, \gamma, \delta$ ↓ $\alpha, \beta, \delta, \gamma$ R1	d. $\alpha, \beta, \gamma, \delta$ ↓ $\beta, \alpha, \delta, \gamma$ M1
① (Very many)	⑩ (None)	⑦ (Many)	④ (None)

Although the G^s of the biquadratic equation in (ii) partially captures half of the fact: ①<Dem, Num, A, N> = very many, ⑦<Dem, Num, N, A> = many, it incorrectly includes two linear orders (④, ⑩), which are not produced by C_{III} . Therefore, C_{III} must be solving a quartic equation E_{III} that is more complex than a simple biquadratic equation.

45 Wallace’s (1889: 467) statement is cited in Chomsky (2005: 16; 2007: 7; 2010: 53). “The significance of such phenomena, however, is far from clear.” (Chomsky 2009: 26, 33). See Chomsky & McGilvray (2012: 16) for relevant discussion, in which Chomsky argues against Butterworth (2000), which argues against Chomsky’s hypothesis that mathematical capacity is derived from language.

46 Chomsky has stricter view than Kronecker in that C_{III} is the origin of natural numbers, not integers. According to Chomsky (2005: 17), the “most restrictive case of Merge applies to a single object, forming a singleton set. Restriction to this case yields the successor function, from which the rest of the theory of natural numbers

can be developed in familiar ways.”

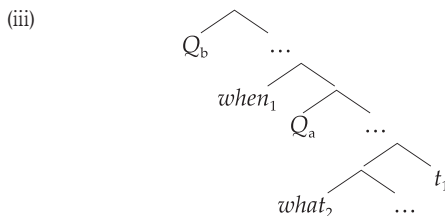
- 47 This is a huge “if.” An anonymous reviewer (a group theorist) asks: Could solving *algebraic* equations be such a fundamental logical operation as to explain whatever symmetry that is found in human brain? The reviewer is inclined to answer *no*.
- 48 As a mathematical amateur, I rely on Nakamura (2010: 91-104), a book that was written for the general public, for an introduction to the Galois theory. To view a sentence as an equation is not so outlandish in the generative syntax study. Ross (1967) initiated the study of constraints on variables in syntax, which has revealed that a wh-trace (variable x , or wh-phrase under the copy theory of movement) appears in structurally restricted positions. For example, consider the following sentence, which is less acceptable. The symbol t stands for a trace, which is a bound variable that is bound by an operator. [...] is the embedded clause.

- (i) ??What₂ did you wonder [when Mary fixed t₂]?

The intended logical meaning is the following.

- (ii) What is x , x a thing, and when is y , y a time, such that you wondered Mary fixed x at y ?

Let us show the relevant structure before the movement of *what*. A structure is built bottom-up. Q is the sentential head that attracts a wh-phrase.



The embedded-clause Q_a has attracted *when*₁. When Q_b appears, the minimality condition (*Attract the closest*. A physical law. Rizzi 1990) requires that Q_b attract the closest wh-phrase, i.e. *when*₁. However, Q_b attracted *what*₂, which is a violation of the minimality condition. Hence, the less acceptability of (i). The sentence has the following structure.

- (iv) ($what_2 + Q_b (when_1 + Q_a (x_1 (x_2)))$) = 0

C_{III} fails to solve the equation: the equation has no solution for x_2 . Ross named the underlined structure an *island*: a structure in which a variable cannot connect with the operator. In this case, the variable x_2 cannot connect with the operator

what. The generative syntacticians have studied structures of island for more than six decades. See Arikawa (2012 a) for a method of calculating island complexity.

49 I thank an anonymous reviewer (a group theorist) for pointing out that it may be meaningless to talk about the rationality of the roots s , o , and v . Mathematicians have extended number system to solve new types of equations. Stewart (2004: 5-6) gives a good illustration. An equation as $x + 2 = 7$ can be solved in \mathbb{N} (natural numbers), $x + 7 = 2$ in \mathbb{Z} (integers), $2x = 7$ in \mathbb{Q} (rationals), $x^2 = 2$ in \mathbb{R} (real numbers), and $x^2 = -1$ in \mathbb{C} (complex numbers). The relation $\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}$ holds. A complex number in \mathbb{C} is expressed as $a + bi$, where a and b are real numbers ($b \neq 0$) and i is an imaginary number. I intentionally misread a question posed by Stewart (2004: 21). “The question then arises: why stop at \mathbb{C} ? Why not find an equation that has no solutions over \mathbb{C} , and enlarge the number system still further to provide a solution?” It may be that the roots s , o , and v are such enlarged numbers. However, this section does not dare to go that far. I try to show that s , o , and v stay in \mathbb{C} . But Stewart seems open-minded when he says that “The answer is that no such equation exists, at least if we limit ourselves to polynomials.” The language equation could be a new type of “polynomials” that needs a new type of “numbers.” Again, I may be talking nonsense, as an anonymous reviewer (a group theorist) keeps warning me.

50 The symbol \Leftrightarrow indicates equivalence; the left and right properties simultaneously must or must not hold.

51 Here we are witnessing spontaneous symmetry forming (obeying the entropy law) battling with spontaneous symmetry breaking (obeying the algebraic and economy principles). Note that the entropy law, algebraic principle, and economy principle belong to Chomsky’s third factor. Several thousand years ago, the ancient C_{III} was less symmetrical. In full compliance with algebraic and economy principles, the ancient C_{III} selected <SOV> as the sole unmarked word order. The current C_{III} is more symmetrical; all six permutations have emerged, despite asymmetrical distribution with a gradient according to the cost difference. If algebraic and economy principles win the battle, the future C_{III} will again select <SOV> as the sole unmarked order, as in the ancient languages. If the entropy law wins the battle, the future C_{III} will become fully symmetrical; all six permutations will appear, each with the ratio of 16.6% of the total languages.

52 I thank an anonymous reviewer (a group theorist) for pointing out that it may

be meaningless to talk about rationality of s , o , and v . If a field expands, G^s will shrink. If a field shrinks, G^s will expand (Lieber 1932, Nakamura 2010). The fact that the G^s of C_{III} has expanded indicates that the field of C_{III} has shrunk. Then, the property of s , o , and v in C_{III} have undergone a change in the following direction: \mathbb{C} (complex numbers) \rightarrow \mathbb{R} (real numbers) \rightarrow \mathbb{Q} (rational numbers) \rightarrow \mathbb{Z} (integers) \rightarrow \mathbb{N} (natural numbers). Given $\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}$, it follows that C_{III} contained everything from the beginning. If so, the direction of this evolution is opposite to the process by which mathematicians have expanded numbers.

- 53 I thank an anonymous reviewer (a group theoretician) for warning me that “of course one is free to consider a hypothetical analogy between symmetry among arbitrary objects and that among the roots of equations – which are elements of fields, i.e., objects on which one can perform $+$, $-$, \times , and \div , e.g., *numbers*. But then we need to ask whether this analogy could be meaningless, or even if it seems meaningless, whether it shows any indication of a hidden structure of the objects in question.” This fundamental question governs the fate of my entire project.
- 54 For example, a cubic equation such as $x^3 - 2 = 0$ has a structure G^s that includes all six permutations (Lieber 1932, Nakamura 2011, Kim 2011). The current C_{III} must be solving an E_{III} that is similar to this cubic equation, which is the Delian problem (doubling the cube), which is unsolvable by compass and straightedge construction. The Egyptians, Greeks, and Indians knew about this problem thousands of years ago.
- 55 Different polynomials as $(s - o)(s - v)(o - v)$ or $-(s - o)(s - v)(o - v)$ break the symmetry.
- 56 The number of possible permutations of Rubik’s Cube is 43,252,003,274,489,856,000. If one does one permutation in one second, it would take more than one trillion years. If one started the permutation right after the Big Ban about 13.7 billion years ago, he or she is still working on it now (Kim 2011: 133).
- 57 An expert does it in less than 10 seconds.
- 58 Therefore, Jenkins (2000: 164) is right when he says “...word order types would be the (asymmetric) stable solutions of the symmetric still-to-be-discovered ‘equations’ governing word order distribution.”
- 59 In a linear matrix equation $Ax = b$, when b is given, how do we get x ? This is an inverse problem. We multiply the inverse element of A , which is A^{-1} , to both sides of the equation. We get $A^{-1} \cdot Ax = A^{-1} \cdot b$. Since $A^{-1} \cdot A = I$, $x = A^{-1} \cdot b$ (Strang

2003). Therefore, a typical solution of an equation is a direct (well-posed/conditioned) problem, where input (an equation or a model) is given and output (roots or data) is unknown. Inverse (ill-posed/conditioned) problems, where output (data) is given and input (model) is unknown, are well-studied mathematical problems.

60 We may not be able to solve it (it may be just too hard for Homo sapiens). Or, the problem is posed incorrectly, as an anonymous reviewer (a group theorist) warns me.

61 Consider a simpler case. The formula for solving a quadratic equation $x^2 + px + q = 0$ is

$$(i) \quad x = \frac{-p \pm \sqrt{p^2 - 4q}}{2}$$

Given the two roots, a and b , the formula expresses the mean between the elementary symmetrical polynomial $(a + b)$ and the elementary difference product $\pm (a - b)$ (Nakamura 2010: 36). The reasoning is as follows. Since the two roots are a and b , we have $(x - a)(x - b) = 0$. Factorizing, we get $x^2 - (a + b)x + ab = 0$. Thus, $a + b = -p$, and $ab = q$. The formula in (i) is expressed with a and b as follows:

$$(ii) \quad x = \frac{-p \pm \sqrt{p^2 - 4q}}{2} = \frac{(a + b) \pm \sqrt{(a + b)^2 - 4ab}}{2} \\ = \frac{(a + b) \pm \sqrt{(a - b)^2}}{2} = \frac{(a + b) \pm (a - b)}{2}$$

Assume $a > b$. The true character of $\sqrt{p^2 - 4q}$ is the elementary difference product $a - b$. Similarly, when we have a quadratic E_{hil} $x^2 + px + q = 0$ with the two roots s and v , the identity of the formula for the E_{hil} is

$$(iii) \quad x = \frac{(s + v) \pm (s - v)}{2}$$

Assume $s > v$. If the field consists of rational numbers (s and v are rational numbers), the G^s of the quadratic E_{hil} includes $\langle sv \rangle$ and $\langle vs \rangle$, and $\langle sv \rangle$ and $\langle vs \rangle$ are the two unmarked word orders that the current C_{hil} demonstrates. Crucially, the E_{hil} version of the numerator of the second equation in (ii) $(s + v) \pm \sqrt{(s + v)^2 - 4sv}$, is a symmetrical polynomial that is not affected by switching s and v . The symmetry of the two unmarked word orders $\langle sv \rangle$ and $\langle vs \rangle$ expresses the symmetry of the formula for the E_{hil} .

62 I speculate that the initial state of the current C_{hil} in a human infant solves an

E_{HL} similar to $x^3 - 2 = 0$, which is irreducible (nonfactorizable) if the field consists of rational numbers. In the initial state, the current C_{HL} is free to choose any permutation as the unmarked word order (a human infant can acquire any permutation as the unmarked word order); this is a linguistic fact. In the course of parameter setting (mother-tongue acquisition), the current C_{HL} narrows the field to solve a particular equation that has a particular permutation as the $G^{\#}$. For example, the current C_{HL} that is exposed to an <SOV> environment develops in a specific way to solve an E_{HL} similar to the cubic equation (9), by which a higher degree of symmetry (the scrambling phenomenon) emerges. It follows that there are six distinct final states of the current C_{HL} , which solves six distinct types of E_{HL} with six distinct types of $G^{\#}$, each of which consists of a particular single member (unmarked word order). However, this ignores the probability variance: <SOV> and <SVO> emerge in almost the same ratio of languages and together make up more than 80% of all languages that have been studied. Concerning the phylogeny of the C_{HL} of homosapiens as a species, the initial state of the ancient C_{HL} was solving an E_{HL} such as (9) with a less symmetrical $G^{\#}$, whereas that of the current C_{HL} solves an E_{HL} similar to $x^3 - 2 = 0$ with a more symmetrical $G^{\#}$. The entropy law explains the symmetry formation.

- 63 We focus on idealized intransitive and transitive verbs other than ellipses.
- 64 I thank an anonymous reviewer (a group theorist) for teaching me the essence of groups as follows. The Galois groups of equations do not necessarily dictate which type of numbers we work on. An equation can be defined over any number field, indeed over any field (not necessarily consisting of numbers), and in most cases, it is easy to cook up examples of equations that have a given subgroup as their Galois groups. Therefore, the types of subgroups do not have their own inherent Galois-theoretical (much less number-theoretical) meanings – they are just groups and nothing more. It follows that the section titles in A.4.1 and A.5.1 have expository purposes only and do not reflect any fundamental classification.
- 65 The reasoning is as follows. By factorization, we obtain

$$(i) \quad x^2 + 3x - 4 = (x + 4)(x - 1) = 0.$$

The roots are -4 and 1 : $x = -4, 1$. Given that the relevant E_{HL} with the roots $\{s, v\}$ is

$$(ii) \quad (x - s)(x - v) = x^2 - (s + v)x + sv = 0,$$

the two real number roots are $x = -4 (= s), 1 (= v)$. More generally, we have

(iii) $x^2 + px + q = 0$.

Given that $p = -(s + v)$, $q = sv$, the equation yields

(iv) $p = -(s + v) = -(-4 + 1) = 3$
 $q = -4 \times 1 = -4$.

This is what we expect from the relationship between roots and coefficients. Let us consider the difference product, Δ , a possible rational expression R. Consider the discriminant D:

(v) $D = \Delta^2 = (s - v)^2 = (s + v)^2 - 4sv = (-p)^2 - 4q = p^2 - 4q$
 $= (3)^2 - 4 \cdot (-4) = 25$.

Therefore, we obtain $\Delta = \pm 5$. Because the values of R ($s - v$) are rational numbers, G^s cannot contain a permutation that changes the value of R. Therefore, G^s must consist of I alone.

66 $K = (sv)$ indicates that K is a permutation that switches s and v .

67 The examples are from Kamei et al. (1989: 584). Among 1498 languages, <SV> accounts for 79.7%, <VS> accounts for 13.0%, and the remaining 7.3% are languages with no dominant order. The Galois group does not distinguish <sv> and <vs>, but in actual calculation, it is economical for C_{HL} to find a simpler root. The root s may be computationally simpler than the root v . There is no corresponding difference between O and V; <OV> accounts for 46.9% of the languages and <VO> accounts for 46.4%, while the remaining 6.6% have no dominant order (Dryer and Martin 2011).

68 The reasoning is as follows. A more general equation is

(i) $x^2 + px + q = 0$.

If E_{HL} had two roots s and v , we would have $(x - s)(x - v) = 0$. By expansion, we obtain

(ii) $x^2 - (s + v)x + sv = 0$.

The relations between the roots and coefficients are $p = -(s + v)$ and $q = sv$. Now, the discriminant D is the square of the difference product $\Delta = (s - v)$. Let us determine D:

(iii) $D = \Delta^2 = (s - v)^2 = (s + v)^2 - 4sv = (-p)^2 - 4q = p^2 - 4q$.

Therefore, $D = \Delta^2 = 3^2 - 4 \cdot 1 = 5$.

Since $D = 5 > 0$, the equation has two real number roots.

69 If the field of R were the set of real numbers (the field contains $\sqrt{5}$), the value of R would have to be preserved under all permutations (Lieber 1932). G^s_{HL} cannot

contain K because K changes the value of R , i.e., the value changes from $(s - v)$ to $(v - s)$. Thus, G_{III}^s contains I alone. The empirical facts about C_{III} are that $\langle SV \rangle$ languages (80%) outnumber $\langle VS \rangle$ languages (20%), while $\langle OV \rangle$ (50%) and $\langle VO \rangle$ (50%) appear in the same percentage of languages (Dryer and Martin 2011). The present C_{III} may be using the field similar to the real number field (an expanded field) for $\{S, V\}$ but the field similar to the rational number field (reduced field) for $\{O, V\}$.

- 70 The reasoning is as follows. Given one rational number and two irrational numbers, by (10b), the G^s must contain a permutation that changes the value of R and another that does not. The permutation that does not change R is the identity permutation I . Thus, G^s contains I . The quadratic part of the cubic equation

$$(i) \quad x^2 + 3x + 1 = 0$$

has two irrational number roots, and, as we have already noted, the G^s consists of I and K , which switches the two roots. Thus, the G^s of the cubic equation in question includes rO and fI .

- 71 The calculation is as follows. Because

$$(i) \quad (s - o)(s - v)(o - v) \\ = -\frac{(-3 + \sqrt{5})}{2} \cdot -\frac{(-3 - \sqrt{5})}{2} \cdot \sqrt{5} = \frac{9 - 5}{4} \cdot \sqrt{5} = \sqrt{5},$$

it follows that $\Delta = R = \pm\sqrt{5}$.

- 72 I thank an anonymous reviewer (a group theorist) for pointing out a mathematical fact that is possibly relevant, namely as the field \mathbb{R} of real numbers only has quadratic extensions (the field \mathbb{C} of complex numbers), there is no cubic equation over the real numbers with the Galois group C_3 or S_3 . That leaves $\{id\}$ and C_2 . This paper considers $\{id\}$ to be $\langle SOV \rangle$ and C_2 to be $\langle SOV \rangle$ and $\langle SVO \rangle$. It is the fundamental nature of cubic equations over the rational field that the equations have $\langle SOV \rangle$ alone as G^s or $\langle SOV \rangle$ and $\langle SVO \rangle$ as G^s . This paper draws a strong connection between this mathematical fact and the linguistic fact that the ancient languages showed $\langle SOV \rangle$ universally and more than 80% of the current languages have either $\langle SOV \rangle$ or $\langle SVO \rangle$ as their unmarked word orders. However, the reviewer warns me that I have gone too far at this point. That is, I cannot deduce from this the fact that an object with a symmetry corresponding to C_2 must have something to do with real numbers rather than other number fields. In addition, the reviewer claims that C_2 is too simple to show anything. However, an object

(unmarked word orders in C_{III}) with a symmetry corresponding to C_2 (however simple) must at least have something to do with that subgroup rather than other subgroups.

Let us keep to this analysis for the moment (although it may be groundless). The analysis leaves $\langle VSO \rangle$ (9.2%) and other unmarked word orders as problems remaining to be explained. What does it mean to say that C_{III} is close to solving an equation with one rational number root and two irrational number roots? I propose that the factorization expresses the algebraic structure of the base vP. According to Larson (1988), it is standard to postulate a double-layer verbal structure:

$$(i) \quad E_{III} = [{}_{vP} x [{}_{vP} (x^2 + px + q)]] = 0$$

The vP edge (the rational root) constitutes s and the quadratic equation part of VP (the two irrational roots) constitutes v and o . C_{III} seems to have evolved such that the system first finds the simpler root s . The Galois theory analysis appears to grasp something about the essence of sentence structure. However, as the reviewer points out, this may be an illusion arising from the group theory, which seems to apply to anything without scientific significance.

73 The calculation is as follows. By the formula $\Delta = \pm \sqrt{-4p^3 - 27q^2}$, we obtain

$$(i) \quad \Delta = \pm \sqrt{-4 \cdot (-3)^3 \cdot -27 \cdot 1^2} = \pm \sqrt{-4 \cdot (-27) \cdot -27 \cdot 1} = \pm \sqrt{(4-1) \cdot 27} = \pm 9$$

74 The reasoning is as follows. Given a difference product: $\Delta = (s-o)(s-v)(o-v)$, $f1$, $f2$ and $f3$ affects Δ as follows.

$$(i) \quad f1: (s-o)(s-v)(o-v) = R \rightarrow (s-v)(s-o)(\underline{v-o}) = (s-o)(s-v)-(o-v) = -(s-o)(s-v)(o-v) = -R.$$

$$(ii) \quad f2: (s-o)(s-v)(o-v) = R \rightarrow (\underline{o-s})(o-v)(s-v) = -(s-o)(s-v)(o-v) = -R.$$

$$(iii) \quad f3: (s-o)(s-v)(o-v) = R \rightarrow (\underline{v-o})(\underline{v-s})(\underline{o-s}) = -(s-o)-(s-v)-(o-v) = -(s-o)(s-v)(o-v) = -R.$$

$f1$, $f2$, and $f3$ change the signs of Δ . We, therefore, exclude $f1$, $f2$, and $f3$ from G_{III}^* .

75 The reasoning is as follows. Let us repeat the difference product:

$$(i) \quad \Delta = \pm (s-o)(s-v)(o-v)$$

$$(ii) \quad I: (s-o)(s-v)(o-v) = R \rightarrow (s-o)(s-v)(o-v) = R.$$

$$(iii) \quad rI: (s-o)(s-v)(o-v) = R \rightarrow (\underline{v-s})(\underline{v-o})(s-o) = (s-o)-(s-v)-(o-v)$$

$$\begin{aligned}
 &= (s - o) (s - v) (o - v) = R. \\
 \text{(iv) } r2: & (s - o) (s - v) (o - v) = R \rightarrow (o - v) \underline{(o - s)} \underline{(v - s)} = -(s - o) - (s - v) (o - v) \\
 &= (s - o) (s - v) (o - v) = R.
 \end{aligned}$$

76 The reasoning is as follows. Let the cubic roots of 1 be:

$$1, \omega = \frac{-1 \pm \sqrt{3}i}{2}$$

The equation has three roots, $s = \sqrt[3]{2}$, $o = \sqrt[3]{2} \omega$, and $v = \sqrt[3]{2} \omega^2$. The question is: What is the G^g ? Given the three roots s , o , and v , the value of the difference product $\Delta = (s - o) (s - v) (o - v)$ is given by the formula $\Delta = \pm \sqrt{-4 \times 0^3 - 27 \times (-2)^2} = \pm \sqrt{-108} = \pm \sqrt{108} i^2 = \pm \sqrt{108} \times i$. Given the field of rational numbers, the value of Δ (a complex number with real and imaginary parts) does not exist in the field. Therefore, by definition, the permutations of G^g must change the value of Δ . Regarding three roots, even permutations (cyclic rotations) do not alter the value, whereas odd permutations (flips) do. Odd permutations are (so) , (sv) , and (ov) . By definition (10b), it is predicted that the G^g in question contains some of these flips. And the G^g may contain some other permutations. The definition (10a) requires that the G^g must be closed under multiplications. There are six G^g s that are closed. They are:

- A = S_3 (all the six permutations: $\{I, (abc), (acb), (so), (sv), (ov)\}$)
- B = C_3 ($\{I, (sov), (svo)\}$)
- C = $\{I, (so)\}$
- D = $\{I, (sv)\}$
- E = $\{I, (ov)\}$
- F = $\{I\}$

Now, the relevant G^g must contain one, two, or all of (so) , (sv) , and (ov) . Therefore, the G^g cannot be B or F. Suppose that the G^g is C. Consider $s + o$. I and (so) do not change the value. By (10b), the value of $s + o$ must be a rational number. Since the second-degree coefficient is 0, $s + o + v = 0$. Hence, $s + o = -v$. It follows that $-v$ must be a rational number. However, $-v = -\sqrt[3]{2} \omega^2$ is not a rational number. A contradiction arises. Therefore, the G^g cannot be C. By the same reasoning, the G^g cannot be D or E. That leaves A. Thus, the G^g is S^3 that contains the maximum members. Recall that when the field extends, the G^g shrinks. The ancient C_{HL} had the extended field and the minimized $G^g \langle sov \rangle$, whereas the current C_{HL} has the shrunken field and the maximized G^g . But what does it mean?

See Nakamura (2011: 122-127) and Kim (2011: 56-58) for plain and amateur-friendly instructions.

More Evidence for Geometrical Cost Approach To Basic Word Order Asymmetry in Human Language

ARIKAWA Koji

Abstract

The computational procedure for human natural language (C_{HL}) produces an asymmetry in unmarked word orders for S , O and V . Capitalizing on the insightful idea of Lyle Jenkins (2000, 2003), I propose that the asymmetry is based on a group-theoretical factor, which is included in Noam Chomsky's third factor: "principles of neural organization that may be even more deeply grounded in physical law" (Chomsky 1965: 59) and "principles of structural architecture and developmental constraints that enter into canalization, organic form, and action over a wide range, including principles of efficient computation, which would be expected to be of particular significance for computational systems such as language" (Chomsky 2005: 6). A mathematician would say that the symmetric group S_3 of order $3! = 6$ is too simple and the Galois theory is irrelevant. However, this very simplicity is the reason that I consider cost differences among the six symmetric operations of S_3 . A mathematician would say that this is not the Galois theory because Galois groups disregard cost. Still, as a biolinguist and a Galois-theory enthusiast, I would like to propose that (a) syntactic relations constitute a group under Merge; (b) the asymmetry reflects the asymmetry of the algebraic structure of sentences taken as expressions of equations that C_{HL} solves; and (c) s , o , and v are the roots of such equations.