Is Word Order Asymmetry Mathematically Expressible?

Koji Arikawa

The computational procedure for human natural language (C_{HL}) shows an asymmetry in unmarked orders for *S*, *O*, and *V*. Following Lyle Jenkins, it is speculated that the asymmetry is expressible as a group-theoretical factor (included in Chomsky's third factor): "[W]ord order types would be the (asymmetric) stable solutions of the symmetric still-to-be-discovered 'equations' governing word order distribution". A possible "symmetric equation" is a linear transformation $f(x) = y$, where function f is a set of merge operations (transformations) expressed as a set of symmetric transformations of an equilateral triangle, x is the universal base *v*P input expressed as the identity triangle, and y is a mapped output tree expressed as an output triangle that preserves symmetry. Although the symmetric group S_3 of order $3! = 6$ is too simple, this very simplicity is the reason that in the present work cost differences are considered among the six symmetric operations of S_3. This article attempts to pose a set of feasible questions for future research.

Keywords: cost; economy; equilibrium; Galois group; geometry; symmetry; third factor; transformation; unmarked word order

1. Introduction

1.1. Problem

I would like to pose the question of whether the following phenomenon can be mathematically (Galois theoretically) expressed.[1]

I am grateful to the editors and anonymous reviewers for their patience in assessing this challenging article over the past two years. I would like to thank Makoto Toma for his valuable comments and suggestions. Without his constructive criticism regarding my amateurish mathematics, I could not have finished this. I thank Massimo Piattelli-Palmarini for allowing me to join his class on biolinguistics at MIT in 2003, which marked the beginning of this project. I am grateful to Lyle Jenkins 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 a campus café. Finally, I would like to thank Enago for editing and proofreading my work, which clarified the reasoning that I wished to express. All remaining errors are my own.

[1] The author does not claim that the geometrical cost calculation proposed here is the 'third factor' (non-genetic and non-environmental) that is actually at work in C_{HL}. Rather, he claims that it may be a mathematically feasible way to express and translate the unmarked word order asymmetry into a language of geometrical cost calculation that leads us to

© The Author (2013)
ISSN 1450–3417
Biolinguistics 7: 276–300, 2013
http://www.biolinguistics.eu

(1) In terms of phylogeny,[2] C_{HL} shows the following language distribution: <*SOV*> = 48.5%, <*SVO*> = 38.7%, <*VSO*> = 9.2%, <*VOS*> = 2.4%, <*OVS*> = 0.7%, <*OSV*> = 0.5% (Yamamoto 2002).[3]

Why do we focus on *S*, *O*, and *V*?[4] There are four reasons. First, many reliable studies since the seminal work of Greenberg (1963) present relatively solid evidence regarding the probability of unmarked word orders. Second, we have reliable data from native speakers, who have relatively clear intuitions about what the unmarked order of the set {*S*, *O*, *V*} is for their languages. The third reason is simplicity: we should start from the simplest possible case. The fourth reason is reducibility: we can and should reduce seemingly complex structures to the simplest possible structures, namely *S* + *V* and *S* + *O* + *V*. *S* and *O* may be complex, but they are reducible to the simple *S* and *O*. *O* may be direct (*DO*) or indirect (*IO*), but we start from the simpler *DO*. Sentence structures contain CP, TP, *v*P, and VP, but the most basic semantic domain is *v*P+VP, in which *S*, *O*, and *V* appear originally. Yamamoto (2002: 85) contains a table that is useful for comparing the relevant percentages that have appeared in previous studies. Here I have included Yamamoto (2002), Dryer & Martin (2011), and Gell-Mann & Ruhlen (2011).[5] This full list is shown in Table 1.

---

algebraic and group-theoretical analyses in the future. I thank an anonymous reviewer for clarifying the issue. With regard to Galois theory, Évariste Galois (a French mathematician; 1811–1832) developed the fundamental mathematical tool, the Galois group (algebraic structure of equations), for examining the symmetry of equations. Modern science would not exist without Galois theory. Group theory is a calculus of symmetry (Stewart 2007: 111). In Chomsky (2002), Fukui and Zushi mention Weil (1969), which is a group-theoretical analysis of an aboriginal kinship structure in Australia (Japanese translation of Chomsky 1982: 356). As regards other group-theoretical analysis on C_{HL}, see Laughren (1982), in which the author attempts a group-theoretical analysis of Walpiri kinship structure (languages of kinship in Aboriginal Australia). See also Jenkins (2013) for the introduction of Laughren (1982). In Chomsky (2002), Fukui and Zushi suggest a possibility of "Galois theory of phrase structure (I-language)" (Japanese translation of Chomsky 1982: 397–398).

2 The phylogeny problem (species puzzle) asks why a language system (the current C_{HL}) behaves in a particular way, "the historical development of languages" (Di Sciullo 2013). However, we are concerned with synchronic phenomena (why the current C_{HL} appears like this; how it has come to have the property; what the cause is) and we put aside the actual diachronic analysis. The ontogeny problem (individual puzzle) asks how a human child acquires his or her mother tongue, i.e. "the growth of language in the individual" (ibid.).

3 Yamamoto (2002) considers the largest number (2,932) of languages for typological analysis to date (gross=6,000). The actual number used for calculating the percentages is 2,537. Given that many previous studies have only considered 20 or 30 to 200 or 300 languages, Yamamoto (2002) offers a significantly reliable sampling. <…> indicates an ordered set of unmarked (basic) word order. The ratio is rounded to the first decimal place.

4 I thank an anonymous reviewer for pointing out this fundamental question.

5 I added Yamamoto (2002; gross: 2,537 languages), Dryer & Martin (2011; gross: 1,377), and Gell-Mann & Ruhlen (2011; gross: 2,011). In Dryer & Martin (2011), 189 languages have no dominant order. Selected language families and samples are provided below.

<*SOV*>: Niger-Congo, Semitic, Turkic, Indo-Aryan, Dravidian, Austonesian, Altaic, Chibchan, Native American languages, …

<*SVO*>: Indo-European, Niger-Congo, Tai-Kadai, Sinae, Austronesian, Arawakan, …

<*VSO*>: Celtic, Semitic, Niger-Congo, Austronesian, Native American languages, Chibchan, …

<*VOS*>: Malagasy, Batak, Seediq (Austronesian languages), Native American languages, Chibchan, …

	SOV	*SVO*	*VSO*	*VOS*	*OVS*	*OSV*
Greenberg 1966	37.0%	43.0%	20.0%	0.0%	0.0%	0.0%
Ultan 1969	44.0%	34.6%	18.6%	2.6%	0.0%	0.0%
Ruhlen 1975	51.5%	35.6%	10.5%	2.1%	0.0%	0.2%
Mallinson & Blake 1981[6]	41.0%	35.0%	9.0%	2.0%	1.0%	1.0%
Tomlin 1986	44.8%	41.8%	9.2%	3.0%	1.2%	0.6%
Matsumoto 1992	49.3%	35.0%	11.2%	2.8%	1.0%	0.6%
Yamamoto 2002	48.5%	38.7%	9.2%	2.4%	0.7%	0.5%
Dryer & Martin 2011	41.0%	35.4%	6.9%	1.8%	0.8%	0.3%
Gell-Mann & Ruhlen 2011	50.1%	38.0%	8.0%	2.0%	0.8%	0.6%
Average	45.2%	37.5%	11.4%	2.1%	0.6%	0.4%

Table 1: Unmarked Word Order Asymmetry Produced by C_{HL}

However, an anonymous reviewer points out many fundamental problems. Why should we focus on the ordering among *S*, *O*, and *V*? Is it not the case that *S*, *O*, *V* are the grossest levels of organization of the clause, hence encompassing the maximal level of complexity? Is it not the case that unmarked orders such as <*SOV*> and <*SVO*> are shadows, not the essential substances? Is it not possible that the unmarked <*SOV*> has many other derivations, hence leading to different varieties of unmarked <*SOV*>?[7] Why is <*SOV*> the base order? Why should the base order be the most common? If <*SOV*> is the cheapest, why is it not the case that all languages show <*SOV*> as the unmarked order? Why does an unmarked order such as <*OSV*> (0.5%) exist at all?[8] I attempt to answer these questions as far as possible. However, the questions are so fundamental that a complete answer is beyond the reach of this paper. Although the article faces many

<*OVS*>: Päri (Niger-Congo), Ungarinjin (moribund Australian aboriginal language), Hixkaryana (Carib language), Tuvaluan (Austronesian)
<*OSV*>: Kxoe (Kalahari), Tobati (Papua New Guinea), Wik Ngathana (Pama-Nyngan), Nadëb (Brazilian Amazon)

6 11% of languages are unclassified in this study. I thank an anonymous reviewer for pointing out that referring Yamamoto alone is insufficient.

7 The reviewer suggests that distinct operations yielding the same superficial <*SOV*> unmarked order, for example, are parametrized.

8 I thank an anonymous reviewer who pointed out this serious problem that my approach should solve. The readers can refer to Yang (2002) and Chomsky (2012) for the method behind the explanation of the statistical duality of irregular verbs. The reviewer's puzzle (a phylogeny issue) is particularly important in that it relates to an important statistical paradox (an ontogeny issue) as follows (Yang 2002, Chomsky 2012): Why do low-probability irregular verbs behave like high-probability regular verbs such that irregular verbs are as naturally and frequently used as regular verbs? Why do irregular verbs exist at all? Yang (2002) has discovered that irregular verbs are in fact 'regular' for they are grouped into distinct classes and the classes obey the relevant regular rules. For example, the blocking effect of the past tense form *went* over *goed* indicates that the 'weight' (probability) of the corresponding rule is 1.0 (must happen) or very close to 1.0 (very likely to happen) as a result of learning. Following his insight, I will argue later that a low-probability unmarked order such as <*OSV*> behaves like a high-probability order because the cost calculation is 'regular': The gross computational cost is within the threshold permitted for C_{HL} (the minimum cost). The blocking effect of unmarked <*OSV*> over <*SOV*> indicates that the 'weight' (probability) of the corresponding cost calculation is 1.0 (must happen) or very close to 1.0 (very likely to happen) as a result of cost equilibrium.

problems, let us first look at what typological studies have found with respect to the probability of unmarked (basic) word order asymmetry and see how far we can go within the geometrical cost approach.

Greenberg (1966) showed that *<SVO>* languages outnumber *<SOV>* languages, and Yamamoto (2002: 85) attributed this unlikely result to the smaller samples (30 languages) and a bias toward Indo-European and African languages, excluding the languages of New Guinea and Melanesia. The general ranking of unmarked word order seems to be clear:

(2) $SOV > SVO > VSO > VOS > OVS >? OSV$

It is significant that *<SOV>* and *<SVO>* account for more than 80%. C_{HL} is strongly biased for these two unmarked word orders. Can we say as follows? Starting from *<SOV>*, *<SVO>* involves flipping *O* and *V*, and *<VSO>* involves rotating one position rightward, *<VOS>* involves flipping *S* and *V* (or it is a one-dimensional mirror image of *<SOV>*). Where does the ranking in (2) arise from? Why does C_{HL} select this particular ranking? The main goal of this study is to show that the ranking is expressible as geometrical cost differences, which will ideally lead to a Galois-theoretic explanation, and that C_{HL} chooses the most cost-effective unmarked word orders with respect to the phylogeny (the issue of why we can observe the current probability regarding unmarked word order asymmetry in human language). However, it is also a fact that all six possible unmarked word orders show symmetry and they are each the result of the most efficient computation with respect to ontogeny (the issue of why all six word orders are respectively the most natural and frequent unmarked orders for the respective native speakers). In a sense, phylogenetically minor unmarked orders such as *<OSV>* are similar to irregular verbs because they show low probability (we do not find many samples) but simultaneously show high probability (they are the most natural, frequent, and unmarked orders for the respective native speakers). Why do minor unmarked orders show low probability but simultaneously show high probability? I will offer a possible answer to this paradox in the last part of Section 3. With regard to the basic statistical data, I tentatively adopt Yamamoto (2002) in the following sections because it contains the largest data set available at present (2,932 languages).

1.2. Chomsky's Third Factor

The biolinguistic approach tackles the problem of whether we can explain C_{HL} by natural laws, which Chomsky calls the third factor. The third factor includes "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).[9] Approximately half a century of biolinguistic

[9] The first factor is the human genome (the DNA and brain that yield properties of C_{HL} such

research has revealed that there are parts of C_{HL} that obey the principle of efficient computation, informally stated as follows:

(3) *Economy Principle (Minimal Computation)*
Select the most cost-effective computation.

Measures of effective computation include the least effort, the shortest distance, the closest element, the fewest steps, the simplest structure, and the minimal search. The initial state of C_{HL} is an organic computational system that includes the Economy Principle that governs an inorganic world. The initial state of C_{HL}, which is given by the human genome, undergoes parameter setting in a linguistic environment until C_{HL} reaches the final state, the point at which the mother-tongue acquisition system deactivates.[10] C_{HL} is a system that exhibits the discrete infinity property, which typically appears at the molecular level or below. A system of discrete infinity obeys the Economy Principle, such as a snowflake's hexagonal shape emerging as the idealized (optimized) realization of the atomic structure of H_2O in midair, free from the noise of gravity and earth's thick air. As Chomsky often mentions, it would be interesting if an inorganic principle were operating on organic matter such as the human brain.[11]

I assume that the group-theoretical principles of an algebraic structure belong to the third factor. Jenkins (2000, 2003) suggested that "word order types would be the (asymmetric) stable solutions of the symmetric still-to-be-discovered 'equations' governing word order distribution" (Jenkins (2000: 164) and that "the tools of group theory may be able to aid in characterizing the symmetries of word order patterns" (ibid.: 164).[12] I believe that a study of the

as discrete infinity and merge) and the second factor is the linguistic environment. The first factor is a force internal to C_{HL}, and the second and third factors are external forces (Yang 2000). The first and second factors are responsible for the ontogeny of C_{HL} (how C_{HL} grows in the brain of a human infant), while the third factor is responsible for the phylogeny (why C_{HL} has evolved in such a way). The interaction of these three factors determines the facts of C_{HL}. Boeckx (2009: 46) points out that Chomsky's 'three factors' resemble Gould's 'adaptive triangle' (Stephen Jay Gould, American paleontologist, evolutionary biologist, and historian of science; 1941–2002), which has three vertexes: (1) historical (chance); contingencies of phylogeny (mutation of DNA, 1st factor), (2) functional; active adaptation (environmental pressure, 2nd factor), and (3) structural constraints; rules of structure (physical laws, 3rd factor) (Gould 2002). See Uriagereka (2010) and Longa *et al*. (2011) for relevant discussions.

10 C_{HL} is generally active for mother-tongue acquisition until approximately the appearance of secondary sex characteristics. Many mysteries exist regarding the issue.

11 With regard to the connection between Hamilton's principle of least action in physics and the third factor in C_{HL}, see Fukui (1996). I thank an anonymous reviewer for suggesting that I should mention Hamilton's principle in this connection.

12 The assumption here is that an asymmetric state is stable; a symmetric state is too tense and expensive to maintain and such an unstable symmetric state becomes stabilized (costless to maintain) when the symmetry is broken. For example, Kayne (1994) proposes that syntactic terms must be in an antisymmetric c-command relation. Moro (2000: 15–29) claims that a symmetric structure (a point of symmetry) is too unstable for C_{HL} to tolerate and that symmetry must be broken, and this drives movement, stabilizing the structure. Di Sciullo (2005, 2008) investigates symmetry breaking (as a result of 'fluctuating asymmetry (oscillation)') in merge and morphology. In contrast, from the viewpoint of physics, a symmetric situation is stable (highly probable). An example is a gas, in which every direction appears the same. Symmetry forming is information diffusion and obeys the

algebraic structure of equations (Galois group) will help us to express the phylogeny problem concerning permutation asymmetry in C_{HL}. I attempt to express and translate the unmarked word order asymmetry into Galois-theoretic language, by considering cost.

The rest of this paper is organized as follows. In Section 2, I claim that C_{HL} produces the universal base *v*P, where *S* c-commands *O*, and *O* c-commands *V*, and that this base *v*P corresponds to the identity element (*I*) in mathematics. In Section 3, I propose that geometrical cost asymmetry is a possible "language" to express the unmarked word order asymmetry. I would like to propose that the unmarked ordering asymmetry in C_{HL} can be expressed by Galois-theoretic language: the third factor.[13] In particular, I propose a possible "equation governing [unmarked] word order distribution". Moreover, I also attempt to answer an important question: Why is it not the case that all languages show unmarked <*SOV*> provided that <*SOV*> derives from the most efficient computation? Section 4 summarizes the paper.

2. The Universal Base *v*P as the Identity Element

I propose that C_{HL} creates the universal base *v*P, which is the identity element (identity syntactic relation) under the Merge operation.[14] The base *v*P has the c-command relation $S \gg O \gg V$, as shown in Figure 1.[15] The base *v*P is formed

entropy law: Disorder develops (the second law of thermodynamics). Symmetry breaking is information condensation and disobeys the entropy law, i.e., order develops. An example is a crystal, in which things look different according to the viewpoint. For Kayne, Moro, and Di Sciullo, structure building is symmetry breaking, which produces information, disobeying the entropy law. On the other hand, Fukui (2012a) proposes that F(feature)-equilibrium (symmetry formation) drives structure building. F-equilibrium obeys the entropy law. For Fukui, structure building is symmetry formation: information loss. There is no contradiction. Kayne, Moro, and Di Sciullo discuss how structures produce phonetic (sound) and semantic (meaning) information, which must not be deleted, whereas Fukui talks about how structures lose formal features (structural information), which must be deleted.

The issue is related to a diachronic question that an anonymous reviewer asks as follows: What will happen to the synchronic unmarked order asymmetry? Will all languages become <*SOV*> type, provided that it is the most efficient? Although the diachronic issue is beyond the reach of this paper, at this point, let us tentatively assume as follows. The diachronic change may be determined by the dynamic interaction between the two forces noted above: symmetry breaking and symmetry preservation (formation).

13 An anonymous reviewer suggests that S-initiality is largely areal (geographical proximity of other S-initial languages) (Dryer 2012). If so, we should conclude that it is primarily the environmental factor that induces the unmarked word order asymmetry. Although the issue is beyond the scope of this paper, let us tentatively start with the view that all three factors (genetic, environmental, and physical) are involved in the asymmetry in question.

14 I focus on the structure of a simple matrix transitive sentence (consisting of *S*, *O*, and *V*) that the relevant native speakers judge to be the unmarked (basic) word order (actually their C_{HL} reaction). C_{HL} is what motivates the universal base *v*P. I thank an anonymous reviewer for pointing out this unclarity. I call the universal base *v*P the base *v*P for simplicity.

15 The definition of c-command is as follows (Uriagereka 2012: 121):

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

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 *v*P. If TRANSFER applies to the base *v*P, the phonological component Φ (sensorimotor interface) produces <*SOV*> as the unmarked order.[16]

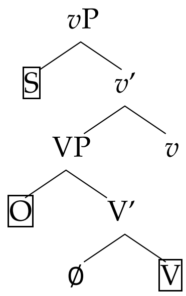

Figure 1: The universal base vP

Why is this structure the universal base *v*P?[17] First, it is the most cost-effective structure: the base *v*P is built by external merges only. If the cost is zero, the base *v*P 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 *v*P at its deepest structure. Third, it gives us semantic universality: The base *v*P 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*.[18] Thus, C_{HL} disallows other possibilities.

C-command expresses a balance (equilibrium) between disconnection (i) and connection (ii) in a tree (Chomsky 1995: 339).

16 TRANSFER (Spell Out) sends a halfway-built tree with sound information to Φ. The relevant derivation may involve movements in later steps. An anonymous reviewer asks an important question in this connection: Is it not the case that <*SOV*>, for example, is always re-derived many times or has many sources? I tentatively assume that the geometrical cost approach mapping a tree to an unmarked word order is compatible with the conception that an unmarked word order (output) derives from many source trees (input) because a function allows many-to-one correspondence (Stewart 1975). For unmarked <*SOV*> and <*SVO*>, let us assume that the c-command relation within the *v*P phase at the point of the first TRANSFER determines the unmarked order.

17 I thank an anonymous reviewer for pointing out this crucial question. In an earlier draft, I adopted the view that *O* moves to Spec, *v*P for feature checking. The reviewer pointed out that such a *v*P 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 universal base *v*P; it is constructed by an external merge alone, which yields the simplest possible architecture for *S*, *O*, and *V*.

For phylogeny, the third factor (geometrically lowest cost) determines the six unmarked word orders. But for ontogeny, capitalizing on Yang (2002: 72), who argued that 'irregular'-verb formation is in fact 'regular' in that a child acquires 'irregular' verbs by applying 'regular'-class-forming rules, I propose that a child reliably associates an 'irregular' (minor) order (*OSV, VOS, OVS*) with its matching 'irregular'-formation rule, and reliably apply the rule over the default <*SOV*>. The ontogeny (acquisition) of 'irregular' (minor) unmarked orders parallels that of 'irregular' verbs. See section 3 for a detailed discussion.

18 There is much evidence which indicates that *V* merges with *O*. *V* selects *O* (e.g., the *V say*

Let us demonstrate how the base *v*P 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 ∅ (an axiom). *V* externally merges with ∅.[19] *V′* and *O* merge, and *V* assigns Patient θ (a semantic role) to *O*.[20] The light verb *v* merges with VP. The *v′* merges with *S* and *v* assigns Agent θ to *S*. Thus, the base *v*P 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 *v*P. Every final structure contains the base *v*P as a subset, and the base *v*P does not affect the usable c-command relations in the final structure. As noted above, the base *v*P is like the identity element 0 (zero) in addition. Probe features in *v* agrees with the goal features in *O*, the relevant structural features are valuated and deleted (Chomsky 2000).[21] The valued

---

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).

19 An anonymous reviewer points out that construing the empty set as a legitimate syntactic object is something new and that it should be justified. The reviewer points out that it poses a problem because in set theory, the empty set is a subset of every set, not an element of every set. I tentatively adopt the following definition of syntactic object in the bare phrase structure model (Chomsky 1995: 243, 262). I reintroduce the relevant definition stated in Uriagereka (2000: 497).

(i) *Syntactic object*
σ is a syntactic object if it is
a. a lexical item or the set of formal features of a lexical item, or
b. the set K = {γ, {α, β}} or K = {$\underline{<\gamma,\ \gamma>}$, {$\alpha$, β}} such that α and β are syntactic objects and γ or $\underline{<\gamma,\ \gamma>}$ is the label of K.

If the *set* of formal features of a lexical item is a syntactic object as in (ia) and if the phonologically empty set lacking any member (phonological feature) is a legitimate phonological object, the syntactically empty set lacking any member (syntactic feature) may also be a legitimate syntactic object. As an alternative, the reviewer suggests 'Self-Merge' that allows vacuous projection, as in Guimarães (2000) and Kayne (2008). I leave open this fundamental problem for future research. See Barrie (2006: 99–100) for the solution adopted here, 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 consists of empty set (∅) before *V* enters, i.e., "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 successor function" (Chomsky & McGilvray 2012: 15). The operation also satisfies the restriction that "Merge cannot apply to a copy: a trace or an empty category that has moved covertly" (Chomsky 2004). The empty set ∅ is not a copy or an empty category that has moved covertly. Therefore, ∅ is allowed to merge with V. "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 *ø* is a set" (Stewart 1975: 48). "[T]he empty set ∅ 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 ∅ which was not an element of *S*. In particular there would have to be an element of ∅. Since ∅ has no elements, this is impossible" (ibid.: 49). See also Fukui (2012b: 259) for the hypothesis that 1 is created by merging 0 with 1. If the natural numbers emerged from the abstraction from merge, the sentence-structure building must involve the empty set merging with V at the first step.

20 An intermediate projection such as *V′* is used for expository purposes.

21 The base *v*P is consistent with the Multiple Spell Out (MSO) hypothesis (which states that there is more than one point when a structure with sound features attached is sent to the PF (Φ) (Uriagereka 2012: 113, fn. 33). According to MSO, a domain, such as *S*, that is moved to

φ-feature is deleted because it is redundant: O contains the same φ-set in the first place. The valued structural Case is deleted as a reflex (side effect) of valued-φ deletion (ibid.: 122). If a formal feature is not deleted within C_{HL} and enters into the external performance systems (Φ and the thought system Σ), the external systems will freeze because such a structural feature is unknown to them.

The base vP is the most economical structure (involving the least effort) 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. LCA demands that a structurally higher term should be pronounced earlier. Let us adopt the following definition of LCA (Uriagereka 2012: 56).[22]

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

The base vP does not influence later structures. For example, suppose we arrived at $V \gg S \gg O$ as the final output structure of TRANSFER. In Φ, LCA notices only the boxed terms in Figure 2.[23] There, TRANSFER (Spell Out) sends the final CP structure to Φ, and LCA maps this structure to the linear unmarked order <*VSO*>.[24] 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 (zero) for addition).[25]

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: it incorrectly predicts that the moved O should exhibit the island effect. The universal base vP hypothesis rejects Kayne's (1994) hypothesis that structure building starts with the base VP in which S c-commands V, which c-commands O. See Fukui & Takano (1998) for arguments for our hypothesis.

22 The original definition of LCA is as follows (Kayne 1994: 6). Given $d(X)$ = the set of terminals T that X dominates and A = the set of ordered pairs $<X_j, Y_j>$ 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 = def. $d(A)$ is a linear ordering of T.

23 With regard to the V-initial unmarked order, there is a debate on the derivation, i.e. remnant-VP movement vs. V-movement. For arguments for the former view, see Alexiadou & Anagnostopoulou (1998) and Massam (2000). I use a V-movement analysis for simplicity. The choice does not affect the discussion. See Carnie *et al*. (2005) for relevant discussions.

24 If T contains EPP and attracts S, V must have reached C at the point of the final TRANSFER for the unmarked order <*VSO*> to be realized.

25 The tree building in C_{HL} constitutes a group. It conforms to the four definitions of a group. First, it is closed: Merge applies to a tree and it creates a tree. Second, it has an identity element: the universal base vP is similar to 1 for multiplication; it does not affect the output. Third, it has inverse elements: there is always a set of remerge operations that returns some c-command relation to the base $S \gg O \gg V$ relation. Fourth, it obeys the associative law, (XY)Z = X(YZ), with respect to structure building (head projectionability); given the head-final property, both (XY)Z and X(YZ) produce a projection of Z. Alternatively, given the head-initial property, both (XY)Z and X(YZ) produce a projection of X. With regard to the fourth condition, Fukui & Zushi hold the view that C_{HL} disobeys the associative law for semantics, i.e., distinct hierarchical (binary) structures produce distinct meanings (Merge disobeying the associative law causes the hierarchical structures). See their comment on pages 19 and 322 of the Japanese translation of Chomsky (1982, 2002).

If Merge is the fundamental operation in C_{HL} and the concept of 'group' applies to any system with the possibility of combining two objects to yield another (Stewart 1975: 1), C_{HL} deserves a group-theoretical analysis. "Thus the concept 'group' has applications to

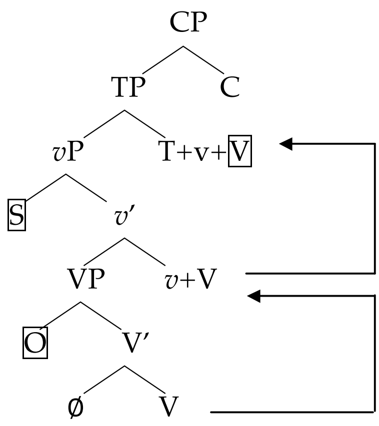

Figure 2: V ≫S ≫O c-command relation mapped to unmarked <VSO>

3. Word Order Asymmetry as Geometrical Cost Asymmetry

The symmetry structure of an equilateral triangle represents the group-theoretical structure of a cubic equation (Stewart 2007).[26] The permutation of three solutions corresponds to that of the three vertexes. Assume counterclockwise rotations, with a 0° rotation serving as the identity *I*.[27] Let us call the original triangle as the identity element or identity triangle.

---

rigid motions in space, symmetries of geometrical figures, the additive structure of whole numbers, or the deformation of curves in a topological space. The common property is the possibility of combining two objects of a certain kind to yield another" (ibid.).

26 I thank an anonymous reviewer for clarifying the issue. That is, the permutation group S_3 of three letters have only 4 isomorphism classes (or conjugacy classes) of subgroups, namely, {id} = *I*, C_2 (a cyclic group of order 2), C_3 (a cyclic group of order 3) and S_3. The reviewer criticizes that the observed broken symmetry corresponds most closely to C_2, amounts to a rather simple observation that *V* and *O* seem to remain symmetric whereas *S* is not symmetric with others. Here is the list of six subgroups of S_3. (23) stands for the permutation that switches 2 and 3, leaving 1 intact, as in (1, 2, 3) → (1, 3, 2). (132) stands for the permutation that changes 1 to 3, 3 to 2, and 2 to 1, as in (1, 2, 3) → (3, 1, 2). *I* is the identity permutation that keeps everything intact, as in (1, 2, 3) → (1, 2, 3). Assume 1 = *S*, 2 = *O*, 3 = *V*.

(i) a. {*I*, (23), (13), (12), (132), (123)} = S_3
 b. {*I*, (132), (123)} = A_3
 c. {*I*, (12)}
 d. {*I*, (13)}
 e. {*I*, (23)}
 f. {*I*}

Every subgroup contains *I*, which is (*S*, *O*, *V*) → (*S*, *O*, *V*). This might partially express the C_{HL} fact that it is the highest probability that the identity transformation maps the universal base *v*P onto <*SOV*> unmarked order.

27 A 0° and 360° cannot be distinguished group-theoretically, but they are distinct if we take the cost difference into consideration.

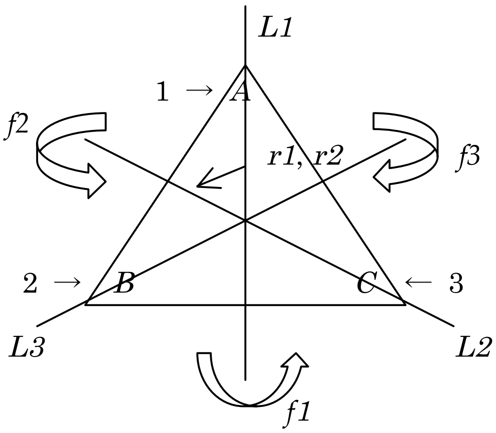

Figure 3: Symmetrical Operations of an Equilateral Triangle

An equilateral triangle has six symmetrical operations: rotations *r* (cyclic permutations) and reflections *f* (flips or non-cyclic permutations) indicated in (5).[28]

(5) a. *r0* = 0° = *I* (do-nothing rotation)
b. *r1* = 120° rotation
c. *r*2 = 240° rotation
d. *f1* = Flip around axis *L1*
e. *f2* = Flip around axis *L2*
f. *f3* = Flip around axis *L3*

The do-nothing operation *r0* changes <*ABC*> to <*ABC*>. The top apex corresponds to the first position, the lower left apex to the second, and the lower right apex to the third. The six transformations are as follows:

(6) a. *r0* changes <*ABC*> to <*ABC*>.
b. *r1* changes <*ABC*> to <*CAB*>.
c. *r2* changes <*ABC*> to <*BCA*>.
d. *f1* changes <*ABC*> to <*ACB*>.
e. *f2* changes <*ABC*> to <*BAC*>.
f. *f3* changes <*ABC*> to <*CBA*>.

The transformation *r0* is the most cost-effective. Although Galois groups are indifferent to cost, geometrical operations do have cost differences, given an appropriate cost function. It is true that the structure of the symmetric group S_3 of order 3! (6) is too simple to imply anything. However, this simplicity is the very reason why I take operational costs into consideration.[29] All six symmetrical

[28] Rotations are linear transformations *T* (or function *f*) in R^2 (two-dimensional real-number space). Flips are *T* of R^2 subspace in R^3 (Strang 2009). *T* or *f* can be translated into a matrix *A*. If the unmarked order asymmetry can be expressed by *T*, we will be able to translate it into the matrix language, which we leave for future research.

[29] Algebraic cost means computing time (Strang 2003: 87). An anonymous reviewer offered the criticism that the structure of the symmetric group is too simple to imply anything. I thank

trans- formations can be expressed using only *r0*, *r1*, and *f1*, that is, *r2*, *f2*, and *f3* are derivable operations (Armstrong 1988).[30]

(7) a. *r0*
b. *r1*
c. *r2 = r1 + r1*
d. *f1*
e. *f2 = f1 + r1*
f. *f3 = r1 + f1*

Why do we select *r0*, *r1*, and *f1* as irreducible atoms for symmetrical transformations?[31] Recall that we started from an empirical (physical) fact about the human brain: C_{HL} produces a sentence structure with the base *v*P as its universal base, in which *S*, *O*, and *V* are externally merged such that *S* asymmetrically c-commands *O*, which in turn asymmetrically c-commands *V*. The base *v*P is the most cost-effective base with a cost of 0: it is built by external merges alone. Therefore, the base *v*P corresponds to *r0*, the identity operation (with a cost of 0). Since we use the cost differences between transformations, we have to rank transformations by their geometrical cost. After *r0*, the next most cost-effective operation is *f1*, which switches two (rather than three) positions, *O* and *V*. Because *f1* switches *O* and *V*, which have a strong bond, as stated earlier, and which form a natural class, *f1* is the most cost-effective transformation among flips (or reflections). Following *r0* (cost 0) and *f1* (cost 1), *r1* (with cost 2; it is a single-step rotation with three (rather than two) positions replaced) is the second most cost-effective transformation within the rotations.

Let us summarize cost calculation. Suppose that the identity operation *r0* has cost 0. The geometrical operation *r0* syntactically corresponds to doing nothing to the least costly base *v*P before spell-out, which in turn sent to Φ where LCA produces the linear order <*SOV*>. The more positions that a computation replaces, the more energy the computation uses.[32] This is the reason why *r1* is costlier than *f1*.[33] Furthermore, single-step operations are cheaper than two-step operations — mathematicians call this the 'length function' in symmetric groups.[34] Hence, *r0* is the cheapest of all, *f1* is the second cheapest, and *r1* is the third.[35] Assuming that *f1* has cost 1, *r1* has cost 2, and that addition is used for

---

the reviewer for clarifying the crucial reason why I should consider geometrical cost, namely it sharpens the tool for observing the phenomena.

30 I stipulate that the vertical axis *L1* is the default (basis). An empirical reason is as follows. Given that the base *v*P corresponds to an equilateral triangle in which *S* is the top vertex, *O* is on the left, and *V* is on the right, the vertical axis *L1* switches *O* and *V*. There is considerable evidence that *V* has an affinity for *O*, rather than *S*. That is, given, *S*, *O*, and *V*, {*O*, *V*} constitutes a natural class excluding *S*, whereas {*S*, *V*} excluding *O* does not. The vertical axis *L1* switches elements in a natural class.

31 I thank an anonymous reviewer for pointing out the necessity of clarifying this reasoning.

32 "[I]n group theory it is the end result that matters, not the route taken to get there" (Stewart 2007: 121). However, the route matters for the geometrical cost approach: A longer route is more expensive.

33 I thank an anonymous reviewer for pointing out unclarity in an earlier draft.

34 I thank an anonymous reviewer for pointing this out.

35 This cost function is consistent with results under the Mobius function, according to which

cost accumulation, the costs for the six transformations are as follows:

(8) $r0 = 0$
$r1 = 2$
$r2 = r1 + r1 = 2 + 2 = 4$
$f1 = 1$
$f2 = f1 + r1 = 1 + 2 = 3$
$f3 = r1 + f1 = 2 + 1 = 3$

The identity operation *r0* is the cheapest (cost 0) followed by *f1* (cost 1) and *r1* (cost 2).[36] This is what we would expect if we replaced *A*, *B*, and *C* with *S*, *O*, and *V*, respectively.[37] The identity triangle looks like the following:

---

the equation for flip is simpler than that for rotation.

36 An anonymous reviewer asks a subtle and extremely important question: Exactly what are the relevant 'costs' to be minimized, provided the economy principle? I adopt the view that algebraic cost means computing time (Strang 2003: 87). The longer the root, the more time it takes. Therefore, the relevant 'cost' to be minimized is computing time. The high probability of unmarked <*SOV*> from phylogenetic point of view emerges from the fact that the identity (do-nothing) transformation is the fastest computation. Also, I thank an anonymous reviewer for pointing out a miscalculation in a previous draft and for clarifying the reason for selecting smaller values. The reasoning is as follows. For *f2*, there are three sets of operations that lead to the same result: $f2 = f1 + r1 = 1 + 2 = 3$, $f2 = r2 + f1 = 4 + 1 = 5$, and $f2 = r1 + f1 + r2 = 2 + 1 + 4 = 7$. For *f3*, there are two sets of operations that lead to the same result: $f3 = r1 + f1 = 2 + 1 = 3$, and $f3 = f1 + r2 = 1 + 4 = 5$. I select the lowest cost for each, assuming that C_{HL} obeys the Economy Principle. Therefore, $f2 = f1 + r1 = 3$, and $f3 = r1 + f1 = 3$.

37 A reviewer points out that "[these] permutations on the SOV 'basic' string as the relevant group-theoretic action" is "the source of the most severe problems". However, what is 'basic' is not the SOV string itself. What is 'basic and universal' is the *v*P structure without internal merge (copy and remerge) at the point of TRANSFER (movements may occur later). The universal base *v*P *per se* is not the unmarked <*SOV*> order. TRANSFER applies to the universal base *v*P and Φ outputs <*SOV*> as a possible unmarked order for a simple matrix transitive sentence. The reviewer also has severe doubts on "the author's technique of considering string permutations rather than movement operations in the tree." However, I do not propose string permutations as a new technique to analyze sentence structures. Rather, I claim that movement operations in a tree (including no movement) can be expressed as the group-theoretical transformations of equilateral triangle. The movement operations and the geometrical transformations are compatible and translatable. If a certain structure (order) is not derivable due to a violation of the movement constraint, there is no geometrical expression for it. We consider how a per- mitted tree structure can be expressed algebraically and geometrically. The group-theoretic action acts on an equilateral triangle in a certain coordinates (which is a geometrical expression of a particular permutation of three solutions of a cubic equation). A triangle undergoes various linear transformations in R^2 (e.g., rotations in two-dimensional real-number space) and R^3 (e.g., reflections (flips) in three-dimensional space). However, I admit that the geometrical cost approach does rely on the universal base *v*P as the identity element. If that approach is untenable (as the reviewer points out), the geometrical cost approach collapses.

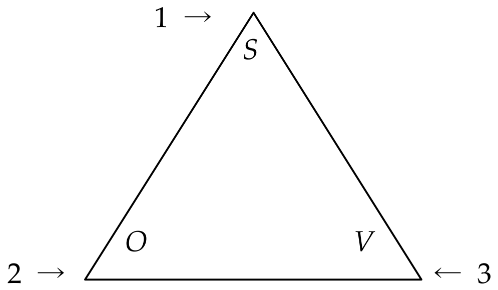

Figure 4: Identity Triangle Expresses the Universal Base v*P*

Internal merge operations including the lack thereof apply to the universal base *v*P, and the LCA produces various unmarked order types in Φ. This situation is geometrically expressed as symmetric transformations applied to the identity triangle, producing various permutations. Table 2 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 2: Transformations and Costs for {S, O, V}

Following Jenkins (2000, 2003), I speculate that the unmarked word order asymmetry is expressible as a group-theoretical factor (included in Chomsky's third factor): "[W]ord order types would be the (asymmetric) stable solutions of the symmetric still-to-be-discovered 'equations' governing word order distribution". The 'symmetric equation' is a linear transformation $f(x) = y$, where function f (or transformation T) is a set of merge operations that is expressed as a set of symmetric transformations of an equilateral triangle (or permutations of three solutions of a solvable cubic equation), x is the universal base *v*P input that is expressed as the identity triangle, and y is a mapped output tree that is expressed as an output triangle that preserves symmetry. The equation $f(x) = y$ can be translated into the matrix language: $Ax = y$, where A is a matrix that performs the transformation, x is a set of input vectors expressing the identity triangle (the universal base *v*P), and y is a set of output vectors expressing the transformed symmetrical triangle (the transformed tree).[38] The Galois theory and the Economy Principle (choose the cheaper operation) can express the current ratio of languages with the top three unmarked word orders:

[38] See Strang (2009) for the basic idea of linear transformations. The condition that a linear transformation must satisfy is as follows: $T(cv + dw) = cT(v) + dT(w)$, where T is a linear transformation, v and w are some vectors, and c and d are some constants. Projections and rotations are examples of linear transformations.

(9) 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:

(10) a. *f2* (cost 3) produces <*OSV*> with a ratio of 0.5%.
b. *f3* (cost 3) produces <*VOS*> with a ratio of 2.4%.
c. *r2* (cost 4) produces <*OVS*> with a ratio of 0.7%.

The geometrical cost approach predicts that <*OSV*> and <*VOS*> should emerge at the same rate, and that <*OVS*> should exhibit the lowest rate, which is not reflected in the actual statistics. We are not able to predict this difference. However, it is significant that the approach predicts the internal ranking of the major (top) three unmarked word orders and the division between the higher three and lower three with respect to unmarked word order in C_{HL}.[39]

What is symmetry? A state is symmetrical when an operation (or a transformation) does not affect (change) the properties of the state. However, some properties are preserved after transformation (symmetry is formed), whereas some properties are not preserved (symmetry is broken). What properties are preserved and not preserved here? The preserved property is the structure of the equilateral triangle itself located in particular coordinates (the entire shape looks the same after symmetrical transformations); information regarding the locations of *S*, *O*, and *V* is irrelevant. We observe the same-looking equilateral triangle after various symmetrical operations. The property not

[39] With regard to <*OSV*>, I tentatively propose that *O* raises and becomes the Spec, TP. The operation is very expensive because C_{HL} must find (and actually finds) a solution to circumvent a violation of the minimality principle; T has attracted *O*, which is more distant than *S*. With regard to <*OVS*>, V further raises to T. With regard to <*VOS*> (e.g., Austronesian languages such as Malagasy, Seediq, and Tzotzil), V further raises to C. However, the analysis wrongly predicts that the probability difference should be *OSV* > *OVS* > *VOS*. As an anonymous reviewer points out, the currently available difference is unexpectedly the opposite: *VOS* > *OVS* > *OSV*. Why should <*VOS*> be the most probable among the three? It may be that V-movement to C facilitates *O*-movement, as in Object Shift phenomena. As for the conditions on Object Shift, see Chomsky (2000). Alternatively, it may be related to the mathematical fact that "Inverses come in reverse order" (Strang 2003: 72). That is, $(SOV)^{-1} = V^{-1}O^{-1}S^{-1}$. In other words, <*VOS*> could be an inverse of <*SOV*>. Therefore, $(SOV)^{-1} \times (SOV) = (V^{-1}O^{-1}S^{-1}) \times (SOV) = I \times I \times I = I$. <*VOS*> shows relatively high probability because it is in inverse relation with the highest probable order, <*SOV*>. However, neither the exact nature of the derivation nor the linear algebraic reasoning is clear at this point.

Furthermore, a question arises as to why the unmarked word orders <*OSV*>, <*VOS*>, and <*OVS*> exist at all; i.e. why do they not show 0% if they are very expensive? From the perspective of phylogeny, I propose that these unmarked word orders are rare (minor) because they have higher geometrical cost. However, from the perspective of ontogeny, I propose, capitalizing on Yang (2002: 69–70), that they exist because they have higher weight (probability that is one or very close to one as a result of learning). These rare (minor) unmarked orders are like 'irregular' verbs: Every 'irregular'-forming rule, which applies to the verb class, is associated with a weight (probability). As a child acquires 'irregular' verbs by applying 'regular' class-forming rules, she acquires a 'minor' basic word order by applying 'regular' transformation (phrasal and head movement) rules.

preserved is the locational information of *S*, *O*, and *V* regarding where *S*, *O*, and *V* end up in the triangle after symmetrical transformations. We observe different arrangements of *S*, *O*, and *V* after various symmetrical operations. However, the identity (do-nothing) transformation is special in that it always preserves all properties after symmetrical operations.

A derivation of a sentence starts out with the universal base *v*P, in which *S* c-commands *O* and *O* c-commands *V*. If the base *v*P (without movement) is transferred to Φ, we obtain <*SOV*> as the unmarked order. This is geometrically expressed as the identity transformation where nothing is done. If *V* raises to *v* (one-step V-movement) before TRANSFER, we obtain <*SVO*> as the unmarked order. This is geometrically expressed as a flip (three-dimensional transformation) where we have *V* in the base-*O* position and *O* in the base-*V* position (*O* and *V* are switched). If *V* raises to *v* and then to *T* (two-step V-movement) before TRANSFER, we obtain <*VSO*> as the unmarked order. This is geometrically expressed as a 120° rotation, where we have *V* in the base-*S* position, S in the base-*O* position, and *O* in the base-*V* position. The structure-building cost corresponds to the geometrical cost. This causes the probability difference among the three major basic word order types from the phylogenetic viewpoint. Let us summarize the C_{HL} geometry correspondence in the following figures. The boxes in the trees are visible to Φ and to the LCA spelling out the unmarked word order.

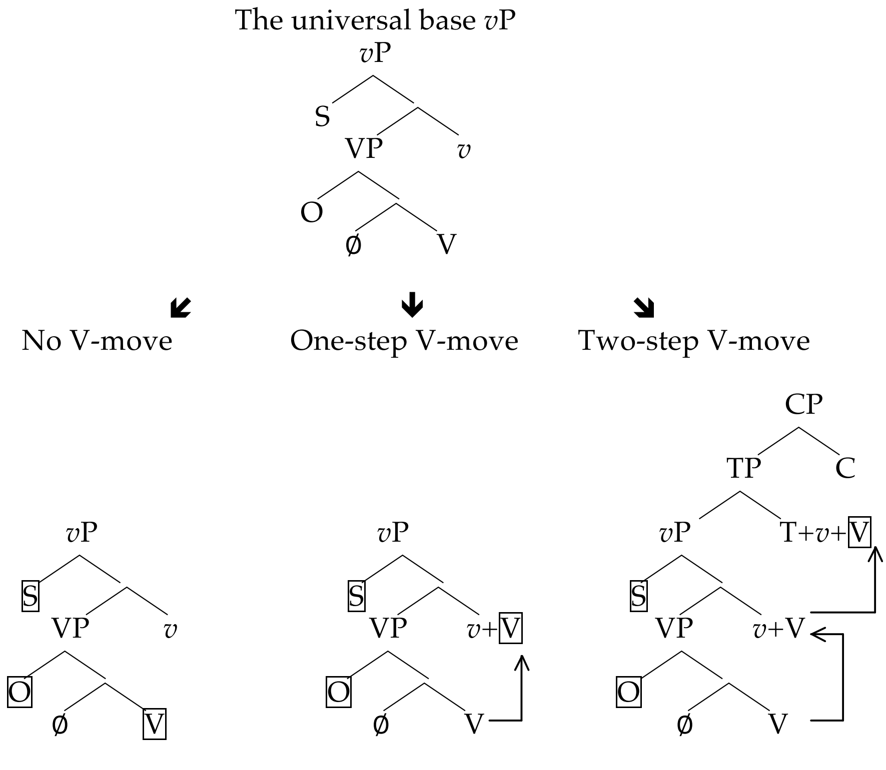

Figure 5: C_{HL} Transformation Deriving the Major Three Unmarked Orders

The above tree-building steps can respectively be expressed as (Galois-theoretic) geometrical transformations (rigid movements) as follows. These geometrical transformations express various permutations of the solutions of the solvable cubic equation.[40]

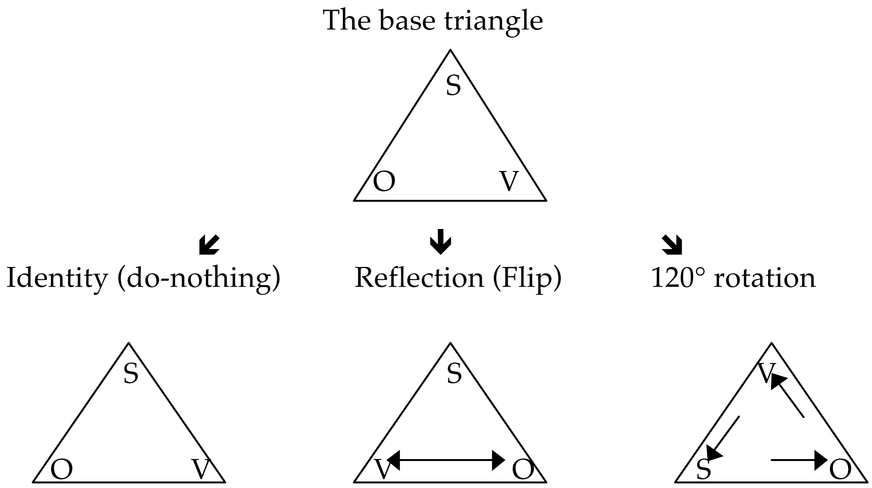

Figure 6: Geometrical Transformations Deriving the Major Three Unmarked Orders

Our analysis is consistent with the conception that "[o]ptimally, linearization should be restricted to the mapping of the object to the SM [sensorimotor] interface [Φ], where it is required for language-external reasons" (Chomsky 2005). The geometrical cost belongs to a mathematical or physical law that is language external. In Addition, our model supports the view that "order does not enter into the generation of the C-I [thought] interface," and that "syntactic determinants of order fall within the phonological component" (Chomsky 2008). In other words, the permutation among *S*, *O*, and *V* does not influence the meaning of the matrix simple transitive sentence in all languages in the thought system: the idea of "John loves Mary" is the same in all languages, whatever the unmarked order is; symmetry is maintained. On the other hand, with regard to the ordering that takes place in Φ, symmetry breaks in a manner that obeys a mathematical or physical law (except the do-nothing (identity) operation). Ordering is not accidental or random, contra Chomsky (2012).

However, it is also a fact that all six unmarked-order types behave alike in that they are all possible mother languages; each type is the most natural, frequent, and unmarked word order for the respective native speakers. The computational cost for basic order formation must be within the permissible level in all types; the relevant computation is equally efficient in all languages.

An anonymous reviewer asks a crucial question: Is it not the case that C_{HL} must produce the unmarked <*SOV*> only, provided that the unmarked <*SOV*> derives from the most efficient computation and that C_{HL} obeys the principle of efficient computation? Why does C_{HL} allow other unmarked orders that derive from less efficient computation? Why does the unmarked <*OVS*> for example exist at all, given that it derives from the least efficient computation? Is it not the case that the unmarked <*OVS*> cannot exist? Why does it exist at all?

---

[40] It is not clear how the relevant cubic equation looks like at this point.

A tentative answer is as follows. Suppose that the gross computational cost is 1.0 in all languages and that C_{HL} allows all possible patterns as long as the gross cost is 1.0.[41] If the basic (unmarked) word order is <*SOV*>, approximately cost 0.1 is used for the unmarked order building and the rest (cost 0.9) is used for other operations. If the basic word order is <*SVO*>, cost 0.2 is used for the unmarked order building and the rest (cost 0.8) is used for other operations. If the basic word order is <*VSO*>, cost 0.3 is used for the unmarked order building and the rest (cost 0.7) is used for other operations.[42] For example, the <*SOV*> type has the greatest cost 0.9 remaining for other operations. Thus, an <*SOV*>-type language such as Japanese tends to allow computationally more complex operations in other domains: this type allows (phonologically) null subjects, null expletives, null agreement morphologies, covert (phonologically null) wh-movement, covert extraction of argument-wh phrases out of islands, and scrambling (relatively free word ordering).[43] The C_{HL} needs more energy to locate where these silent entities are, how they are moving, and where they went because they are not heard (not pronounced); they are difficult to find and keep track of.[44] Therefore, our model predicts that the <*SVO*> type, unlike the <*SOV*> type, is less tolerant toward these phonetically null entities and word permutations. The prediction is borne out as comparative syntactic studies have observed: an <*SVO*>-type language such as English tends not to allow covert subjects, covert expletives, covert agreement morphologies, covert wh-movement, covert extraction of wh-phrases out of islands, and scrambling. In addition, our analysis predicts that the <*VSO*> type is furthermore less tolerant toward these phenomena.[45] We leave the detailed verification for future research. Let us summarize our point in Table 3.

---

[41] Notice that the number 1.0 is tentatively used here for maximum level of computational cost, not the probability 1.0 (it must happen).

[42] The specific numbers expressing cost do not matter. What matters is the difference.

[43] Covert extraction of adjunct-wh phrases out of islands is not allowed even in this type. The computational cost exceeds the threshold level (cost 1.0) at this point.

[44] This idea is the opposite of the standard conception that covert entities and operations need less energy because the costly pronunciation is not necessary.

[45] Unlike <*SVO*>-type languages such as English, <*VSO*>-type languages such as Irish (exclusively <*VSO*>) and Tagalog tend to show severer restrictions on covert elements and word permutations. For example, they require a phonetically realized question marker at the beginning (or the second position) of the question sentence; V-initial languages have pre-V particles (C?), C has a more elaborate system of phonetic realization with respect to feature combination of [±Q] and [±WH], which restricts cyclic wh-movement (Irish), *wh*-fronting is obligatory (Irish), the patient wh-phrase, but not the agent wh-phrase, is fronted in the matrix simple transitive question (Tagalog) (Aldridge 2002: 394), an argument movement to the left edge is strictly disallowed (Irish), and null subject is more strictly constrained; a pronoun must appear when V takes an analytic form (Irish), and ordering within nominals is more restricted (strictly head-initial), i.e., nouns must precede demonstratives, adjectives, or relative clauses; and inverted order is prohibited in questions. These observations indicate that the <*VSO*> type is much less tolerant toward covert elements and word permutations than the <*SVO*> type. See Carnie *et al*. (2005) for more information.

	<*SOV*>	<*SVO*>	<*VSO*>
Cost of basic order formation	0.1	0.2	0.3
Cost left for other operations	0.9	0.8	0.7
Gross cost used in C_{HL}	1.0	1.0	1.0

Table 3: Cost is balanced

Assume that the gross cost-level for C_{HL} operations is the same in all languages. In addition, assume that the number of parameters is the same in all languages, i.e., the cost for language acquisition is the same. With regard to <*SOV*>, less parameters are fixed for determining the unmarked order, and more parameters must be fixed for other operations. With regard to <*VSO*>, more parameters must be fixed for determining the unmarked order, and less parameters are fixed for other operations. However, the gross cost is the same in all languages. Our analysis is compatible with the conceptions that "[c]omplexities [expensive computation] in one domain of language are balanced by simplicity [inexpensive computation] in another domain", "[a]ll languages are necessarily equally complex [the gross cost is 1.0]", and that "[c]omplexity trades off between the subsystems of language."[46]

4. Conclusion

I am grateful to the anonymous reviewers for teaching me reality: My approach may be too simple, immature, groundless, and without promise, and my research has a long way to go even if it should turn out to be tenable. The reviewers pointed out several faults. First, S_3 is 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_{HL} works group theoretically. Third, the classification based on S, O, and V may be too crude for samples. Fourth, it may be too simple to assume that the derivation of the unmarked <*SOV*>, for example, is done in only one way; there may be many ways to derive the unmarked <*SOV*>. The reviewers 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. I hope that my approach will lead to possible future research from the combined perspective of applied mathematics and biolinguistics.

Despite tons of difficulty, let us ask the following question. What would it mean for the geometrical cost approach to express the basic word order asymmetry in C_{HL}? What does it mean to say that the basic word order asymmetry can be expressed as solving a cubic (or complex quadratic, whatever) equation? What does it mean for the categories as *S*, *O*, and *V* to be described as the roots of an equation?[47] Following Noam Chomsky, I speculate that these

[46] Fenk & Fenk (2008), Nematzadeh (2013). See p. 329 in the Japanese translation of Chomsky (1982) for a possible hypothesis that every individual language shows the same cost level.

[47] I thank an anonymous reviewer for pointing out the necessity of asking these questions in order to provide the *raison d'être* for this project. According to the reviewer, in Jenkins' for-

questions lead us to a partial answer to the traditional question that troubled Alfred Russel Wallace (a British naturalist, explorer, geographer, anthropologist, and biologist; 1823–1913), co-author of the evolutionary theory of natural selection, 124 years ago. Chomsky (2005: 16, 2007: 7, 20, 2010: 53) quotes Wallace's puzzlement: 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.[48] In favor of Leopold Kronecker (a German mathematician; 1823–1891), who said that God (Mother Nature) 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."[49] Considering Merge within the context of the evolutionary theory, Chomsky (2007: 7) proposed the following hypothesis:

(11) 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.

(12) A simple matrix transitive sentence consisting of *S*, *O*, and *V* can be expressed as a solvable equation with an algebraic-geometrical structure.

If this is true, we can study C_{HL} with Galois-theoretic tools.[50] As a Galois group characterizes the algebraic (or symmetry-related) structure of an equation, it can also characterize the algebraic (or symmetry-related) structure of a sentence at a relevant level.

Let us summarize the discussion through the key points listed in (13):

---

mulation, the idea was that *the word orders themselves* (not the coarse individual categories *S*, *O*, or *V*) are the solutions to the equations governing syntactic structure and that the group theory could shed light on the algebraic properties of those equations. This paper attempts a very preliminary study to see how far we can proceed with permutations of these coarse but sufficiently simple categories.

48 Wallace's (1889: 467) statement is cited in Chomsky (2007: 7). "The significance of such phenomena, however, is far from clear." (Chomsky 2009: 26, 33). See Chomsky & McGilvray (2012: 16) for relevant discussion.

49 Chomsky has stricter view than Kronecker in that C_{HL} is the origin of nutural 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."

50 This is a huge 'if'. An anonymous reviewer 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*. But at the same time, the reviewer states that "it is always worthwhile pointing out that every discrete structure in human language deserves group-theoretic analysis," and that "at least it must have value if it encourages future research in this direction."

(13) a. The cost hierarchy among the six geometrical operations that express the six unmarked word orders in C_{HL} is:

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

where *r0* corresponds to <*SOV*>, *f1* to <*SVO*>, *r1* to <*VSO*>, *f2* to <*OSV*>, *f3* to <*VOS*>, and *r2* to <*OVS*>. The geometrical cost approach predicts the current percentages of languages that have the top three word orders:

<*SOV*> (48.5%), <*SVO*> (38.7%), and <*VSO*> (9.2%).

b. 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*>).

c. As Lyle Jenkins suggests, the unmarked word order asymmetry is expressible as a group-theoretical factor (included in Chomsky's third factor): "word order types would be the (asymmetric) stable solutions of the symmetric still-to-be-discovered 'equations' governing word order distribution." The "symmetric equation" is a linear function (transformation) $f(x) = y$, where the mapping function f consists of various internal merge operations that are expressed as various symmetric transformations (rigid movements) of an equilateral triangle, the input x is the universal base *v*P that is expressed as the identity triangle, and y is the respective output tree that is expressed as the output triangle that preserves symmetry after transformation.

d. The gross computational cost is the same in all languages. The more energy the system uses for the basic word order formation, the less energy is left for other operations (the law of conservation of energy). Our model predicts that the <*SOV*> type is the most tolerant toward phonetically null entities and operations in which C_{HL} needs more energy to locate them and keep track of the result, the <*SVO*> type less tolerant, and the <*VSO*> type still less tolerant.

e. The unmarked ordering asymmetry obeys a physical law that has algebraic and geometrical expressions. Ordering is not accidental, contra Chomsky (2012).

References

Alexiadou, Artemis & Elena Anagnostopoulou. 1998. Parameterizing AGR: Word order, V-movement and EPP-checking. *Natural Language and Linguistic Theory* 16, 491–539.

Aldridge, Edith. 2002. Nominalization and WH-movement in Seediq and Tagalog. *Language and Linguistics* 3, 393–426.

Armstrong, Mark A. 1988. *Groups and Symmetry*. New York: Springer-Verlag.
Barrie, Michael J. M. 2006. Dynamic antisymmetry and the syntax of noun incorporation. Toronto: University of Toronto dissertation.
Boeckx, Cedric. 2009. The nature of Merge: Consequences for language, mind, and biology. In Massimo Piattelli-Palmarini, Pello Salaburu & Juan Uriagereka (eds.), *Of Minds and Language: A Dialogue with Noam Chomsky in the Basque Country*, 44–57. Oxford: Oxford University Press.
Carnie Andrew, Heidi Harley & Sheila Ann Dooley (eds.). 2005. *Verb First —On the Syntax of Verb Initial Languages*. Amsterdam: John Benjamins.
Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
Chomsky, Noam. 1982. *Noam Chomsky on the Generative Enterprise: A Discussion with Riny Huybregts and Henk van Riemsdijk*. Dordrecht: Foris. (Japanese translation by N. Fukui & M. Zushi. *Seesee bunpoo-no kuwadate* [generative enterprise]. 2011. Tokyo: Iwanami publication.)
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. 2002. Linguistics in the 21st century (An interview with Noam Chomsky). Interviewers: N. Fukui & M. Zushi, Nov. 22 at MIT. (Japanese translation by N. Fukui & M. Zushi. *Seesee bunpoo-no kuwadate* [Generative Enterprise]. 2011. Tokyo: Iwanami publication.)
Chomsky, Noam. 2004. Beyond explanatory adequacy. In Adriana Belletti (ed.), *Structures and Beyond*, vol. 3: *The Cartography of Syntactic Structures*, 104–131. Oxford: Oxford University Press.
Chomsky, Noam. 2005. Three factors in language design. *Linguistic Inquiry* 36, 1–22.
Chomsky, Noam. 2007. Approaching UG from below. In Uli Sauerland & Hans-Martin Gärtner (eds.) *Interfaces + Recursion = Language?*, 1–29. Berlin and New York: Mouton de Gruyter.
Chomsky, Noam. 2008. On phases. In Robert Freidin, Carlos P. Otero and Maria Luisa Zubizarreta (eds.), *Foundational Issues in Linguistic Theory. Essays in Honor of Jean-Roger Vergnaud*, 133–166. Cambridge, MA: MIT Press.
Chomsky, Noam. 2009. Opening remarks. In Massimo Piattelli-Palmarini, Juan Uriagereka & Pello Salaburu (eds.), *Of Minds & Language: A Dialogue with Noam Chomsky in the Basque Country*, 13–43. Oxford: Oxford University Press.
Chomsky, Noam. 2010. Some simple Evo-Devo theses: How true might they be for language? In Richard K. Larson, Viviane Déprez & Hiroko Yamakido (eds.), *The Evolution of Human Language: Biolinguistic Perspectives*, 45–62. Cambridge: Cambridge University Press.
Chomsky, Noam. 2012. On the poverty of the stimulus. A lecture at University College London. [http://www.youtube.com/watch?v=068Id3Grjp0 (22 November 2013)]
Chomsky, Noam & James McGilvray. 2012. *The Science of Language: Interviews with James McGilvray*. Cambridge: Cambridge University Press.

Di Sciullo, Anna Maria. 2005. *Asymmetry in Morphology*. Cambridge, MA: MIT Press.

Di Sciullo, Anna Maria & Daniela Isac. 2008. The asymmetry of merge. *Biolinguistics* 2, 260–290.

Di Sciullo, Anna Maria. 2013. Evo-devo language universals. A talk at *19th International Congress of Linguists (19ICL), Workshop 102: Advances in Biolinguistics*. 25-27 July 2013, University of Geneva, Geneva, Switzerland.

Dryer, Matthew S. & Martin Haspelmath (eds.). 2011. *The World Atlas of Language Structures Online*. Munich: Max Planck Digital Library. [http://wals.info/ (8 October 2011)]

Dryer, Matthew, S. 2012. Order of subject, object and verb. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Munich: Max Planck Digital Library. [http://wals.info/chapter/81 (20 August 2013)]

Fenk-Oczion, G. & A. Fenk. 2008. Complexity trade-offs between the subsystems of language. In M. Miestamo, K. Simemäki & F. Karisson (eds.), *Language Complexity—Typology, Contact, Change*, 43–65. Amsterdam: John Benjamins.

Fukui, Naoki. 1996. On the nature of economy in language. *Cognitive Studies* 3, 51–71.

Fukui, Naoki & Yuji Takano. 1998. Symmetry in syntax: Merge and Demerge. *Journal of East Asian Linguistics* 7, 27–86.

Fukui, Naoki. 2012a. Merge and (a)symmetry. A talk and handout at *Kyoto Conference on Biolinguistics*. March 12, 2012.

Fukui, Naoki. 2012b. *Shin shihzen kagaku toshiteno gengogaku: Seesee bunpoo towa nanika* [New edition, Linguistics as natural science: What is generative grammar]. Tokyo: Chikuma.

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, 17290–17295. [http://www.tv.com/web/ted-talks/watch/murray-gell-mann-do-all-languages-have-a-common-ancestor-1545550/ (22 November 2013)]

Gould, Stephen Jay. 2002. *The Structure of Evolutionary Theory*. Harvard, MA: Harvard University Press.

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, MA: 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 edn., 73–113. Cambridge, MA MIT Press.

Guimarães, Maximiliano. 2000. In defence of vacuous projections in bare phrase structure. *University of Maryland Working Papers in Linguistics* 9, 90–115.

Jenkins, Lyle. 2000. *Biolinguistics: Exploring the Biology of Language*. Cambridge: 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.

Jenkins, Lyle. 2013. Language typology and language universals. A talk at *19th International Congress of Linguists (19ICL), Workshop 102: Advances in*

Biolinguistics. 25-27 July 2013, University of Geneva, Geneva, Switzerland.
Jurka, Johannes. 2010. The importance of being a complement: CED effect revisited. College Park, MD: University of Maryland dissertation.
Kayne, Richard S. 1994. *The Antisymmetry of Syntax*. Cambridge, MA: MIT Press.
Kayne, Richard S. 2008. Antisymmetry and the lexicon. *Linguistic Variation Yearbook*, 8, 1–31.
Laughren, Mary. 1982. Warlpiri kinship structure. In J. Heath, F. Merlan & A. Rumsey (eds.), *Languages of Kinship in Aboriginal Australia*. *Oceania Linguistic Monographs* 24, 72–85. Sydney: University of Sydney.
Longa, Víctor M., Guillermo Lorenzo & Juan Uriagereka. 2011. Minimizing language evolution: The Minimalist Program and the evolutionary shaping of language. In Cedric Boeckx (ed.), *The Oxford Handbook of Linguistic Minimalism*, 595–616. Oxford: Oxford University Press.
Mallinson, Graham & Barry J. Blake. 1981. *Language Typology: Cross-Linguistic Studies in Syntax.* North-Holland Linguistic Series 46. Amsterdam: North-Holland Publishing Company.
Massam, Diane. 2000. VSO is VOS: Aspects of Niuean Word Order. In Andrew Carnie & Eithne Guilfoyle (eds.), *The Syntax of Verb Initial Languages*, 97–116. Oxford: Oxford University Press.
Matsumoto, Katsumi. 1992. Distributions and variations of word-order: A typological and areal study. *Kansai Linguistic Society* 12, 155–164.
Moro, Andrea. 2000. *Dynamic Antisymmetry*. Cambridge, MA: MIT Press.
Nakamura, Akira. 2010. *Garoa no Riron: Hooteeshiki wa Naze Tokenakattanoka* [Galois Theory: Why the Equation Could Not be Solved]. Tokyo: Kodansya.
Nematzadeh, S. 2013. Perspective on complexity. A talk at *19th International Congress of Linguists (19ICL), Workshop 121: Perspectives on complexity*. University of Geneva, Geneva. [25–27 July 2013]
Ruhlen, Merritt. 1975. *A Guide to the Languages of the World*. Stanford, CA: Language Universals Project, Stanford University.
Stewart, Ian. 1975. *Concepts of Modern Mathematics.* New York: Dover Publications.
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.
Strang, Gilbert. 2009. *Introduction to Linear Algebra*, 4th edn. Wellesley, MA: Wellesley-Cambridge Press. [http://ocw.mit.edu/courses/mathematics/18-06sc-linear-algebra-fall-2011/ (22 November 2013)]
Tomlin, Russel Schenk. 1986. *Basic Word Order: Functional Principles*. London: Croom Helm.
Ultan, Russell. 1969. Some general characteristics of interrogative systems. *Working Papers on Language Universals* 1, 41–63. Stanford, CA: Language Universals Project, Stanford University.
Uriagereka, Juan. 2010. The structure of the language faculty. A lecture at the Verbum Summer School of Linguistics. Mediateca, uvigo-tv, recursos multimedia da Universidade de Vigo. [http://tv.uvigo.es/video/30570.html (22 November 2013)]

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

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

Weil, André. 1969. On the algebraic study of certain types of marriage laws (Murngin system). In C. Lévi-Strauss (ed.), *The Elementary Structure of Kinship*, 221–229. Beacon Press. (Original version published as Sur l'étude algébrique de certains types de lois de marriage (Systèm Murngin). In C. Lévi-Strauss. 1947. *Les structures élémentaires de la parenté*. Paris: Mouton.)

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.

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

Koji Arikawa
St. Andrew's University (*Momoyama Gakuin University*)
Department of International Studies and Liberal Arts
1-1, Manabino, Izumi
Osaka
Japan
karikawa@andrew.ac.jp