Does Gaussian Elimination Teach About Uninterpretable Feature Elimination in CHL?

ARIKAWA Koji

Abstract

Are uninterpretable features (uF) similar to vanishing unknowns, such as x or y in linear algebra? Does Gaussian elimination (GE) teach us anything about uF elimination (uFE) in the computational procedures of human natural language (CHL)? It may teach us nothing. It may be pointless to compare GE to uFE. To answer the question, we compare GE with uFE and raise questions that have not been posed seriously and perform a toy experiment of linear algebraic uFE with $u\varphi$ and $uCase$ weighted by coefficients and constants, which express the ratio (symmetry degree) of $u\varphi$ to $uCase$ in the probe (P) and goal (G). The three types of solvability in GE parallel those in uFE. Namely, ① a unique solution (successful GE), ② no solution (failed GE), and ③ infinitely many solutions (failed GE) respectively correspond to ① complete AGREE, ② incomplete AGREE, and ③ internal merge. ③ is recycled in uFE, but not in GE. The experiment answers (or deepens) many uFE puzzles. We also investigate whether graph theory teaches us anything about P, G, and sentence structures.

1. Introduction

Alfred Russel Wallace (1823-1913), the co-founder of the theory of evolution, was puzzled: "the present gigantic development of the mathematical capacity is wholly unexplained by the theory of natural selection, and must be due to some altogether distinct cause" (Wallace 1889: 467). Leopold Kronecker (1823-1891), a German mathematician, stated that God made integers; all else is the work of man.[1] Noam Chomsky speculates that arithmetic derives from the computa-

1) "Die ganzen Zahlen hat der liebe Gott gemacht, alles andere ist Menschenwerk." Cited in Weber (1893). More naturally, nature made the human brain and natural numbers therein; all else is the work of man.

tional procedures of human natural language (C$_{HL}$) (Chomsky 2007: 7).[2]

The human design specifies that C$_{HL}$ is connected to the conceptual-intentional (thought) system (CI) and the sensorimotor system (SM). The CI reads and uses semantic information. The SM reads and uses phonetic information.[3] C$_{HL}$ builds binary-branching structures with uninterpretable (formal, structural) features (uF) that are neither meaning nor sound. Three types of uF exist: *uφ*, *uCase*, and edge feature (EF).[4] C$_{HL}$ deletes uF before it (uF) enters the CI or the SM (uF elimination; uFE). If a uF enters the CI or the SM, the CI or the SM freeze, i.e., they cannot compute uFs.

We compare uFE with Gaussian elimination (GE), the most elementary algorithm for solving simultaneous linear equations, e.g.,

$$（1）\quad \begin{cases} ax + by = c & \cdots\text{(a)} \\ dx + ey = f & \cdots\text{(b).}^{5)} \end{cases}$$

GE attempts to make the coefficient d zero, which eliminates the unknown x in (1b) using $0x = 0$. We compare uFE with GE and investigate whether a uF is similar to the unknown x.

Chomsky speculates that AGREE, internal merge (IM), and TRANSFER to SM are specific to C$_{HL}$; they are not built into an artificial language as the computational procedures of mathematics (C$_M$) (Chomsky 2000b).[6]

（2）a. [AGREE, IM, and TRANSFER to SM are] what appears to be "design flaws" that are not necessary for language-like systems (Chomsky 2000b: 117).

　　b. [T]wo "imperfections" uF and the dislocation property (in fact, morphology altogether)

2) Chomsky (2005: 3; 16, 2007, 2008: 139).

3) "For each language L (a state of FL [faculty of language]), the expressions generated by L must be "legible" [readable] to systems that access these objects at the interface between FL and external systems [CI and SM]－external to FL, internal to person" (Chomsky 2001a: 1). TRANSFER sends information to the CI and SM. C$_{HL}$ is primarily perfect for the CI, secondarily for the SM (Chomsky 2007; Epstein, Kitahara, and Seely 2010: 132-139).

4) Chomsky (2008). A special case of EF is EPP "extended projection principle: "I want to become a Spec, [TP]" (Bošković 2007). A complement bears occurrence (sister) feature (OCC): "I must be an occurrence [sister] of some ß" Chomsky (2001b: 11). These features drive structure growth.

5) GE is the most commonly used technique in computer programming to solve equations (Strang 2011). The idea of elimination appeared in the Chinese mathematical text *Chapter Eight Rectangular Arrays of The Nine Chapters on the Mathematical Art*, published approximately BCE 150 (Calinger 1999). In Europe, the method was given by Isaac Newton (English physicist and mathematician, 1642-1726), and a useful method was devised by Carl Friedrich Gauss (German mathematician and physicist, 1777-1855) (Grcar 2011).

6) Given a term X in a structure, external merge (EM) merges X with Z, which originates outside the structure. Internal merge (IM) merges X with Y, which originates inside the structure.

are never built into special-purpose symbolic systems (ibid. 442).

   c. [S]omething like Theta Theory [EM] is a property of any language-like system, while checking theory [AGREE, IM] is specific to human language, motivated (we are speculating) by legibility conditions [CI and SM pose on C_{HL}] (ibid. 452).

   d. [S]ymbolic systems designed for special purposes (mathematics, computers, etc.) dispense entirely with a phonological component [phonetic form: PF], not facing the need to meet the legibility conditions for human language at the sensorimotor interface [SM] (ibid. 438).

In contrast to Chomsky, we argue against (2): GE has operations similar to AGREE and IM. Some general questions arise.

(3) General questions

   a. Does C_M lack AGREE, IM, and TRANSFER to the SM?

   b. Is uFE distinct from GE?

   c. Are $u\varphi$ and $uCase$ unknowns, similar to x and y in C_M?

For (3a) and (3b), we answer no; for (3c), we answer yes.[7]

The remainder of this article is organized as follows. In Section 2, we raise questions regarding uFE, introduce GE, explain what coefficients and constants express, and show how uFE and GE seem similar. We perform a toy experiment of linear-algebraic uFE, where C_{HL} solves a series of a system of P and G equations with unknowns, such as $uCase$ and $u\varphi$, weighted by coefficients. We also investigate whether graph and network theories teach us anything about the dynamics of P and G, and the equilibrium (balance control) of a sentence structure. Section 3 confesses risks and hopes of our attempt. Section 4 concludes the paper and positions our approach in a wider frame.

2. Elimination in C_{HL}

2.1. Persistent look-ahead problem

Consider the following sentence.[8]

(4) Whom does he like?

---

7) The answer is consistent with the view that the instruction for a uF is "enter F without value" (Chomsky 2001b: 16).
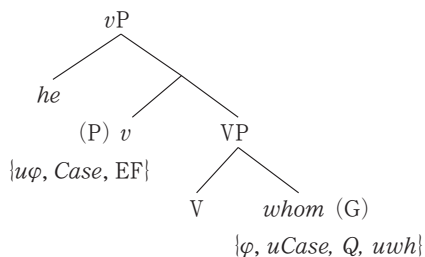8) The example is adapted from Epstein, Kitahara, and Seely (2010: 131).

vP

he

(P) v

$\{u\varphi, Case, \text{EF}\}$

VP

V　　*whom* (G)

$\{\varphi, uCase, Q, uwh\}$

Figure 1. Before AGREE and IM

vP (＝phase)

whom

$\{\varphi, \text{uCase}, Q, uwh\}$　*he*

v (P)

$\{\text{u}\varphi, Case, \text{EF}\}$

VP

V　　　*whom* (G)
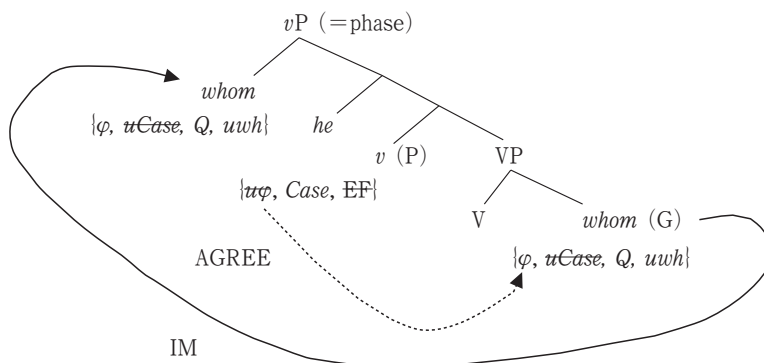
$\{\varphi, \text{uCase}, Q, uwh\}$

AGREE

IM

Figure 2. After AGREE and IM

Here, we focus on vP before feature checking *whom* (G) by v (P).[9]

P c-commands G. The wh-determiner phrase (wh-DP) *whom* bears $\{\varphi, uCase\}$ for A-movement and $\{Q, uwh\}$ for $\overline{\text{A}}$-movement (Chomsky 2000b: 128). P and G undergo AGREE (dotted arrow). G undergoes MOVE (solid arrow). We put aside move of V and *he* for expository purposes. Assume valuation-transfer (Chomsky 2000b: 122-123, 2001: 15-17, 2007, 2008, Bošković 2007: 593, 618-621). The algorithm is as follows.

（5）AGREE and IM algorithm in Figure 2

　　① $u\varphi$ in P finds the closest matching partner (complete-φ) in G.

　　② φ in G valuates $u\varphi$ in P as φ.

　　③ $u\varphi$ in P deletes.[10]

　　④ φ in P deletes because two identical φ's are redundant.

---

9) A probe (P) is a uF seeking the value. A goal (G) is an interpretable feature (iF) that valuates the P. More correctly, P and G refer to features. However, in the literature, P and G also refer to terms bearing P- and G-features. We use P and G both ways unless usage requires clarification.

10) The requirement "the closest" comes from the third factor (minimal search or minimal computation (MC)) (Chomsky 2005).

⑤ *Case* in P valuates *uCase* in G as [ACC].

⑥ *uCase* in G deletes.

Condition: a. If *uCase* of G is a reflex of φ in G, it deletes in situ.

b. If it is a reflex of $u\varphi$ in P, it deletes when G undergoes IM and become a *v*P-edge.

⑦ *Case* ([ACC]; inherent uF) in P and [ACC] (derivational uF) in G delete when φ and $u\varphi$ in P delete because *uCase* is a reflex of φ or $u\varphi$.

⑧ EF in P causes G internally merges with a projection of *v* (P) (IM). IM obeys the Phase Impenetrability Condition (PIC).[11]

⑨ EF deletes.

⑩ TRANSFER deletes uF in *v*P.

⑪ TRANSFER sends VP (*v*P-phase complement) to the CI and to the SM for semantic and phonetic interpretation, respectively.[12]

The following steps contain a look-ahead. ①: How does P know its valuator is G?[13] ⑤: How does G know its valuator is P? ⑧: How does C$_{HL}$ know *v* must bear EF to satisfy PIC? ⑩: How does TRANSFER know that undeleted uF will cause crash in SM and CI? Later steps contain a look-ahead. Finite T *s* in *does* bears $u\varphi$ (P) agreeing with the subject DP *he* (G). How does C$_{HL}$ know *s*'s partner is *he*? *He* values *s*'s $u\varphi$ as [3rd person, singular]. Is it true? Does finite T bear [NOM]? $u\varphi$-deletion values $u\varphi$ and *uCase*-deletion values *uCase*. The deletion is one fell swoop,

---

11) PIC: The domain of a head X of a phase XP is not accessible to operations outside XP; only X and its edge are accessible to such operations. See Chomsky (2000b: 108; 2001: 13). PIC foresees future: a look-ahead problem.

12) See Epstein, Kitahara, and Seely (2010) for problems with the analysis. TRANSFER recognizes uF ([-Int]) (*uCase*, $u\varphi$, EF, and phonological codes) as "CI-offending features that must be removed from the CI-bound object" (ibid.: 138). Assume the law of conservation, i.e., in narrow syntax (NS), features cannot be created or destroyed (ibid. 134). The valued (ex-unvalued) $u\varphi$ of *v* and valued (ex-unvalued) *uCase* of *whom* at *v*P-edge are "culprits" prosecuted for being [-Int] (ibid. 132). TRANSFER needs to see their "criminal record" to delete them, i.e., once unvalued. However, TRANSFER cannot erase valued $u\varphi$ and valued *uCase* because they are indistinguishable from inherent [+Int]. VP contains evidence of crime, i.e., AGREE. Evidence is lost by transferring VP. PIC is an accomplice.

Three way-outs are proposed (ibid. 132, 139-141): first, feature-splitting, i.e., lethal [-Int] remains inside the phase-complement, which will disappear (Obata and Epstein 2011), second, a perfect TRANSFER, i.e., CI-servant TRANSFER has a "clairvoyant power" reading the criminal record inside the lost phase complement (Epstein, Kitahara, and Seely 2010: 132-139; see fn. 45), and third, blind crash-proof CI. A look-ahead is persistent. How do TRANSFER and CI know that [-Int] will offend CI?

13) Chomsky attempts to avoid look-ahead. He states "the requirement that the matched probe delete" (Suicidal Greed), which "does not have the 'look-ahead' property of Greed, a complexity reduction that could be significant" (Chomsky (2000b: 127)). However, the matching contains a look-ahead. Bošković (2007: 618-621) observes a look-ahead. When DP becomes Spec, *v*P, DP behaves as if it knows that *v* will appear with EF to attract DP and delete *uCase*. Bošković discards EF and proposes that *uCase* knows that it must move to c-command the goal valuator *Case*, and respect PIC. How does *uCase* know its destiny?

Table 1: Three types of uF

	$u\varphi$	$uCase$	EF
Location	C, v^*	N	C, v^*
Optional	NO	NO	YES
Selectional[15]	NO	NO	YES
Activate AGREE	YES	YES	NO
Reflex of $u\varphi$ or φ	NO	YES	NO
Match with valuator	YES	NO	NO
Matching effect	YES	NO	NO
Valued	YES	YES	NO
Valuator	φ-Complete N	φ-Complete v^*, C	Non-existent
Intervention effect	YES	YES	NO
Deleted	YES	YES	YES
Deleter	φ-Complete N	φ-Complete v^*, C	φ-Complete v^*, C
IM required	NO	NO	YES

i.e., every relevant $u\varphi$ is deleted in all the copies simultaneously.[14] Table 1 summarizes the three uF types.

More specifically, we ask following questions.

（6）Specific questions

　　a. Can we avoid look-ahead problems in C$_{HL}$?

　　b. What does it means that $uCase$ is a reflex of φ or $u\varphi$ under AGREE? Is it redundant that $uCase$ is deleted in two ways?

　　c. Why does EF differ from $u\varphi$ and $uCase$? Why is it optional, selectional（Chomsky 2001a：40）, indifferent to AGREE, immune to matching/valuation/intervention effects（Chomsky 2005：18-19）, and indiscriminate：it can seek any goal in its domain（Chomsky 2008：151）, allowing free Merge to the edge, indefinitely（ibid. 629）? Why does it require IM, i.e., something occupy Spec-T［or any edge］（Chomsky（2000b：104））?

2.2. Elimination in C$_M$－How different are they?

2.2.1. Successful GE

To solve the system in（7）,（a）＋（b）appears a simple method as in（8）.

14）"We take deletion to be a 'one fell swoop' operation, dealing with the φ-set as a unit. Its features cannot selectively delete: either all or none"（Chomsky（2000b：124））. "Properties of the probe/selector α must be exhausted before new elements of the lexical subarray are accessed to drive further operation"（ibid. 462）. See also Chomsky（2001：15）. Chomsky（2000b：115-117）defines chain based on occurrence（sister）to assure one-fell-swoop deletion.

15）"［A］property … P is the EPP property of H［head］, commonly taken to be a selectional feature satisfied by merging K. K may be an expletive or a category determined by probe-goal agreement and pied-piping; the latter is the case of multiple Merge"（Chomsky 2001a：40）.

(7) $\begin{cases} x-2y=1 & \cdots(a) \\ 3x+2y=11 & \cdots(b) \end{cases}$

(8) $\begin{array}{l} \quad x-2y=1 \quad \cdots(a) \\ +)\ \underline{3x+2y=11} \quad \cdots(b) \\ \quad 4x+0y=12 \end{array}$

$-2y$ cancels $2y$. $0y$ eliminates y. This yields $x=3$. Back substitution derives $y=1$. What about distinct coefficients for y? What about five equations in five unknowns? Supercomputers use GE to solve millions of equations in millions of unknowns. GE is simple. To eliminate x: subtract a multiple of (a) from (b) (cf. Strang 2009: 45).

(9) $\begin{array}{l} \quad 3x+2y=11 \quad \cdots(b) \\ -)\ \underline{3x-6y=3} \quad \cdots(a)\times3 \\ \quad 0x+8y=8 \end{array}$

Here, $3x$ matches and cancels $3x$. The coefficient 0 eliminates x. The augmented matrix A_{aug} with coefficients and constants is as follows.

(10) $\begin{bmatrix} \boxed{1} & -2 & 1 \\ 3 & 2 & 11 \end{bmatrix} = A_{\mathrm{aug}}$

A_{aug} contains everything we need. GE and uFE share algorithmic flavor, which we notate by underlining. The purpose is to change 3 to 0, i.e., $3x$ to $0x$. GE equilibrates $\boxed{1}$ and 3. $\boxed{1}$ is the closest to 3 in the local area (the closest pair in the same column). To obtain 0 in the position of 3, match 3 and $\boxed{1}$, by multiplying $\boxed{1}$ with an appropriate multiplier (matchmaker M), in this case $M=3/1=$ ③.[16] Then, $\boxed{1}\times$③ and 3 match. We obtain $3=3$ or $3-3=0$, i.e., identity matching or maximized matching.[17] Multiply every entry in row one by $M=$③ and then column by column subtract the result from the entry in the second row. Row one is copied as it is.

(11) $\begin{bmatrix} \boxed{1} & -2 & 1 \\ 3-\boxed{1}\times③ & 2-(-2\times③) & 11-1\times③ \end{bmatrix}$

---

16) GE contains MATCH. In uFE, "the uninterpretable features of α [P] and β [G] render them *active*, so … matching leads to agreement" (Chomsky 2001a: 4). GE is activated by the difference between coefficients a and d.
17) "Maximize matching effects" (Chomsky 2001a: 15; (14)).

Every entry in row one is <u>modified</u>, <u>copied</u>, and <u>internally merged with</u> each original entry in row two. In $3-\boxed{1}\times\mathcal{3}$, $\boxed{1}$ reappears; $\boxed{1}$ is modified, in that it is multiplied by $\mathcal{3}$ and -1. The result is <u>internally merged with</u> the original entry 3, i.e., $3-\boxed{1}\times\mathcal{3}=3+(\boxed{1}\times\mathcal{3}\times(-1))$.[18] For coefficients, we obtain an upper triangular matrix U. 1, -2, and 8 form U.

(12) $\begin{bmatrix} \boxed{1} & -2 & 1 \\ 0 & \boxed{8} & 8 \end{bmatrix}$　U exists.

The new second row contains <u>new information generated by IM</u>. GE reveals that the original system hides the following.

(13) $\begin{cases} x-2y=1 & \cdots(a) \\ 8y=8 & \cdots(b) \end{cases}$

(13b) yields $y=1$. The solution is <u>transferred to CI for reasoning and SM for externalization</u>.[19] Back substituting $y=1$ in (13a) affords $x=3$. GE has AGREE, IM, and TRANSFER to SM. This answers general questions (3a) and (3b), reproduced in (14).

(14) Q: Does Cᴍ lack AGREE, IM, and TRANSFER to SM? (= 3a)
　　　A: No, it does not. They are built into Cᴍ.

　　　Q: Is uFE distinct from GE? (=3b)
　　　A: No, it is not. GE and uFE are translatable.

2.2.2. Failed GE

Elimination can fail. Consider the following (Strang 2009: 46-47).

(15) $\begin{cases} x-2y=1 & \cdots(a) \\ 3x-6y=11 & \cdots(b) \end{cases}$

---

18) An anonymous reviewer points out that merge is a set formation and is distinct from addition. Merge may have produced addition. See section 3.

19) An anonymous reviewer points out that it makes no sense to say that the value of $y=1$ is phonetically realized as [wʌn], i.e., English math would be different from Spanish math. We intended to show a simple example of externalization of solutions. Externalization can be done in various ways, including no sound. Solutions in Cᴍ can be externalized in some way.

Here GE fails to produce U.

$$(16) \quad \begin{bmatrix} \boxed{1} & -2 & 1 \\ 3 & -6 & 11 \end{bmatrix} = A_{\text{aug}} \quad \underset{\text{GE}}{\rightarrow} \quad \begin{bmatrix} \boxed{1} & -2 & 1 \\ 0 & 0 & 8 \end{bmatrix} \quad \text{No } U \text{ exists.}$$

The original system hides the following system.

$$(17) \quad \begin{cases} x - 2y = 1 & \cdots(a) \\ 0y = 8 & \cdots(b) \end{cases}$$

(17b) yields $0 = 8$, which is a contradiction. GE can fail in another way.

$$(18) \quad \begin{cases} x - 2y = 1 & \cdots(a) \\ 3x - 6y = 3 & \cdots(b) \end{cases}$$

GE fails to produce U.

$$(19) \quad \begin{bmatrix} \boxed{1} & -2 & 1 \\ 3 & -6 & 3 \end{bmatrix} = A_{\text{aug}} \quad \underset{\text{GE}}{\rightarrow} \quad \begin{bmatrix} \boxed{1} & -2 & 1 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{No } U \text{ exists.}$$

GE reveals the following system hidden in the original system.

$$(20) \quad \begin{cases} x - 2y = 1 & \cdots(a) \\ 0y = 0 & \cdots(b) \end{cases}$$

Any value satisfies y; infinite solutions, no unique solution.

2.3. A more general look at GE

We reproduce the system with its augmented matrix.

$$(21) \quad \begin{cases} ax + by = c & \cdots(a) \\ dx + ey = f & \cdots(b) \end{cases}$$

$$\begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} = A_{\text{aug}}$$

GE equilibrates the difference between a and d. Since $M=d/a$, $d-aM=d-a\cdot d/a=d-d$ $=0$. $d=d$. The matching is maximized. Perform the same operation for e and f to obtain U. 0 means $d-aM$.

(22)　$\begin{bmatrix} a & b & c \\ 0 & e-bM & f-cM \end{bmatrix}$　U exists.

The original system hides the following system. Here M brings it out.

(23)　$\begin{cases} ax+by=c & \cdots(a) \\ 0x+(e-bM)y=f-cM & \cdots(b) \end{cases}$

Zero modifies x. Since $0x=0$, x is eliminated, thereby valuating y.

(24)　$y=\dfrac{f-cM}{e-bM}=\dfrac{f-c\cdot d/a}{e-b\cdot d/a}=\dfrac{af-cd}{ae-bd}$

A unique solution requires the denominator and numerator to be nonzero.[20] Eliminating x valuates y. y-valuation is a reflex of x-elimination. GE does not delete every x. Back substitution valuates x.

(25)　$x=\dfrac{c-b\cdot\dfrac{f-cM}{e-bM}}{a}=\dfrac{c-b\cdot\dfrac{af-cd}{ae-bd}}{a}=\dfrac{ec-bf}{ae-bd}$

The value of x is a derivative of y.[21]

What is A_{aug} with no solution? GE reveals the following.

(26)　$\begin{bmatrix} a & b & c \\ 0 & 0 & f-cM \end{bmatrix}$　U does not exist.

---

20) The value of the denominator is called the determinant of A, expressed as det A or $|A|$. The dot product of the upper triangular matrix U is $|A|$. We require $|A|$ in eigenvalue problems, where symmetry is calculated.

21) The value of x is simplified as $(ec-bf)/(ae-bd)$. The numerator contains y's coefficients. The ratio of coefficients of y in the two equations matters to the value of x. The value of y is $(af-cd)/(ae-bd)$. The numerator contains x's coefficients. The ratio of the coefficients of x in the two equations matters to the value of y. In a sense, each value is a derivative of the other. See footnote 33.

We obtain $0y = f - cM$, where $f - cM \neq 0$. GE halts because no value satisfies y. Here, we double-check how GE fails.

$$(27) \quad y = \frac{f - cM}{e - bM} = \frac{f - c \cdot d/a}{e - b \cdot d/a} = \frac{af - cd}{ae - bd}$$

$e - bM = 0 = e - b \cdot d/a = 0$. Multiply by a, $ae - bd = 0$. The denominator is zero, which is not permitted, and the numerator is nonzero.[22]

What is A_{aug} with infinitely many solutions? GE reveals the following.

$$(28) \quad \begin{bmatrix} a & b & c \\ 0 & 0 & 0 \end{bmatrix} \quad U \text{ does not exist.}$$

We obtain $0y = 0$.[23] Any value satisfies y, i.e., infinitely many solutions. $e - bM = 0$ means $e - b \cdot d/a = 0$. Multiply by a, $ae - bd = 0$. The denominator is zero, which is disallowed. The numerator is zero.

2.4. What do coefficients and constants express?

Let us repeat A_{aug}.

$$(29) \quad \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} = A_{\mathrm{aug}}$$

Elimination changes A_{aug} to the following.

$$(30) \quad \begin{bmatrix} a & b & c \\ d - aM & e - bM & f - cM \end{bmatrix}$$

When GE yields a unique solution, $d - aM = 0$, $e - bM \neq 0$, and $f - cM \neq 0$. Assume $M = d/a$. Then $d - aM = d - a \cdot d/a = d - d = 0$. $0x$ eliminates x. $e - bM = e - b \cdot d/a \neq 0$. Multiply by a, $ae - bd \neq 0$. $ae \neq bd$, i.e., $a : b \neq d : e$. The ratio of the coefficients in one equation differs from that

---

22) Suppose $a/b = c$, where $a \neq 0$, $b = 0$, $c \neq 0$. Then $0c = a$, which means 0 equals nonzero a, i.e., a contradiction. The denominator cannot be zero (QED).

23) The equation $0y = 0$ hides a contradiction. Given $0y = 0$, any value satisfies this equation. On the other hand, $y = 0/0 = 0$, which means that the value of y is 0. The value of y can be anything and simultaneously it must be zero.

in the other. $f-cM=f-c\cdot d/a\neq0$. Multiply by a, $af-cd\neq0$. $af\neq cd$, i.e., $a:c\neq d:f$. The ratio of the coefficient and constant in one equation differs from that in the other. Here, the two equations differ; thus, symmetry is fully broken.

When GE yields no solution, $d-aM=0$, $e-bM=0$, and $f-cM\neq0$. $e-bM=0$ derives $a:b=d:e$. The two equations have the same ratio of coefficients. $f-cM\neq0$ derives $a:c\neq d:f$. The two equations have distinct ratio of a coefficient and constant. The two equations are identical and distinct, a contradiction; thus, symmetry is half broken.

When GE yields infinite solutions, $d-aM=0$, $e-bM=0$, and $f-cM=0$. $e-bM=0$ derives $a:b=d:e$. The two equations have identical ratio of coefficients. $f-cM=0$ derives $a:c=d:f$. The two equations have identical ratio of a coefficient and constant. The two equations are identical; symmetry is fully preserved. We summarize the correlation.

Table 2: Solvability, ratio difference, and symmetry degree

	$d-aM$	$e-bM$	$f-cM$	Solvability	Ratio of x to y	Symmetry
①	Zero	Nonzero	Nonzero	Unique solution	Different $a:b\neq d:e, a:c\neq d:f$	Fully broken
②	Zero	Zero	Nonzero	No solution	Equal & distinct $a:b=d:e, a:c\neq d:f$	Half broken
③	Zero	Zero	Zero	Infinitely many solutions	Equal $a:b=d:e, a:c=d:f$	Fully preserved

2.5. Geometry of three types of solvability

An equation with two unknowns is a line. Three solvability types exist.

(31) Three types of solvability with two lines:

　　① Two lines intersect. (A unique solution)

　　② Two lines are parallel. (No solution)

　　③ Two lines pile up. (Infinitely many solutions)

These are the only possibilities (Stewart 1995: 216). In ①, elimination of x is equivalent to tilting the line toward the x-axis until it becomes parallel to the x-axis, thus valuating y.
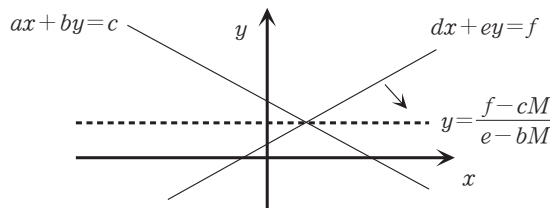

Figure 3: Successful GE (a unique solution)
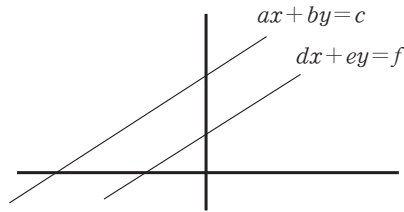
In ②, the parallel lines yield no intersection.

$$ax+by=c$$
$$dx+ey=f$$

Figure 4: Failed GE (no solution)

Elimination of x produces zero, which becomes the new coefficient of y.

(32) $$0y=\frac{f-cM}{e-bM} \qquad \left(\text{where} \frac{f-cM}{e-bM}\neq0\right)$$

Here, no value satisfies y, i.e., no solution. In ③, the two lines are identical. GE produces $0y=0$. Here, y is "free" (Strang 2009: 47). The two lines meet at infinite points (Strang 2003: 36–38).[24]
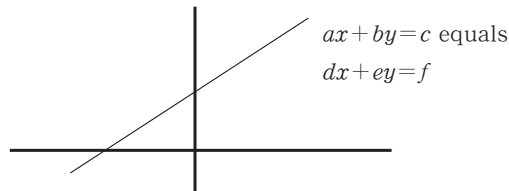
$$ax+by=c \text{ equals}$$
$$dx+ey=f$$

Figure 5: Failed GE (infinitely many solutions)

2.6. Linear algebraic uFE

2.6.1. If uF were unknown weighted by coefficient, …

Here, we perform linear algebraic uFE, a toy experiment to observe how far we can push the idea that a uF is an unknown bearing a coefficient.[25]

(33) Linear algebraic uFE

---

24) These row pictures suffice for two unknowns. For more unknowns, column pictures are more useful where column vectors play roles.

25) What about unifying uCase and $u\varphi$ as [-Int], i.e., $auF=b$ (Chomsky 2000b: 96; Epstein, Kitahara, and Seely 2010: 137)? Three types of solvability are available. However, we lose P and G distinction.
    For wh-movement, the system to solve is $auwh+buQ=c$ (G) and $duwh+euQ=f$ (P), where uwh corresponds to $uCase$ and uQ to $u\varphi$, (Chomsky (2000b: 128-129)). What about quadratic equation $ax^2+bx+c=d$, where the solution for (x_1, x_2) is that for $(u\varphi, uCase)$? Given $d=0$, the value of the discriminant D (b^2-4ac) distinguishes three types of roots (solutions), i.e., D>0: two distinct real-number solutions; D=0: one real-number solution; D<0: no real-number solution or two distinct complex-number solutions.

a. $u\varphi$ and $uCase$ are unknowns modified（weighted）by coefficients.

b. C_{HL} solves a series of a system of simultaneous linear equations that expresses a P and G interaction at each step.

$$\begin{cases} auCase + bu\varphi = c : \text{G equation（G attempts to assign value to P）} \\ duCase + eu\varphi = f : \text{P equation（P seeks value from G）} \end{cases}$$

c. As GE, uFE shows three solvabilities. uFE, not GE, recycles ③.

① Complete AGREE: successful uFE; ratio of $uCase$ to $u\varphi$ in P and G is different; a unique solution.

② Incomplete AGREE: failed uFE; ratio of $uCase$ to $u\varphi$ in P and G is equal and different: contradiction; no solution.

③ No AGREE; successful uFE; IM; ratio of $uCase$ to $u\varphi$ in P and G is equal; infinitely many solutions.

The coefficients and constants express the symmetry of uF in P and G. What are the contents of coefficients and constants?[26] A preliminary approximation is that coefficients are interpretable features（iF）.

Table 3: Substance of coefficients and constants in P and G equations

	Coefficient of $uCase$	Coefficient of $u\varphi$	Constant
G equation	$a = Case$-related iF in G, e.g., *inherent case*	$b = \varphi$-related iF in G, e.g., φ as [3rd person, singular]; θ [patient] assigned by V[27]	c interacts with a and b.
P equation	$d = Case$-related iF in P, e.g., *inherent case*; focus feature[28]	$e = \varphi$-related iF in P, e.g., finiteness of C; transitivity of v^*, θ [agent] assigned by v^*	f interacts with d and e.

Regarding a, we distinguish *Case* and *case*. *Case*（uF）is structural and lacks meaning, i.e., it does not contribute to semantic interpretation. The distinction between *him* bearing accusative Case [ACC] and *he* bearing nominative Case [NOM] does not yield difference in meaning.

(34) a. Mary killed *him*.　　　（*him*＝patient）

b. *He* was killed by Mary.　（*he*＝patient）

26）Assume that the field consists of real numbers.

27）An anonymous reviewer warns about oversimplification: "even if we grant that person, number, and gender are all conflated into a single φ-feature, this is a highly unrealistic assumption," e.g., for adjectives in German and Icelandic, the definiteness of D is responsible for the strong/weak distinction in AGREE. φ contains definiteness. Strong property reduces to IM（EF）, and weak property to AGREE. We consider it seriously that "a suggestive fact is that internal Merge requires just these three kinds of information [$u\varphi$, $uCase$, and EF]"（Chomsky 2001b: 16）.

28）For [FOCUS] triggering IM, see Watanabe（2005）and Miyagawa（2010）.

A preposition such as *by* assigns *inherent case* (iF) to nouns, i.e., *by* is like a predicate. Meanings of *by* are sensitive to nominal choice.

(35) a. He was killed *by Mary*.   (*Mary*＝agent)

　　　b. He was killed *by the house*.   (*the house*＝vicinity)

　　　c. He was killed *by three*.   (*three*＝the latest)

A set of φ-features [3rd person, singular] (iF) and θ [patient] (iF) of *him* in (34a) are examples of coefficient *b*. With respect to coefficient *d*, the following demonstrates that the focus feature in T (P) is relevant to IM of DP (G).

(36) This book, I asked Bill to read.   (Chomsky 1977)

Coefficient *e* includes transitivity and θ [agent] of v^* in (34a).

Graph-theoretically, a current flows from a high-potential node to a low-potential node (Strang 2009: 428). G φ-valuates P. Information flows from G to P. The potential of G must be greater than that of P, i.e., $b > e$ before AGREE. Why? To make TRANSFER delete uF and send phase-complements to the external systems, C and v^* respectively transmit information to T and V before TRANSFER (feature-inheritance; Richards 2007, Chomsky 2001b, 2007, 2008).[29] C and v^* lose information, i.e., $b > e$. P *Case*-valuates G. Information flows from P to G. It must be $d > a$ before *uCase* valuation. Why? *Case* is a reflex of φ and AGREE precedes *uCase* valuation. P gains iF (φ value) from G, i.e., $d > a$.

2.6.2. ① Successful uFE: Full asymmetry (P≠G)

Consider the following with the structure of complete AGREE.

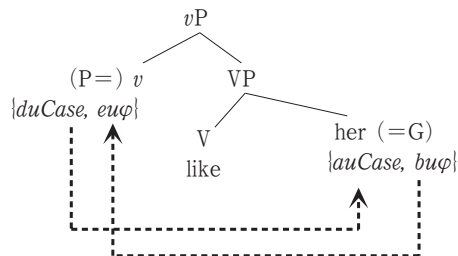(37) He likes her.

Figure 6: Complete AGREE for P seeking φ and for G seeking *Case*

---

29) Feature inheritance (Chomsky 2001b: 16) and TRANSFER contain look-ahead. How do they know uF in C causes a crash at CI unless T inherits the uF?

The system of equations to solve at this step is as follows.

(38)
$$\begin{cases} auCase+bu\varphi=c & \cdots(a): G \ (uCase \ \text{unvalued}; \varphi \ \text{complete}) \\ duCase+eu\varphi=f & \cdots(b): P \ (u\varphi \ \text{unvalued}; Case \ \text{complete}) \end{cases}$$

$uCase$ and $u\varphi$ bear coefficients. The augmented matrix is as follows.[30]

(39)
$$\begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}=A_{\text{aug}}$$

uFE attempts to balance a and d, changing A to U.

(40)
$$\begin{bmatrix} a & b & c \\ 0 & e-bM & f-cM \end{bmatrix}$$

The nonzero a, b, and $e-bM$ form U. 0 means $d-aM$, which is $d-a\cdot d/a$. This is $d-d=0$.[31] Matching is maximized by making d zero. However, here, matching is minimized; only one entry in row two is zero. Matching is minimized in U. uFE reveals the following system.

(41)
$$\begin{cases} auCase+bu\varphi=c & \cdots(a): G \ \text{equation} \\ 0uCase+(e-bM)u\varphi=f-cM & \cdots(b): \text{Hidden P equation} \end{cases}$$

From (41b), we obtain the value of $u\varphi$, which is a unique nonzero value.

(42)　$u\varphi=\dfrac{f-cM}{e-bM}=\dfrac{f-c\cdot d/a}{e-b\cdot d/a}=\dfrac{af-cd}{ae-bd}$

---

30) Provided that A is not singular, i.e., it yields a unique solution, A has the inverse $A^{-1}=\dfrac{1}{ae-bd}\begin{bmatrix} e & -b \\ -d & a \end{bmatrix}$, where the determinant (det) of $A=ae-bd\neq0$. A has two eigenvalues λ_1 and λ_2. Eigenvalues indicate how symmetry is preserved in A. The two eigenvalues are two solutions of the equation $\lambda_2+(a+e)\lambda+(ae-bd)=0$. The number $(a+e)$ is called the trace of A. When det $A=ae-bd=0$, one eigenvalue is 0, i.e., A is singular: it does not yield a unique solution. When $ae-bd\neq0$, the quadratic formula solves the equation: $\lambda=\dfrac{-(a+e)\pm\sqrt{(a+e)^2-4(ae-bd)}}{2}$. See the next section for more discussion.

31) The view is consistent with Chomsky (2001: 6): "the best case" of a match is identity rather than "nondistinctness" (Chomsky 2001a: 6; 2001b: 16).

The value only contains b, which is [3rd person, singular].

Importantly, $u\varphi \neq b$, i.e., $u\varphi$'s value is not G's φ: [3rd person, singular]. The value of $u\varphi$ is more complex; it contains all coefficients and constants.[32] To valuate $u\varphi$, what is matched is d (*Case*-related iF in P) and $a \cdot d/a$ (*Case*-related iF in G times (*Case*-related iF in P over *Case*-related iF in G)) to eliminate $uCase$. Since $d > a$, $d/a > 1$.

The unique nonzero solution needs $ae - bd \neq 0$ and $af - cd \neq 0$, i.e., $ae \neq bd$ and $af \neq cd$; thus, $a:d \neq b:e$ and $a:d \neq c:f$. Symmetry in P and G is fully broken. The $uCase$-deletion and $u\varphi$-valuation take place simultaneously. P does not foresee its value. Back substitution valuates $uCase$.

$$(43) \quad uCase = \frac{c - b \cdot \dfrac{f - cM}{e - bM}}{a} = \frac{c - b \cdot \dfrac{af - cd}{ae - bd}}{a} = \frac{ec - bf}{ae - bd}$$

$(ec - bf)/(ae - bd)$ is the value of [ACC]. The value of $uCase$ contains the value of $u\varphi$. The value of $uCase$ is a reflex (side effect) of $u\varphi$-valuation.[33] Now we have an answer to question (6a), which is repeated as (44).

(44) Q: Can we avoid look-ahead problems in C$_{HL}$? (=6a)

　　A: Yes, we can. A $u\varphi$ (P) does not know its value until uFE solves the system of equations. $uCase$ is valuated automatically.

The value of $uCase$ contains the value of $u\varphi$, which indicates that $uCase$ is a reflex of $u\varphi$. This answers question (6b), repeated as (45).

(45) Q: $uCase$ is a reflex of φ or $u\varphi$ under AGREE. What does it mean? Is it not redundant that $uCase$ is deleted in two ways? (=6b)

　　A: The value of $uCase$ contains the value of $u\varphi$. $uCase$ deletion and $u\varphi$ valuation occur simultaneously. $uCase$ is deleted in one way.

---

32) The value of $u\varphi = \dfrac{af - cd}{ae - bd}$ should contain "phi+transitive" that Case-values Accusative in the sense of Epstein, Kitahara, and Seely (2012).

33) The value of $uCase$ is simplified as $(ec - bf)/(ae - bd)$. The numerator contains $u\varphi$'s coefficients. The ratio of coefficients of $u\varphi$ in P and G matters to the value of $uCase$. The value of $u\varphi$ is $(af - cd)/(ae - bd)$. The numerator contains $uCase$'s coefficients. The ratio of coefficients of $uCase$ in P and G matters to the value of $u\varphi$. Each value is a reflex of the other. See footnote 21.

We answer question（3c）, which is repeated as（46）.

（46）Q：Are $u\varphi$ and $uCase$ unknowns, similar to x and y in GE?（＝3c）

　　　A：Yes, they are.

Let us look at the geometry.

$$auCase+bu\varphi=c \quad \text{(G-line)} \qquad u\varphi \qquad duCase+eu\varphi=f \quad \text{(P-line)}$$
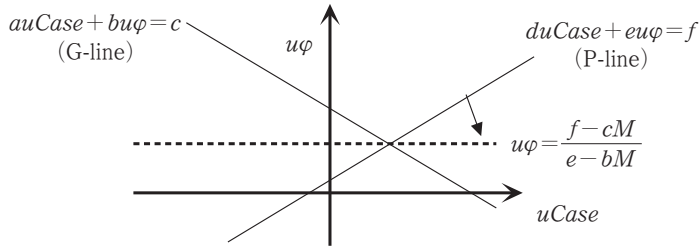
$$u\varphi=\frac{f-cM}{e-bM}$$

$$uCase$$

Figure 7：Complete AGREE（a unique solution）;
elimination of $uCase$ in P makes the P-line horizontal, thereby valuing $u\varphi$

The P- and G-line cross; thus, a unique solution exists.[34]

2.6.2.1. Linear algebraic uFE in simple numbers：complete AGREE

Let us tentatively create a P-G equation system for vP with specific simple numbers, such as 1 and 2, which is simply to aid presentation and must be revised for more elaborate feature calculi.[35] Assume $d>a$ and $b>e$ before AGREE（see Section 2.6.1）, and P and G, equally qualifying as agree-to-be terms, bear identical constant.

$$（47）\begin{cases} uCase+2u\varphi=1 & \cdots\text{(a)：G（}uCase\text{ unvalued; }\varphi\text{ complete）} \\ 2uCase+u\varphi=1 & \cdots\text{(b)：P（}u\varphi\text{ unvalued; }Case\text{ complete）} \end{cases}$$

A_{aug} is as follows. Notice that the coefficient matrix A is symmetric.[36]

34) An anonymous reviewer asks the following. What about multiple AGREE（Hiraiwa 2005）? In multiple AGREE, two parallel G-lines cross a P-line. Anaphoric pronominal PRO is a unique zero solution（0, 0）. PRO at the edge of infinitival TP bears the value（$uCase$, $u\varphi$）＝（NULL, NULL）, read as null Case and null φ（e.g., Chomsky and Lasnik（1995）and Martin（1999, 2001））.

35) We thank an anonymous reviewer who suggested a tentative application with specific numbers is necessary. It turns out that such simple numbers lead us to many far-reaching conclusions, given appropriate assumptions.

36) When each column of A adds to 1, e.g. $A=\begin{bmatrix} 0.3 & 0.7 \\ 0.7 & 0.3 \end{bmatrix}$, a symmetric Markov matrix, $\lambda=1$ is an eigenvalue, as in $Ax=x$, indicating steady state: the special vector（eigenvactor）x is "left unchanged"（Strang 2009: 285）. The output and the input are identical. When $\lambda=1$, the eigenvector is $x=(1, 1)$.

(48) $\begin{bmatrix} \boxed{1} & 2 & 1 \\ 2 & 1 & 1 \end{bmatrix} = A_{\mathrm{aug}}$

An intermediate step of GE is as follows. $2/1$ is the matchmaker (multiplier) M that equilibrates $\boxed{1}$ and 2 that is under $\boxed{1}$.

(49) $\begin{bmatrix} \boxed{1} & 2 & 1 \\ 2-\boxed{1}\times 2/1 & 1-(2\times 2/1) & 1-1\times 2/1 \end{bmatrix}$

GE generates U as follows.

(50) $\begin{bmatrix} \boxed{1} & 2 & 1 \\ 0 & \boxed{-3} & -1 \end{bmatrix} U$ exists.

GE reveals that the original system hides the following.

(51) $\begin{cases} uCase+2u\varphi=1 & \cdots(a): \text{G equation} \\ 0uCase-3u\varphi=-1 & \cdots(b): \text{Hidden P equation} \end{cases}$

Equation (51b) yields the value of $u\varphi$.

(52) $u\varphi=1/3$

Back substitution yields the value of $uCase$.

(53) $uCase=1/3$

We obtain $(u\varphi, uCase)=(1/3, 1/3)$. The value of $u\varphi$ and the value of $uCase$ are identical, i.e.,

---

A^{100} will approach $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. This is P=G and IM. IM is a result of powers of Markov matrix within a particular probe-goal calculation. If A were a symmetric Markov matrix, the symmetry pattern of A hides IM. Chomsky (1957: 18-21) and Chomsky and Miller (1958) refuted a claim that C_{HL} contains "finite state Markov processes" expressed as state diagrams with assignment of "a probability to each transition from state to state" (Shannon and Weaver 1949). According to Chomsky, such "a finite state grammar … runs into serious difficulties and complications at the very outset" and that "it is necessary to define the syntactic properties" (Chomsky 1957: 20-21). It is impossible to assign a probability to how one starts a sentence. However, if IM is a fundamental syntactic property of C_{HL} and a good probe-goal Markov matrix hides IM, we need to shed a new light to Markov matrix. See the next footnote for more discussion.

they literally（arithmetically）realize "complete AGREE" between the two values, not that be-
tween P and G.[37]

2.6.3. ② **Failed uFE: half symmetry（P＝G and P≠G; contradiction）**

Consider an ungrammatical example with the relevant structure.

（54）　＊He likes she.

vP

（P＝）v　　　VP
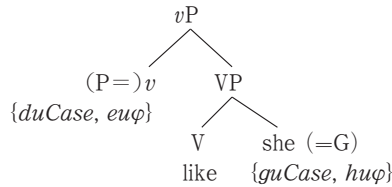{*duCase, euφ*}

　　　　V　　　she（＝G）
　　　like　　{*guCase, huφ*}

Figure 8: Incomplete AGREE: P and G fail to be valued

No valuation occurs. The system to solve is as follows.

（55）　$\begin{cases} guCase + hu\varphi = i & \cdots(a): G\ (uCase\ \text{unvalued};\ \varphi\ \text{complete}) \\ duCase + eu\varphi = f & \cdots(b): P\ (u\varphi\ \text{unvalued};\ Case\ \text{complete}) \end{cases}$

The augmented matrix A_{aug} is as follows.

（56）　$\begin{bmatrix} g & h & i \\ d & e & f \end{bmatrix} = A_{\text{aug}}$

---

37) Certain exceptional vectors x (eigenvectors; input) are in the same direction as Ax (output). Eigenvalues and eigenvectors reveal the symmetry properties of a matrix. What are the eigenvalues and eigenvectors of a symmetric non-Markov matrix as $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ yielding complete AGREE? The basic equation is $Ax = \lambda x$, which can be rewritten as $(A - \lambda I)x = 0$. We use the fact that $A - \lambda I$ is not invertible if $(A - \lambda I)x = 0$ yields a nonzero solution. In that case, the determinant of $A - \lambda I$ must be zero. Computing the determinant of $A - \lambda I$, we obtain $(1 - \lambda)^2 - 4 = 0$, which is $\lambda^2 - 2\lambda - 3 = 0$. Factoring, we obtain $(\lambda - 3)(\lambda + 1) = 0$. $\lambda = 3$ or $\lambda = -1$. Solve $(A - \lambda I)x = 0$ to find eigenvectors. $\lambda = 3$ reveals that A preserves symmetry when x is stretched by a factor of 3 and the eigenvector is $x = (1, 1)$. $\lambda = -1$ reveals that A preserves symmetry when x is reversed and the eigenvector is $x = (1, -1)$. The two eigenvectors are orthogonal. A symmetric matrix has orthogonal eigenvectors. Permutation, projection, and reflection matrices have the same eigenvectors. The former exchanges rows, and the latter yields the closest solution with minimal error. See Strang (2009: 283-288) for an introduction to eigenvalues and eigenvectors.
What do $\lambda = 3$ and $\lambda = -1$ indicate? When A becomes A^{100}, $\lambda = 3$ becomes $\lambda^{100} = 3^{100} = 515377520732011331036461129765621272702107522001$, an exponetial growth that seems to be "infinity" to the human eyes. When A becomes A^{100}, $\lambda = -1$ goes digital, i.e., $\lambda^1 = -1$, $\lambda^2 = 1$, $\lambda^3 = -1$, $\lambda^4 = 1$, ..., $\lambda^{99} = -1$, $\lambda^{100} = 1$. The two eigenvalues indicate that C$_{\text{HL}}$ demonstrates "the property of discrete infinity," which is among "unexpected features of complex biological systems, more like what one expects to find （for unexplained reasons）in the study of the inorganic world" (Chomsky 1995: 154).
If a complete AGREE involves a non-Markov matrix as above, our analysis supports Chomsky's （1957） argument against C$_{\text{HL}}$ as finite state Markov processes.

GE changes A_{aug} to the following matrix, where lower left $0 = d - gM$, lower middle $0 = e - hM$ $= 0$, and $f - iM \neq 0$. $M = d/g$.

(57) $\begin{bmatrix} g & h & i \\ 0 & 0 & f - iM \end{bmatrix}$

Two entries in row two are zeros (medium matching). The original system hides the following, where $f - iM \neq 0$.

(58) $\begin{cases} guCase + hu\varphi = i & \cdots\text{(a)}: \text{G equation} \\ 0u\varphi = f - iM & \cdots\text{(b)}: \text{Hidden P equation} \end{cases}$

No $u\varphi$ satisfies (58b): uFE yields no solution. $g:h = d:e$, i.e., P and G have the same ratio of co-efficients (G=P), and $g:i \neq d:f$, i.e., P and G have distinct ratio of coefficients and constants (G≠P): a contradiction. Crash is defined as no solution in NS.[38] Let us observe the geometry. No intersection exists. No solution exists.
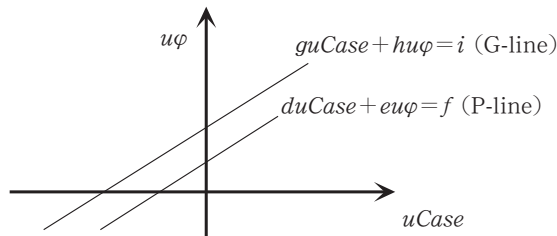

Figure 9: Incomplete AGREE: G- and P-line are parallel (no solution)

2.6.3.1. Linear algebraic uFE in simple numbers: incomplete AGREE

Suppose that the system to solve was as follows.

---

38) Linear algebraic uFE is compatible with two types of crash, i.e., compositional crash (e.g., valued $u\varphi$ [3rd person, singular] and PAST Tense cause crash at CI) vs. single-feature crash (e.g., unvalued $uCase$ alone causes crash at CI). The uFE conforms with compositional crash, i.e., crash means no solution in NS due to compositional contradiction. The uFE is compatible with single-feature crash, i.e., no $u\varphi$ satisfies $0u\varphi = f - iM$, where $f - iM$ is nonzero. A single unvalued $u\varphi$ causes crash. Epstein, Kitahara, and Seely (2010) argue for single-feature crash. Crash occurs in NS, which avoids look-ahead relative to the perspective that "the transfer operation "knows" that the feature that has been valued is uninterpretable and has to be erased at (or before) CI" (Chomsky 2007: 19). See Frampton and Gutmann (2002) and Putnum (2010) for crash-proof syntax. Epstein, Kitahara, and Seely (2014: 464) refute: "Merge is free: it optionally applies (and [hence,] crashing happens; [i.e.,] the system is not crash-proof)."

(59) $\begin{cases} 2uCase + u\varphi = 2 & \cdots(a): G \ (uCase \ \text{unvalued}; \ \varphi \ \text{complete}) \\ 2uCase + u\varphi = 1 & \cdots(b): P \ (u\varphi \ \text{unvalued}; \ Case \ \text{complete}) \end{cases}$

Only G equation has changed, i.e., 2 modifies $uCase$, 1 modifies $u\varphi$, and constant is 2. In G and P, $uCase$ coefficient is identical and $u\varphi$ coefficient is the same. The constants differ in G and P. A_{aug} is as follows, which is antisymmetric.

(60) $\begin{bmatrix} \boxed{2} & 1 & 2 \\ 2 & 1 & 1 \end{bmatrix} = A_{\text{aug}}$

$\boxed{2}/2$ is M that equilibrates $\boxed{2}$ and 2 below.

(61) $\begin{bmatrix} \boxed{2} & 1 & 2 \\ 2 - \boxed{2} \times \boxed{2}/2 & 1 - (1 \times \boxed{2}/2) & 1 - 2 \times \boxed{2}/2 \end{bmatrix}$

GE generates U as follows.

(62) $\begin{bmatrix} \boxed{2} & 1 & 2 \\ 0 & \boxed{0} & -1 \end{bmatrix} U$ exists.

GE reveals that the original system hides the following.

(63) $\begin{cases} 2uCase + u\varphi = 2 & \cdots(a): \text{G equation} \\ 0uCase + 0u\varphi = -1 & \cdots(b): \text{Hidden P equation} \end{cases}$

(63b) is $0u\varphi = -1$. No value is possible for $u\varphi$. With proportionate coefficients and disproportionate constants, GE breaks down.[39]

2.6.4. ③ Internal merge: full symmetry (P＝G)

Consider the following example.

---

39) What are the eigenvalues and eigenvectors of A yielding incomplete AGREE? Computing the determinant of $A - \lambda I$, we obtain $(2-\lambda)(1-\lambda) - 2 = 0$, i.e., $\lambda^2 - 3\lambda = 0$. Factoring, we obtain $\lambda(\lambda-3)=0$, i.e., $\lambda=0$ or $\lambda=3$. When A is unsolvable, $\lambda=0$ is an eigenvalue. Solve $(A-\lambda I)x=0$ to find eigenvectors. $\lambda=0$, i.e., A preserves symmetry when x is zero. The eigenvector is $x=(1, -2)$. $\lambda=3$, i.e., A preserves symmetry when x is stretched by a factor of 3. The eigenvector is $x=(1, 1)$. The two eigenvectors are not orthogonal.

(64) Whom does he like?

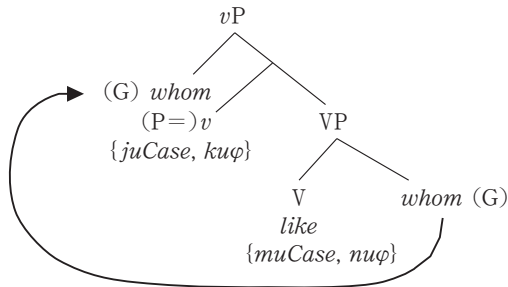After AGREE, the DP *whom* internally merges with a projection of *v*.[40]

Figure 10: IM and no AGREE

EF in P causes IM. No AGREE exists. EF cannot save AGREE failure.

(65) * Whose did he like?

Consider the equation system for the grammatical example as follows.

(66) $\begin{cases} muCase + nu\varphi = o & \cdots(a): \text{G }(uCase \text{ valued}) \\ juCase + ku\varphi = l & \cdots(b): \text{P }(u\varphi \text{ valued}) \end{cases}$

The augmented matrix A_{aug} is as follows.

(67) $\begin{bmatrix} m & n & o \\ j & k & l \end{bmatrix} = A_{\text{aug}}$

GE yields a zero row, i.e., $j - mM = k - nM = l - oM = 0$, where $M = j/m$.

(68) $\begin{bmatrix} m & n & o \\ 0 & 0 & 0 \end{bmatrix}$

Matching is maximized. GE reveals the following hidden system.

---

40) G moves to *v*P-edge to form an operator-variable structure. How does G know G must form such a structure? Another look-ahead.

(69) $\begin{cases} muCase + nu\varphi = 0 \quad \cdots(a)\text{: G equation} \\ 0u\varphi = 0 \quad \cdots(b)\text{: Hidden P equation} \end{cases}$

Any value of $u\varphi$ satisfies (69b). $m:n=j:k$, i.e., P and G have equal ratio of coefficients. $m:o=j:l$, i.e., P and G have the same ratio of the coefficients and constants. G＝P; thus, symmetry is preserved. uFE recycles infinite solutions as IM. IM realizes maximized matching. P＝G causes G internally merging with P. No further matching is necessary. EF yields infinitely many solutions; a free gift from elimination failure. This answers question (6c), repeated as (70).

(70) Q: Why does the EF differ from $u\varphi$ and $uCase$? (＝6c)

　　　A: IM realizes maximized matching: G＝P. EF is dispensable. EF-effect is an apparent side effect due to elimination failure.[41]

EF is either redundant or makes incorrect predictions (Epstein and Seely (2006: 52–53).[42] One equation with one unknown yields EF ($0x=0$) but it cannot differentiate P and G. Two equations with two unknowns yield EF-effect, differentiating P and G. Nature has chosen two unknowns, which are minimally sufficient. Consider the geometry.
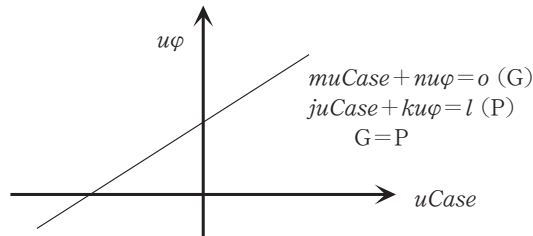

$u\varphi$

$muCase + nu\varphi = o$ (G)
$juCase + ku\varphi = l$ (P)
G＝P

$uCase$

Figure 11: IM; G-line and P-line pile up

Infinite intersections exist; thus, infinitely many solutions exist.[43]

---

41) EPP-effect is accounted for without an EPP-feature. Counterexamples against EPP-less analysis are not problematic. Epstein and Seely (2006: 114-115) list such counterexamples of Condition A, Q-float, and re-construction.

　(ⅰ) * Bill₂ appears to Mary₁ [t₂ to seem to herself₁ to like physics].
　(ⅱ) The students₁ seem t₁ all to know French.
　(ⅲ) a. * [His₁ mother₂'s bread]₃ seems to her₂ t₃ to be known by every man₁ to be the best.
　　　b. [His₁ mother₂'s bread]₃ seems to every man₁ t₃ to be known by her₂ to be the best.

42) For arguments against EF, see Jacobson (1992), McCloskey (1996), Castillo, Drury and Grohmann (1999), Martin (1999), Manzini and Roussou (2000), Bošković (2002), and Chomsky (2001), as cited in Epstein and Seely (2006: 53).

43) Let ＞indicate c-command and G₁ inactive. The intervention effect * P＞G₁＞G₂ is expressed as P- and G₁-line intersecting while P- and G₂-line are parallel. Other examples of infinitely many solutions are defective G/P in raising, ECM, passive, unaccusative, participles lacking [person] (Chomsky 2001b: 12),

2.6.4.1. Linear algebraic uFE in simple numbers: IM

Suppose that the system to solve was as follows.

(71) $\begin{cases} 2uCase+u\varphi=1 & \cdots(a):G\ (uCase\ \text{valued};\ \varphi\ \text{complete}) \\ 2uCase+u\varphi=1 & \cdots(b):P\ (u\varphi\ \text{valued};\ Case\ \text{complete}) \end{cases}$

In G, the coefficients remain the same and the constant returns to 1. In fact, P=G. A_{aug} is as follows, which is antisymmetric.

(72) $\begin{bmatrix} \boxed{2} & 1 & 1 \\ 2 & 1 & 1 \end{bmatrix}=A_{\text{aug}}$

$\boxed{2}/2$ is the matchmaker M that equilibrates $\boxed{2}$ and 2 below.

(73) $\begin{bmatrix} \boxed{2} & 1 & 1 \\ 2-\boxed{2}\times\boxed{2}/2 & 1-(1\times\boxed{2}/2) & 1-1\times\boxed{2}/2 \end{bmatrix}$

GE generates U as follows.

(74) $\begin{bmatrix} \boxed{2} & 1 & 1 \\ 0 & \boxed{0} & 0 \end{bmatrix} U$ exists.

GE reveals that the original system hides the following.

(75) $\begin{cases} 2uCase+u\varphi=1 & \cdots(a):G\ \text{equation} \\ 0uCase+0u\varphi=0 & \cdots(b):\text{Hidden P equation} \end{cases}$

(75b) is $0u\varphi=0$. Any value satisfies $u\varphi$; thus, there are infinite solutions. P=G causes IM.[44]

2.6.4.2. IM as loop formation

Infinite solvability is statistically rare. It shows full symmetry and a fractal structure where infor-

---

Quirky Dative G (ibid: 557), and successive cyclic move.

44) The eigenvalues and eigenvectors of A yielding IM are the same as those of A yielding incomplete AGREE. Both A's preserve symmetry when the output is zero or stretched by a factor of 3. Eigenvalues cannot distinguish incomplete AGREE and IM. Unlike incomplete AGREE (no solution), IM (infinite solutions) employs $\lambda=3$ to save elimination breakdown.

mation is optimally packed and error is minimized.[45] EF creates loops and fractal structures.
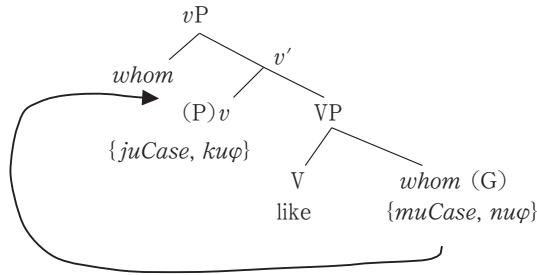
Figure 12: IM creates loop

C_HL employs IM to yield loops. Here, the loop is {*whom, whom, v*P, *v′*, VP}. A loop current is a solution to the balance law (Strang 2009: 420-428). Loops balance structures. IM is an optimal solution to the legibility problem, i.e., minimize error. SMT is reinforced. C_HL recycles EF as the engine for structural growth. C_HL is a flexible tinkerer.[46]

2.6.5. P c-commands G: How does graph theory describe it?

An anonymous reviewer asks: What about P c-commanding G? How does a matrix express a structural relation as *c-command*? We need another matrix for dynamics of P and G. This section demonstrates an application of GE and graph theory to calculate inherent balance force in sentence structures. Figure 2 is reproduced.
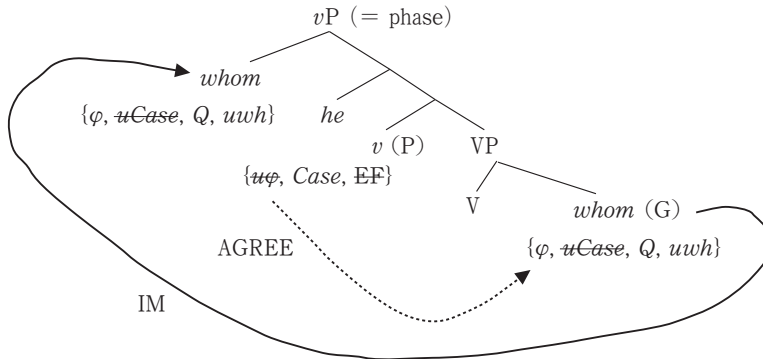
Figure 13. After AGREE and IM (＝Figure 2)

---

45) "[T]he [fractal] shapes ... tend to be *scaling*, implying that the degree of their irregularity and/or fragmentation is identical at all scales. The concept of *fractal ... dimension* plays a central role ... (Mandelbrot 1977: 1).

46) Evolution is a "tinkerer, not an engineer" (Jacob 1977). The mathematical system (C_M) contains symmetry-calculating equations such as $Ax = \lambda x$, which investigate how the input x and the output x (eigenvectors) preserve symmetry in, say, A^{100}, and $A = S\Lambda S^{-1}$ to calculate A^{100} in a fast and easy way, where the matrix A turns into a diagonal eigenvalue matrix Λ when we use eigenvector matrix S properly. As C_HL, C_M recycles infinitely many solutions to observe how a system of equations preserves symmetry.

Number nodes and edges (features are not nodes) in an upward directed graph, and assign -1 to a starting node, and 1 to an end node. vP is translatable into the incidence matrix A.

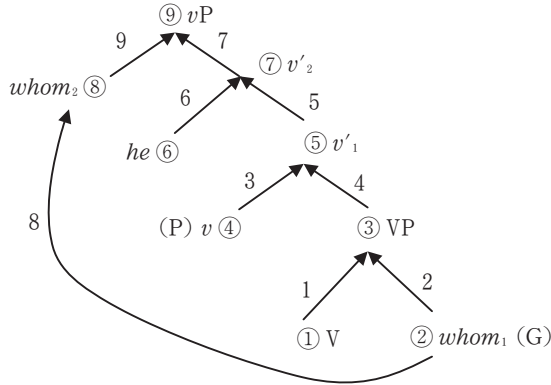Figure 14: Upward directed graph of vP

Table 4: Incidence matrix A

	① V	② wh_1 (G)	③ VP	④ v(P)	⑤ v'_1	⑥ he	⑦ v'_2	⑧ wh_2	⑨ vP
1	-1	0	1	0	0	0	0	0	0
2	0	-1	1	0	0	0	0	0	0
3	0	0	0	-1	1	0	0	0	0
4	0	0	-1	0	1	0	0	0	0
5	0	0	0	0	-1	0	1	0	0
6	0	0	0	0	0	-1	1	0	0
7	0	0	0	0	0	0	-1	0	1
8	0	-1	0	0	0	0	0	1	0
9	0	0	0	0	0	0	0	-1	1

The column vector of G is $(0, -1, 0, 0, 0, 0, 0, -1, 0)$ and that of P is $(0, 0, -1, 0, 0, 0, 0, 0, 0)$ in 9-dimensional real-number space $\mathbf{R}^9$. GE reveals the true size of the column space, i.e., 8 dim, not 9. GE yields U as follows.[47]

Table 5: Upper triangular matrix U of A

	① V	② wh_1 (G)	③ VP	④ v(P)	⑤ v'_1	⑥ he	⑦ v'_2	⑧ wh_2	⑨ vP
1	-1	0	1	0	0	0	0	0	0
2	0	-1	1	0	0	0	0	0	0
3	0	0	-1	0	1	0	0	0	0
4	0	0	0	-1	1	0	0	0	0
5	0	0	0	0	-1	0	1	0	0
6	0	0	0	0	0	-1	1	0	0
7	0	0	0	0	0	0	-1	0	1
8	0	0	0	0	0	0	0	1	-1
9	0	0	0	0	0	0	0	0	0

47) We use the Reshish matrix calculator (reshish.matrix.com). The true size of a matrix is called the rank r. For A, $r=8$. The computation time was 0.012 sec.

GE reveals that one edge is redundant to form an economical tree.[48]

Transpose A to obtain A^T, i.e., exchange rows and columns. It is "the beauty of the framework, that A^T appears along with A (Strang 2009: 412)," thereby hiding information about equilibrium at each node.

Table 6: Transpose matrix A^T

	1	2	3	4	5	6	7	8	9
① V	-1	0	0	0	0	0	0	0	0
② wh_1 (G)	0	-1	0	0	0	0	0	-1	0
③ VP	1	1	0	-1	0	0	0	0	0
④ v (P)	0	0	-1	0	0	0	0	0	0
⑤ v'_1	0	0	1	1	-1	0	0	0	0
⑥ he	0	0	0	0	0	-1	0	0	0
⑦ v'_2	0	0	0	0	1	1	-1	0	0
⑧ wh_2	0	0	0	0	0	0	0	1	-1
⑨ vP	0	0	0	0	0	0	1	0	1

The balance equation is $A^T y = 0$.[49] Let y_n be the current on edge n. Row G yields $-y_2 - y_8 = 0$, i.e., current entering node ② through edge 2 and 8 is zero.[50] The equation is $-y_2 = y_8$, i.e., current entering ② equals the current exiting ②. Flow in equals flow out; there is no accumulation (traffic jam) at nodes. Row P yields $-y_3 = 0$, i.e., no current enters ④.

It is "the beauty of the framework" that the graph Laplacian matrix $A^T A$ hides information about the equilibrium of the system.[51] The potentials and currents are calculable in the balanced system. $A^T A$ is as follows.

---

48) A tree is a graph without loops. The dot product of diagonal is the determinant (det) of A. Det $A = 0$. The inverse A^{-1} does not exist: A is not invertible. A is singular, i.e., it fails to yield a unique solution because infinitely many constant vectors $(c, c, c, ..., c)$ solves $Ax = 0$.

49) Kirchhoff's current law (KCL): Flow in equals flow out at each node (Strang 2009: 425). This law deserves first place among the equations of applied mathematics. It expresses "conservation" and "continuity" and "balance." Nothing is lost, nothing is gained (ibid). $A^T y = 0$ when the system is closed. $A^T y = f$ when the system is open, where f is an external power source. A loop is a solution to KCL. Here, $y = (0, 1, 0, 1, 1, 0, 1, -1, -1)$ is a solution to KCL. C_{HL} has IM because IM creates loops and loops equilibrate structures.

50) The minus sign indicates opposite direction.

51) $A^T A$ comes from KCL and Ohm's law: $y = -CAx$, i.e., current y along an edge = conductance c times potential difference Ax (Strang 2009: 426). C measures how easily flow gets through. Substitute $y = -CAx$ for y in KCL. $A^T \cdot -CAx = 0$ yields. Let $C = 1$ and multiply both sides by -1. We obtain $A^T Ax = 0$.

Table 7: Graph Laplacian matrix $A^{\mathrm{T}}A$

	① V	② wh_1 (G)	③ VP	④ v (P)	⑤ v'_1	⑥ he	⑦ v'_2	⑧ wh_2	⑨ vP
① V	1	0	−1	0	0	0	0	0	0
② wh_1 (G)	0	2	−1	0	0	0	0	−1	0
③ VP	−1	−1	3	0	−1	0	0	0	0
④ v (P)	0	0	0	1	−1	0	0	0	0
⑤ v'_1	0	0	−1	−1	3	0	−1	0	0
⑥ he	0	0	0	0	0	1	−1	0	0
⑦ v'_2	0	0	0	0	−1	−1	3	0	−1
⑧ wh_2	0	−1	0	0	0	0	0	2	−1
⑨ vP	0	0	0	0	0	0	−1	−1	2

$A^{\mathrm{T}}A$ is symmetric, sparse, and banded.[52] The diagonal entry of column G and row G is 2, i.e., two edges meet G. The other numbers express whether the node connects with G; −1 if connected, 0 if disconnected.

$A^{\mathrm{T}}A$ is not solvable.[53] To make $A^{\mathrm{T}}A$ solvable, we ground a node. Imagine that $A^{\mathrm{T}}A$ is a system of springs (edges) and masses (nodes) lying on a table. Gravity does not reveal the inherent balance force. We hang it to the ceiling at a node. Gravity shakes the system and reveals the inherent balance. Grounding a node corresponds to fixing the node to the ceiling or fixing absolute zero (0K). The potential of the fixed node is zero. Let us ground ② (wh_1 (G)) because it is unpronounced. The quiet graph (Figure 14) becomes a busy network (Figure 15), where inherent balance controls the entire system. The network is hung upside-down to the ceiling at ②. The ceiling (support) holds and pulls the system against gravity at ②. S expresses the external source power (reaction force) caused by gravity. Note that S is not an edge.

---

52) Symmetric: transposing changes nothing, i.e., $A^{\mathrm{T}}A = (A^{\mathrm{T}}A)^{\mathrm{T}}$. Namely, $(A^{\mathrm{T}}A)^{\mathrm{T}} = A^{\mathrm{T}}(A^{\mathrm{T}})^{\mathrm{T}} = A^{\mathrm{T}}A$. Sparse: most entries are zero when n (the number of rows and columns) gets large (Strang 2012: 2). Banded: the nonzeros lie in a "band" around the main diagonal (ibid). Our $A^{\mathrm{T}}A$ is nearly banded.

53) We solve $A^{\mathrm{T}}Ax = 0$, where $x = (x_1, x_2, x_3, ..., x_9)$ and $0 = (0, 0, ..., 0)$. If $A^{\mathrm{T}}A$ is solvable, an inverse $(A^{\mathrm{T}}A)^{-1}$ exists such that $(A^{\mathrm{T}}A)^{-1} \cdot A^{\mathrm{T}}Ax = (A^{\mathrm{T}}A)^{-1} \cdot 0$, i.e., x must be 0. However, nonzero solutions exist, i.e., infinitely many constant vectors $(c_1, c_2, c_3, ..., c_9)$ satisfy $A^{\mathrm{T}}Ax = 0$, which is a contradiction.
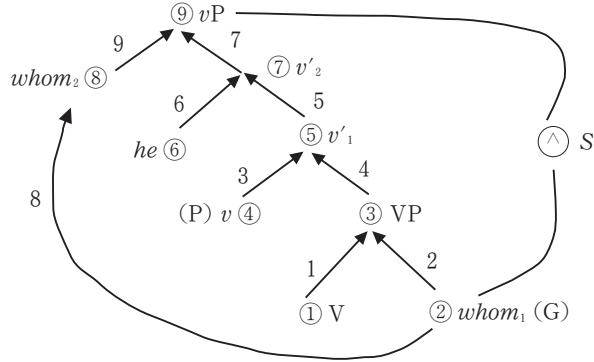
Figure 15: Network with a source S leaving ② and entering ⑨

$A^{\mathrm{T}}A$ is reduced, i.e., ② is fixed as zero. The reduced $A^{\mathrm{T}}A$ $(A^{\mathrm{T}}A_{\mathrm{reduced}})$ lacks row ② and column ②.

Table 8: $A^{\mathrm{T}}A_{\mathrm{reduced}}$

	① V	③ VP	④ v (P)	⑤ v'_1	⑥ he	⑦ v'_2	⑧ wh_2	⑨ vP
① V	1	−1	0	0	0	0	0	0
③ VP	−1	3	0	−1	0	0	0	0
④ v (P)	0	0	1	−1	0	0	0	0
⑤ v'_1	0	−1	−1	3	0	−1	0	0
⑥ he	0	0	0	0	1	−1	0	0
⑦ v'_2	0	0	0	−1	−1	3	0	−1
⑧ wh_2	0	0	0	0	0	0	2	−1
⑨ vP	0	0	0	0	0	−1	−1	2

$A^{\mathrm{T}}A_{\mathrm{reduced}}$ is solvable. We removed all-one solution $(1, 1, 1, …, 1)$ or $(c_1, c_2, c_3, …, c_8)$ from solutions to $A^{\mathrm{T}}Ax=0$. GE yields U of $A^{\mathrm{T}}A_{\mathrm{reduced}}$.

Table 9: U of $A^\mathrm{T}A_{\text{reduced}}$

	① V	③ VP	④ v (P)	⑤ v'_1	⑥ he	⑦ v'_2	⑧ wh_2	⑨ vP
① V	1	−1	0	0	0	0	0	0
③ VP	0	2	0	−1	0	0	0	0
④ v (P)	0	0	1	−1	0	0	0	0
⑤ v'_1	0	0	0	3/2	0	−1	0	0
⑥ he	0	0	0	0	1	−1	0	0
⑦ v'_2	0	0	0	0	0	4/3	0	−1
⑧ wh_2	0	0	0	0	0	0	2	−1
⑨ vP	0	0	0	0	0	0	0	3/4

GE reveals that the rank r of $A^\mathrm{T}A_{\text{reduced}}$ is $r=8$.[54] Let x_n be potential at node ⓝ. We solve $A^\mathrm{T}Ax=b$, where $x=(x_1, x_2, x_3, …, x_9)$ and $b=(0, 0, 0, …, S)$. The potentials at nodes are as follows.[55]

(76) Potentials at nodes in the balanced vP

$x_9=4/3S$

$x_8=2/3S$

$x_7=S$

$x_6=S$

$x_5=2/3S$

$x_4=2/3S$

$x_3=1/3S$

$x_2=0$ (grounded)

$x_1=1/3S$

The net potential is $6S$, which is used to balance vP. The potential of P is $2/3S$; the potential of G is 0; and the potential of internally merged G is $2/3S$. The vP node contains the greatest po-

---

54) Reshish matrix calculator. The computation time was 0.016 sec. The dot product of diagonal of U is det A, which is 6 for non-wh-in-situ languages.

55) The calculation is as follows. $3/4x_9=S$, $x_9=4/3S$; $2x_8-x_9=0$, $x_8=1/2x_9=1/2\cdot4/3S=2/3S$; $4/3x_7-x_9=0$, $x_7=3/4x_9=3/4\cdot4/3S=S$; $x_6-x_7=0$, $x_6=x_7=S$; $3/2x_5-x_7=0$, $x_5=2/3x_7=2/3S$; $x_4-x_5=0$, $x_4=x_5=2/3S$; $2x_3-x_5=0$, $x_3=1/2x_5=1/2\cdot2/3S=1/3S$; $x_2=0$ (grounded); $x_1-x_3=0$, $x_1=x_3=1/3S$.

tential. Ohm's law $y=-CAx$ yields the nine currents.[56] Here, assume $C=1$ or the identity matrix I for simplicity.

(77) Currents on edges in the balanced vP

$$y_1=-(1/3S-1/3S)=0$$

$$y_2=-(1/3S-0)=-1/3S$$

$$y_3=-(2/3S-2/3S)=0$$

$$y_4=-(2/3S-1/3S)=-1/3S$$

$$y_5=-(S-2/3S)=-1/3S$$

$$y_6=-(S-S)=0$$

$$y_7=-(4/3S-S)=-1/3S$$

$$y_8=-(2/3S-0)=-2/3S$$

$$y_9=-(4/3S-2/3S)=-2/3S$$

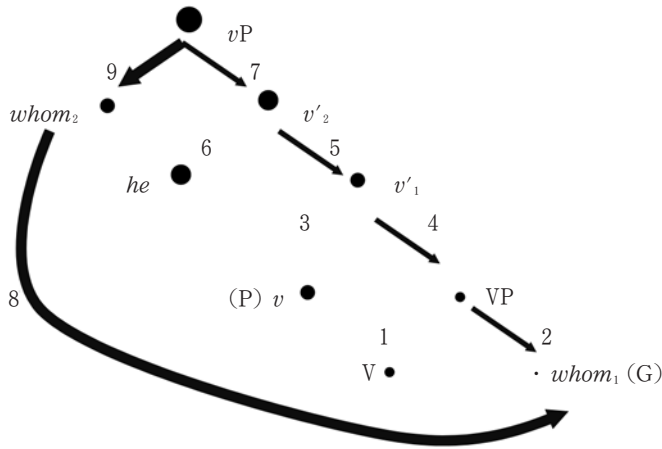Let us visualize potential and current difference by node and edge size.

Figure 16: Graph-theoretical vP: the essence of vP in equilibrium

The absolute net current is $8/3S$. The currents on edges 8 and 9 are respectively two times stronger than those on the edges forming the loop (edge 7, 5, 4, 2). An edge disappears when the current is zero. Edge 3 (head projection of P) disappears in vP in equilibrium. P uses its projec-

---

56) The reason for the minus sign in Ohm's law is as follows. A current flows from a higher-potential node ⊕ to a lower-potential node ⊖. We assign -1 to a starting node and 1 to an end node. Potential difference $Ax=(⊕-⊖)$ becomes negative. We require the minus sign to make it positive.

tion to search G under AGREE but the head projection is deleted in the balanced vP. IM yields edge 8 and forms the loop, which equilibrates the entire structure. Nature distributes currents to minimize error (Strang 2009: 428). The halfway vP phase has minimal error, thereby obeying the principle of minimal computation (MC).[57]

What about wh-in-situ languages such as Chinese where the wh-phrase is pronounced at the original position? As for wh-in-situ vP, we do not ground ② (wh_1) because it is pronounced. We ground ⑧ because wh-phrases internally merged at higher positions are not pronounced in wh-in-situ language. S exits ⑧ and enters ⑨.
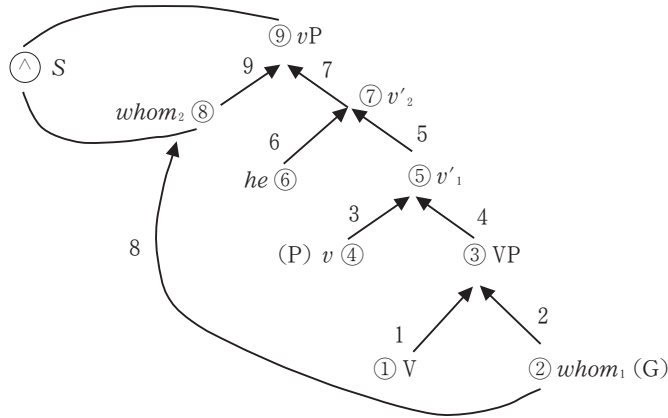

Figure 17: Network with a source S leaving ⑧ and entering ⑨ (wh-in-situ vP)

$A^{\mathrm{T}}A$ is reduced, i.e., ⑧ is fixed as zero. The reduced $A^{\mathrm{T}}A$ ($A^{\mathrm{T}}A_{\text{reduced}}$) lacks row ⑧ and column ⑧.

Table 10: $A^{\mathrm{T}}A_{\text{reduced}}$ (wh-in-situ vP)

	① V	② wh_1 (G)	③ VP	④ v (P)	⑤ v'_1	⑥ he	⑦ v'_2	⑨ vP
① V	1	0	−1	0	0	0	0	0
② wh_1 (G)	0	2	−1	0	0	0	0	0
③ VP	−1	−1	3	0	−1	0	0	0
④ v (P)	0	0	0	1	−1	0	0	0
⑤ v'_1	0	0	−1	−1	3	0	−1	0
⑥ he	0	0	0	0	0	1	−1	0
⑦ v'_2	0	0	0	0	−1	−1	3	−1
⑨ vP	0	0	0	0	0	0	−1	2

---

57) We demonstrate a mere simplified halfway vP for expository purposes. A more elaborate analysis must consider the full CP with additional IM of V, he, T ($does$), and $whom$, obeying PIC. The essence remains.

$A^{\mathrm{T}}A_{\mathrm{reduced}}$ is solvable. We removed the all-one solution $(1, 1, 1, ..., 1)$ or $(c_1, c_2, c_3, ..., c_8)$ from solutions to $A^{\mathrm{T}}Ax=0$. GE yields U of $A^{\mathrm{T}}A_{\mathrm{reduced}}$.

Table 11 : U of $A^{\mathrm{T}}A_{\mathrm{reduced}}$ (wh-in-situ vP)

	① V	② wh_1 (G)	③ VP	④ v (P)	⑤ v'_1	⑥ he	⑦ v'_2	⑨ vP
① V	1	0	−1	0	0	0	0	0
② wh_1 (G)	0	2	−1	0	0	0	0	0
③ VP	0	0	3/2	0	−1	0	0	0
④ v (P)	0	0	0	1	−1	0	0	0
⑤ v'_1	0	0	0	0	4/3	0	−1	0
⑥ he	0	0	0	0	0	1	−1	0
⑦ v'_2	0	0	0	0	0	0	5/4	−1
⑨ vP	0	0	0	0	0	0	0	6/5

GE reveals the rank r of $A^{\mathrm{T}}A_{\mathrm{reduced}}$ as $r=8$.[58] Here, let x_n be the potential at node ⓝ. We solve $A^{\mathrm{T}}Ax=b$, where $x=(x_1, x_2, x_3, ..., x_9)$ and $b=(0, 0, 0, ..., S)$. The potentials at nodes are as follows.[59]

(78) Potentials at nodes in the balanced vP (wh-in-situ vP)

$x_9=5/6S$

$x_8=0$ (grounded)

$x_7=2/3S$

$x_6=2/3S$

$x_5=1/2S$

$x_4=1/2S$

$x_3=1/3S$

$x_2=1/6S$

$x_1=1/3S$

---

58) Reshish matrix calculator. The computation time was 0.251 sec. Det $A=6$ for wh-in-situ languages; the same number for non-wh-in-situ languages.

59) The calculation is as follows. $6/5x_9=S$, $x_9=5/6S$; $x_8=0$ (grounded); $5/4x_7-x_9=0$, $x_7=4/5x_9=4/5\cdot 5/6S=2/3S$; $x_6-x_7=0$, $x_6=x_7=2/3S$; $4/3x_5-x_7=0$, $x_5=3/4x_7=3/4\cdot 2/3S=1/2S$; $x_4-x_5=0$, $x_4=x_5=1/2S$; $3/2x_3-x_5=0$, $x_3=2/3x_5=2/3\cdot 1/2S=1/3S$; $2x_2-x_3=0$, $x_2=1/2x_3=1/2\cdot 1/3S=1/6S$; $x_1-x_3=0$, $x_1=x_3=1/3S$.

Here, the net potential is $4S$, which is used to balance vP. The potential of P is $1/2S$; the potential of G is $1/6S$; and the potential of internally merged G is 0 (grounded). The vP node contains the largest potential. Ohm's law $y=-CAx$ yields nine currents. Assume $C=1$ or the identity matrix I for simplicity.

(79) Currents on edges in the balanced vP (wh-in-situ vP)

$y_1=-(1/3S-1/3S)=0$

$y_2=-(1/3S-1/6S)=-1/6S$

$y_3=-(1/2S-1/2S)=0$

$y_4=-(1/2S-1/3S)=-1/6S$

$y_5=-(2/3S-1/2S)=-1/6S$

$y_6=-(2/3S-2/3S)=0$

$y_7=-(5/6S-2/3S)=-1/6S$

$y_8=-(0-1/6S)=1/6S$

$y_9=-(5/6S-0)=-5/6S$

Let us visualize potential and current difference by node and edge size.
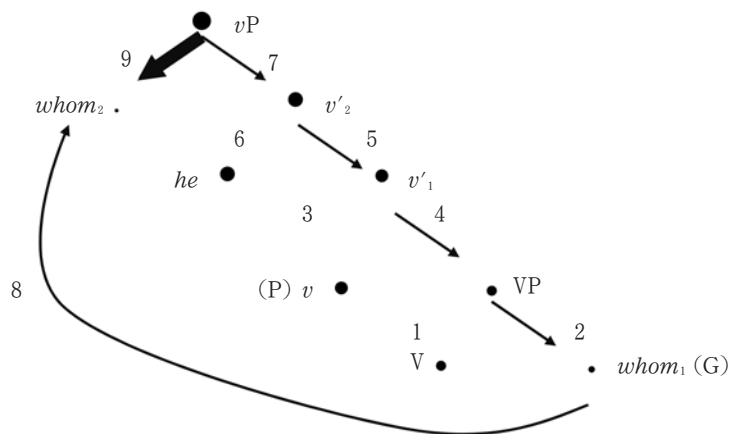

Figure 18: Graph-theoretical wh-in-situ vP: the essence of wh-in-situ vP in equilibrium

The following table summarizes the differences between wh-in-situ vP and non-wh-in-situ vP.

Table 12：Network properties of wh-in-situ vP vs. non-wh-in-situ vP

	Wh-in-situ vP	Non-wh-in-situ vP
GE time for $A^{\mathrm{T}}A_{\mathrm{reduced}}$	0.251 sec	0.016 sec
Net potential	$4S$	$6S$
Potential of P	$1/2S$（$=0.5S$）	$2/3S$（$\fallingdotseq 0.67S$）
Potential of G	$1/6S$（$\fallingdotseq 0.17S$）	0 (grounded)
Absolute net current	$10/6S$（$\fallingdotseq 1.67S$）	$8/3S$（$\fallingdotseq 2.67S$）
Absolute current on edge 8	$1/6S$（$\fallingdotseq 0.17S$）	$2/3S$（$\fallingdotseq 0.67S$）
Direction of current on edge 8	$\mathrm{wh}_2 \leftarrow \mathrm{wh}_1$	$\mathrm{wh}_2 \rightarrow \mathrm{wh}_1$

The time required for GE for $A^{\mathrm{T}}A_{\mathrm{reduced}}$ in wh-in-situ vP is approximately 16 times longer than that in non-wh-situ vP. Here, assume that longer computation time means greater complexity. With respect to GE time, wh-in-situ vP is more complex than non-wh-in-situ vP. The net potential of wh-in-situ vP is approximately 67% that of non-wh-in-situ vP. The absolute net current of wh-in-situ vP is approximately 60% of that of non-wh-in-situ vP, and the absolute current on edge 8 of wh-in-situ vP is approximately 25% of that of non-wh-in-situ vP. The potential and current required to form a loop in wh-in-situ vP are less than those in non-wh-in-situ vP. With respect to potential and current, wh-in-situ vP is more economical and less complex, i.e., it requires less energy. This is a mathematical proof of Huang's（1982）hypothesis that covert wh-movement is costless. A notable result is that the current on edge 8 in wh-in-situ vP is the reverse of that in non-wh-in-situ vP. In wh-in-situ vP, information flows into wh_2, which is internally merged at a higher position, while in non-wh-in-situ vP, information flows into wh_1（G）, which is externally merged at the original position.

2.6.6. Summary：Have we learned anything?

Let us summarize what we have learned. Our argument relies on a crucial mathematical fact as follows.

（80）Coefficients and constants express ratio（symmetry）of unknowns.

Two important proposals of the paper are as follows.[60]

(81) a. If uFs were unknowns weighted by coefficients, then

（ⅰ） P and G are fully distinct (asymmetrical) in complete AGREE, identical and distinct (contradictory) in incomplete AGREE, and fully identical (symmetrical) in IM.

（ⅱ） IM in uFE corresponds to infinite solutions in GE. C$_M$ cannot tolerate infinite solutions but C$_{HL}$ can and recycles it as IM.

b. IM yields loops. Loops balance sentence structures. Error is minimized in the balanced structures.

Additional claims are the following.

(82) a. C$_M$ contains operations as AGREE, IM, and TRANSFER.

b. uFE and GE are translatable.

c. Linear algebraic uFE avoids look-ahead problem.

d. The value of $u\varphi$ contains the value of $uCase$, and vice versa. They are reflex of each other.

Graph theory reveals how wh-in-situ vP and non-wh-in-situ vP differ.

(83) Differences between wh-in-situ vP and non-wh-in-situ vP

a. With respect to GE time for $A^{\mathrm{T}}A_{\mathrm{reduced}}$, wh-in-situ vP is more complex than non-wh-in-situ vP.

b. With respect to potential and current, wh-in-situ vP is less complex than non-wh-in-situ vP.

c. The current on edge 8 reverses in the two types of vP in equilibrium. In wh-in-situ vP, information flows into wh$_2$, whereas in non-wh-in-situ vP, information flows into wh$_1$ (G).

3. Risks and hopes: Limitations, problems, and future work

Does GE teach us anything about uFE? It may teach us nothing. It may be pointless to compare GE to uFE, as an anonymous reviewer points out. The reviewer warns us that linear algebraic uFE has the following fundamental shortcomings and risk.

(84) Shortcomings

a. It overgeneralizes Chomsky's claim that mathematical capacity (meant as arithmetic)

_____

look-ahead problem, i.e., symmetry must break in order to make the syntactic object interpretable at CI.

stems from language capacity.

b. It does not consider the basic ontology of C_M and GE, or their fundamental differences with C_{HL}.

c. It does not fully consider the underlying assumptions of the feature checking theory.

d. It overemphasizes the apparent similarity between the feature-checking mechanism and GE based on cherry-picked examples and methods, e.g., elimination and merge.

(85) Risk

It runs the risk of practicing pseudoscience and pseudolinguistics.

With respect to (84b), we have discarded the coarse claim of the last draft in which it was stated that C_M stems from C_{HL}. The main attempt of this paper is narrower; compare uFE to GE. We have focused on GE, which is the most commonly used technique in software to solve equations. Computers use GE, which is an elementary calculation for scientific computing. We compared two types of elimination, i.e., natural ("nature-made") uFE with artificial (manmade) GE.[61]
Let us assume that the instruction for uF is "enter F without value" (Chomsky 2001b: 16). When uF is eliminated successfully, we have complete AGREE. Success or failure of elimination matters to GE. Is a uF similar to an unknown modified by a coefficient? The reviewer warns that asking such a question overgeneralizes Chomsky's speculation. Chomsky speculates that human arithmetical capacity stems from Merge.

(86) "[T]here happen to be very simple ways to get arithmetic from Merge. Take the concept Merge, which simply says, take two things, and construct a thing that is the set of the two things; that's its simplest form. Suppose you restrict it, and 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. ... [I]t's just a trivial complication of Merge, which restricts it and says, when you put everything in just this way, it does give you arithmetic. When you've got the successor function, the rest comes" (Chomsky and McGilvray 2012: 15).

---

61) The reviewer states that syntactic operations such as uFE are mere metaphors rather than actual procedural applications, even though they are presented as such in most linguistic publications. The reviewer's understanding is consistent with methodological naturalism (Chomsky 2000: 76), i.e., we may never understand the "actual procedures" due to our biological/cognitive limitations, and all we can do is to create explanatory theories ("metaphors" including mathematical models) that are intelligible to us. See the next section.

Pointing out the lack of clarity of evidence in Butterworth (2000) with "many arguments against thinking that the language and arithmetical capacities are related," Chomsky states the following.

(87) "This is an old problem. Alfred Russell Wallace was worried about it. He recognized that mathematical capacities could not have developed by natural selection; … a natural expectation is that they're an offshoot of something. They're an offshoot of, probably like most of the rest of what's called 'human intellectual capacity' [or reason], something like language."

　　"So what we're left with is speculation, but when you don't have enough evidence, you pick the simplest explanation. And the simplest explanation that happens to conform to all the evidence we have is that it[arithmetic]'s just an offshoot of language derived by imposing a specific restriction on Merge" (Chomsky and McGilvray (2012: 15-16).

As the reviewer correctly points out, Chomsky refers to human's arithmetic capacity, i.e., the use of natural number for counting (Hauser, Chomsky, and Fitch 2002: 1576, Chomsky 2007: 7). However, Wallace (1889 [2009]: 466-467) mentioned the computational procedures of mathematics (C_M) in a broad sense. Wallace questioned as follows.

(88) "[Let us inquire] how this rudimentary faulty [of counting of savages] became rapidly developed into that of a Newton, a La Place, a Gauss, or a Cayley. … What motive power caused its development?" (ibid. 466)

Wallace claimed that Euclidian mathematics has nothing to do with the theory of "the fittest to survive in the great struggle of races."

(89) "The Greeks did not successfully resist the Persian invaders by any aid from their few mathematicians, but by military training, patriotism, and self-sacrifice." (ibid. 467)

Wallace concluded as follows.

(90) "We conclude … that the present gigantic development of the mathematical faculty is wholly unexplained by the natural selection, and must be due to some altogether distinct cause." (ibid. 467)

The term "gigantic" seems to indicate that he was assuming more general C_M, not just arithme-

tic. Wallace did not hint that "some altogether distinct cause" was "human intellectual capacity [or reason], something like language." As the reviewer correctly points out, linear algebraic uFE overgeneralizes Chomsky's speculation. The issue is whether the overgeneralization is meaningful. This paper compares uFE to GE to provide hints for this debate.

What is striking is that the three solvability patterns in GE seem to correspond to AGREE and IM in a simple way.

(91) Three types of solvability

① Successful GE and complete AGREE (uFE): the ratio of coefficient and constant of uF in Probe (P) and Goal (G) is asymmetrical (symmetry fully broken); a unique solution (a unique pair of values of $uCase$ and $u\varphi$); invertible (the computation is reversible);[62] successful elimination.

② Failed GE and incomplete AGREE (uFE): the ratio of coefficient and constant of uF in P and G is symmetrical and asymmetrical (contradiction); symmetry is half broken; no solution; no unique pair of values of $uCase$ and $u\varphi$; not invertible; elimination breaks down.

③ Failed GE and no AGREE in uFE (IM): the ratio of coefficient and constant of uF in P and G is symmetrical; symmetry is fully preserved; infinitely many solutions; infinitely many pairs of values of $uCase$ and $u\varphi$; not invertible; elimination breaks down in GE but not in uFE.

For P and G to completely agree, they must be fully different. When P and G are identical and different (contradiction), they agree incompletely. When P and G are identical, G internally merges with P. This paper claims that this parallelism teaches us something. The parallelism may be more than a seeming similarity between uFE and GE. The analogy is so general that it cannot be induced from cherry-picked examples and methods.[63]

---

62) Successful GE means that the computation can be undone. For a successfully eliminated matrix A, the inverse A^{-1} exists, where $A \cdot A^{-1} = I$ (identity matrix). The inverse solves $Ax = b$ efficiently as in $A^{-1}Ax = Ix = x = A^{-1}b$. Most importantly, successful GE means that A is factorized into $A = LU$, where L is the lower triangular matrix. Below the diagonal, L contains M (multipliers or matchmakers) and our memory of GE (Strang 2009: 97). L records the steps of GE by storing the multipliers (Strang 2007: 29). L reverses the GE steps. L takes U back to A. U records the final result (ibid). Here is a connection to the perfect TRANSFER (Epstein, Kitahara, and Seely 2010: 132), which contains L with memory of AGREE. The perfect TRANSFER serves CI and has a "clairvoyant power" to read the "criminal record" inside the lost phase complement (Epstein, Kitahara, and Seely 2010: 132-139; See fn. 14). The factorization $A = LU$ is an algebraic proof of perfect TRANSFER, which is realized as L that reverses computation. The availability of the inverse and factorization means successful and efficient computation.

63) The reviewer points out the following. "[I]t is pointless to examine if the mechanism involved in GE mirrors that of the feature checking mechanism [uFE]. Chances are that any two cognitive domains may

An anonymous reviewer suggests that linear algebraic uFE is not so outlandish. The reviewer points out that "the fact that feature checking can be represented in algebraic terms isn't all that surprising. … In a certain sense, …, all of syntax can already be understood as sequences of algebraic manipulations." According to the reviewer, "minimalist grammars (Stabler 1997) with the Agree operation can be described in terms of regular derivation tree languages (Graf 2011, Kobele 2011), which in turn can be converted to context-free grammar (Thatcher 1967)." The reviewer suggests that we must define (ⅰ) a formal agreement algebra consisting of a carrier set and some operations over it, (ⅱ) a syntactic algebra, e.g., in terms of tree languages, and (ⅲ) a full worked-out mapping from (ⅰ) to (ⅱ).

However, this formal legwork is beyond the author's ability. The paper can only point out that linear algebraic uFE dispenses with features, such as $+f$ and $+F$ (strong features) that are triggers for phrasal movement (Stabler 1997). We propose that phrasal movement is a uFE-recycle of infinitely many solutions in GE. The author is a linguist who is self-educating mathematics with support by colleague mathematicians and Professor Strang through the Massachusetts Institute of Technology Open Courseware. The lack of mathematical formality comes from the author's inability to comprehend academic literature in computational linguistics. However, this paper aims to connect the comparison with formal research. The reviewer states, "I expected a full formal treatment that clearly states its primitives, axioms and domain of operation and then proceeds to construct a syntactic agreement algebra to which the standard methods of GE can be adapted in a straight-forward manner." Another reviewer states "If I am mistaken, and the parallels between GE and syntactic feature checking are actually stronger than I could understand, then this project will need drastic revision to be clearer, more methodical, and understandable to linguists. A great start would be a (hopefully completely formal) method for translating a syntactic structure into a series of equations with actual numbers." These are beyond our ability. Reading the last draft, the reviewer pointed out, "I am still intrigued by the basic idea, but the paper in its current form does a bad job at advocating for it." Despite many shortcomings, we hope this version is at least clearer than the previous version in showing the goal of this paper, i.e., a rudimentary comparison of uFE to GE.

4. Concluding remarks: Linear algebraic uFE in a bigger picture

The reviewer's criticism is important and connects with cognitive and evolutionary puzzles about C_{HL}. We reproduce the criticism (84c).

---

share some properties, but it does not entail that a direct parallelist thesis in one fell swoop is in order." This paper inherits Ross' (1967) insights into syntactic variables, as shown in the dissertation title "Constraints on variables in syntax." What is the nature of variables (unknowns) in C_{HL}? How different are they from those used in GE?

(92) [The paper] does not fully consider the underlying assumptions of the feature checking theory.

We believe that the fundamental mystery of uFE is as follows.

(93) "A major problem is why uninterpretable features and Agree exist at all" (Chomsky 2001b: 16).

"Why do [−Int] (uninterpretable) features such as EF and unvalued features (such as *phi* on T and Case on N) enter this model in the first place?" (Epstein, Kitahara, and Seely 2010: 134).

Mathematicians do not ask why unknowns or GE exist. Linguists ask why the human brain, a natural object, contains C$_{HL}$ with uF and uFE. Piattelli-Palmarini and Uriagereka (2004) suggest that C$_{HL}$ is a virus-checking system evolved in the human brain, which is a typical immune system. Humans have created computers. Computers use GE. Computers are vulnerable to computer viruses, i.e., electrical information that causes computational turbulence. The human brain is similar (both use electrical (digital) information) and different from computers (only the human brain uses chemical (analogue) information). The human brain is a typical immune system. It is natural that C$_{HL}$ has emerged as a virus-checking system in the immune system (the human brain). The total human body is a virus-checking system. The human language system, being a computational organ evolved in an immune system, disguises itself as a virus-checking system. Chomsky speculates that free Merge and undeleted EF have produced unbounded Merge.

(94) "At the minimum, some rewiring of the brain, presumably a small mutation or a by-product of some other change, provided Merge and undeleted EF (unbounded Merge), yielding an infinite range of expressions constituted of LIs (perhaps already available in part at least as conceptual atoms of CI), and permitting explosive power of the capacities of thought" (Chomsky 2007: 14).

We claim that undeleted EF is unnecessary to explain the "EPP-effect." Chomsky claims that uF contributes to efficient computation, thereby obeying the principle of minimal computation (MC).

(95) "[T]hey [uFs] compel phases to be as small as possible consistent with IM and (possibly) assignment of argument structure, CP and v*P, and they impose cyclicity of Transfer (strict

cyclicity, given PIC, thus reducing memory load in computation. Hence, they contribute to SMT" (ibid.: 24).

Linear algebraic uFE claims that CHL solves a series of a system of simultaneous P and G equations that consists of unknowns (uFs or viruses) modified by coefficients (iF) that yields a unique solution (complete AGREE), no solution (incomplete AGREE), or infinitely many solutions (IM).

There are two attitudes toward CHL research, i.e., methodological dualism and methodological naturalism (Chomsky 2000). Methodological dualism is summarized as follows.

(96) Methodological dualism

"The view that we must abandon scientific rationality when we study humans 'above the neck' (metaphorically speaking), becoming mystics in this unique domain ..." (Chomsky 2000: 76). It gives up "the hope of eventual integration" of CHL research and "the 'core' natural sciences" (ibid.).

Methodological naturalism is as follows.

(97) Methodological naturalism

An approach that seeks "to construct intelligible explanatory theories, with the hope of eventual integration with the 'core' natural sciences" (Chomsky 2000: 76).

It may be too early to imagine "the hope," as Chomsky often warns us; the current level of CHL research corresponds to pre-Galilean physics (Chomsky 2014; lecture 2). We may need to accumulate descriptive generalization and construct intelligible explanatory theories in the next five centuries before realizing the hope of eventual integration with the 'core' natural sciences. As Chomsky mentions, chemistry teaches us to be patient.

(98) "Into the 1920s, ... leading scientists would have just ridiculed the idea of taking any of this seriously, ... They though of [atoms, [molecules] and other such 'devices'] as ways of calculating the results of experiments. Atoms can't be taken seriously, because they don't have a physical explanation, which they didn't. Well, it turned out that the physics of the time was seriously inadequate; you had to radically revise physics to be unified with and merged with an unchanged chemistry" (Chomsky and McGilvray 2012: 19).

The reviewer also warns us that we shouldn't be too hasty to discard an alternative claim that GE and uFE may have much less in common than the author claims. Based on the inconsistency of run-time complexity, the reviewer doubts the feasibility of algorithmic parallels between C$_{HL}$ and C$_M$ (or the former even subsuming the latter). Although it is beyond the author's ability to propose a model that calculates the minute cost of GE and uFE, we speculate that EF ("infinite solutions") decreases computational complexity radically, thereby answering Plato's Problem for C$_{HL}$: Why does a child acquire the mother language(s) so easily, i.e., why is the initial state so costless? EF characterizes human natural language.[64]

We are impatient and compare uFE and GE. This paper imagines a glimpse at what "the hope" may look like. Like Chomsky, the reviewer is correct; we should do real linguistics, i.e., construction of intelligible explanatory theories of C$_{HL}$, rather than pseudolinguistics or pseudoscience. However, even if linear algebraic uFE turns out to be pseudolinguistics, pseudoscience and pointless, we believe that the comparison, as a negative exemplum, will contribute to the construction of intelligible explanatory theories of C$_{HL}$, i.e., it shows us how and why linear algebraic uFE is meaningless and fails. In this sense, linear algebraic uFE can contribute to methodological naturalism; the construction of intelligible explanatory theories of C$_{HL}$.

References

Bošković, Ž. 2002. A-movement and the EPP. *Syntax* 5: 167-218.

Bošković, Ž. 2007. On the locality and motivation of Move and Agree: An even more minimal theory. *Linguistic Inquiry* 38, 589-644. Cambridge: MIT Press.

Butterworth, B. 2000. *The Mathematical Brain*. London: Macmillan.

Calinger, R. 1999. *A Contextual History of Mathematics*, Prentice Hall.

Castillo, J. C., J. Drury and K. K. Grohmann. 1999. Merge over move and the extended projection principle. In Aoshima, S., J. Drury, T. Neuvonen (eds.) *University of Maryland Working Papers in Linguistics* 8: 63-103. University of Maryland, College Park: Department of Linguistics.

Chomsky, N. 1957. *Syntactic Structures*. The Hague: Mouton.

Chomsky, N. 1977. On WH movement. In Culicover, T., Wasow, T., and A. Akmajian (eds.) *Formal Syntax*. Academic Press. New York.

Chomsky, N. 1995. *The Minimalist Program*. Cambridge: MIT Press.

Chomsky, N. 2000a. *New Horizons in the Study of Language and Mind*. Cambridge University Press.

Chomsky, N. 2000b. Minimalist inquiries: the framework. In R. Martin, D. Michaels & J. Uriagereka (eds.) *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik*, 89-155. Cambridge, MA: MIT Press.

Chomsky, N. 2001a. Derivation by phase. In M. Kenstowicz (ed.) *Ken Hale: A Life in Language*, 1-52.

---

64) See footnote 44. IM occurs when uFE uses $\lambda=3$. When A preserves symmetry, i.e., the output parallels the input, the output is stretched by a factor of 3. The stretch property causes IM: structural growth. More generally, C$_{HL}$ is expressed as $Ax=\lambda x$, where A indicates the initial state of C$_{HL}$, the unknown x indicates degenerate linguistic input, λ indicates the final state, and the unknown x indicates mother language(s).

Cambridge, MA: MIT Press.

Chomsky, N. 2001b. Beyond explanatory adequacy. *MIT Occasional Papers in Linguistics* 20. 1-28. Cambridge, MA: MIT Press.

Chomsky, N. 2005. Three factors in language design. *Linguistic Inquiry* 36: 1-22.

Chomsky, N. 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, N. 2008. On phases. In R. Freidin, C. P. Otero and M. L. Zubizarreta (eds.) *Foundational Issues in Linguistic Theory. Essays in Honor of Jean-Roger Vergnaud*, 133-166. Cambridge, MA: MIT Press.

Chomsky, N. 2013. Problems of projection. *Lingua*, 130: 33-49.

Chomsky, N. 2014. Chomsky lecture series. MIT TechTV, MIT libraries.

Chomsky, N. and G. A. Miller. 1958. Finite state languages. *Information and Control* 1, 91-112.

Chomsky, N. and H. Lasnik. 1995. The theory of principles and parameters. In Chomsky, N. (ed.) 2015. *The Minimalist Program: 20TH Anniversary Edition*. 11-116. Cambridge: MIT Press.

Chomsky, N. and J. McGilvray. 2012. *The Science of Language: Interviews with James McGilvray*. Cambridge: Cambridge University Press.

Epstein, S. D. and T. D. Seely. 2006. *Derivations in Minimalism*. Cambridge: Cambridge University Press.

Epstein, S. D., H. Kitahara and T. D. Seely. 2010. Uninterpretable features: what are they and what do they do? In Putnam, M. T. (ed.) *Exploring Crash-Proof Grammars*. 125-142. Amsterdam: John Benjamins.

Epstein, S. D., H. Kitahara and T. D. Seely. 2012. Structure building that can't be! In Uribe-etxebarria, M. and Valmala, V. (eds.) *Ways of Structure Building*, 253-270. Oxford: Oxford University Press.

Epstein, S. D., H. Kitahara and T. D. Seely. 2014. Labeling by minimal search: implications for successive-cyclic A-movement and the conception of the postulate "phase" *Linguistic Inquiry* 45: 463-481.

Frampton, J. and S. Gutmann. 2002. Crash proof syntax. In *Derivation and explanation in the Minimalist Program*, Epstein, S. D. and T. D. Seely (ed.) 90-105. Oxford: Blackwell.

Graf, T. 2011. Closure properties of minimalist derivation tree languages. In Pogodalla, S. and J.-P. Prost (eds.) *LACL 2011*, 6736: 96-111. Lecture Notes in Artificial Intelligence. Heidelberg: Springer.

Grcar, J. F. 2011. Mathematicians of Gaussian elimination, *Notices of the American Mathematical Society* 58 (6): 782-792.

Hauser, M., N. Chomsky, and T. Fitch. 2002. The faculty of language. What is it, who has it, and how did it evolve? *Science*, 298: 1569-1579.

Hiraiwa, K. 2005. Dimensions of symmetry in syntax: agreement and clausal architecture. Doctoral dissertation, MIT.

Huang, C.-T. J. 1982. Logical relations in Chinese and the theory of grammar. Doctoral dissertation, MIT.

Jacob, F. 1977. Evolution and tinkering. *Science* 196: 1161-1166.

Jacobson, P. 1992. Raising without movement. In Larson, R. *et al.* (eds.) *Control and Grammatical Theory*, Dordrecht: Kluwer, 149-194.

Kobele, G. M. 2011. Minimalist tree languages are closed under intersection with recognizable tree languages. In Pogodalla, S. and J.-P. Prost (eds.) *LACL 2011*, 6736: 129-144. Lecture Notes in Artificial Intelligence. Heidelberg: Springer.

Mandelbrot, B. B. 1977. *The Fractal Geometry of Nature*. New York: W. H. Freeman and Company.

Manzini, M. R. and A. Roussou. 2000. A minimalist theory of A-movement and control. *Lingua* 110: 409-447.

Martin, R. 1999. Case, the extended projection principle, and minimalism. In Epstein, S. D. and N. Hornstein (eds.) *Working Minimalism*, 1-26. Cambridge, MA: MIT Press.

Martin, R. 2001. Null Case and the distribution of PRO. *Linguistic Inquiry* 32, 141-166.

McCloskey, J. 1996. On the scope of verb movement in Irish. *Natural Language and Linguistic Theory* 14: 47-104.

Miyagawa, S. 2010. *Why Agree, Why Move: Unifying Agreement-Based and Discourse-Configurational Languages*. Cambridge: MIT Press.

Moro, A. 2000. *Dynamic Antisymmetry*. Cambridge, MA: MIT Press.

Obata, M. and S. Epstein. 2011. Feature-splitting internal merge: improper movement, intervention, and the A/A' distinction. *Syntax* 2.2: 122-147.

Piattelli-Palmarini, M. and J. Uriagereka. 2004. The immune syntax: the evolution of the language virus. In Jenkins, L. (ed.) *Variation and Universals in Biolinguistics*. 341-377. Elsevier.

Putnam, M. T. (ed.) 2010. *Exploring Crash-Proof Grammars*. Amsterdam: John Benjamins.

Richards, M. 2007. On feature inheritance: an argument from phase impenetrability condition. *Linguistic Inquiry* 38: 563-572.

Ross, J. R. 1967. Constraints on variables in syntax. Doctoral dissertation. MIT. Published as *Infinite Syntax!* Norwood, N. J.: Ablex (1986).

Shannon, C. E. and W. Weaver. 1949. *The mathematical theory of communication*. Urbana.

Stabler, E. P. 1997. Derivational minimalism. In Retoré, C. (ed.) *Logical Aspects of Computational Linguistics,* 1328: 68-95. Lecture Notes in Computer Science. Berlin: Springer.

Strang, G. 2005. Linear Algebra, Lecture 02: Elimination with matrices. MITOCW 18.06, Spring.

Strang, G. 2008. Mathematical Methods for Engineers I. MITOCW 18.085, Fall.

Strang, G. 2009. *Introduction to Linear Algebra* (4th edition). Wellesley-Cambridge Press.

Strang, G. 2011. Linear Algebra. MITOCW 18.06SC, Fall.

Stewart, I. 1995. *Concepts of Modern Mathematics*. Dover Publications. New York.

Thatcher, J. W. 1967. Charactering derivation trees of context-free grammars through a generalization of finite automata theory. *Journal of Computer and System Sciences* 1: 317-322.

Weber, H. 1893. Leopold Kronecker, *Mathematische Annalen*. (Springer Berlin/ Heidelberg) 43: 1-25.

Wallace, A. R. 1889 [2009]. *Darwinism*. London: Macmillan.

Watanabe, A. 2005. *Minimarisuto Puroguramu Josetsu* [Introduction to minimalist program]. Tokyo: Taisyuukan publisher.