Fora 1 (2017): 51-60

On Grammaticality and Selection toward the Simplest Syntax*

Yushi SUGIMOTO

1. Introduction

At the advent of generative grammar, it has been discussed that what is the adequate grammar for explaining the phenomena of language with the general criterion such as "simplicity." In fact, the evaluation measure of the grammar was "simplicity" in the standard theory (Chomsky 1965), which is specific to the linguistic system. In the end of 1980's / early 1990's, the idea that language is the perfect solution to interfaces puts the theory to beyond to the explanatory adequate theory of language, which is called minimalist program(MP). Minimalism is a program to try to explain the necessity of the universal grammar (UG): why it exists in this way, not the other way; how is it evolved (Chomsky 2004)? My view of generative grammar, however, is that minimalism is not a new idea, but it's the core idea under the generative grammar with the consistent notion of "simplicity." To unify some theories to one theory is better, much is less than more, etc.[1]

In this brief paper, I will clarify the notion of grammaticality and acceptability and I will point out a potential problem of selection to regard it as a property of grammar to seek the "simplest" minimalist approach to grammar. The organization of this paper follows: section 2 briefly reviews the relation between linguistics and cognitive science. Section 3 introduces the notion of the grammaticality and acceptability to understand what exactly narrow syntax (NS) works on. Section 4 shows the problem of the selection and argues that selection must not be introduced in the NS. Section 5 concludes this paper.

2. Language, Grammar and Theory

One of the contribution to the cognitive revolution in 1950's is the study of modern linguistics. The most key incidents were in several Chomsky's seminal works, which claims that language is internalized in our brain/mind, and it is insufficient to examine only the externalized objects. One of his early works is to show that phrase structure grammar is not enough to capture human language, let alone finite state grammar and introduced the transformational grammar (Chomsky 1956). Assuming that grammar generates the set of the grammatical strings, grammar has to generate only grammatical sentences ('grammatical'

*Part of this research is supported by Grant-in-Aid for JSPS Fellows (17J03420).

[1] See Chomsky (1965) for the details. See also Epstein (1990) for the case in which two theories cannot unify into one. In the framework of MP, the issue still remains. Whether a certain condition can be a rule/condition in narrow syntax or it's just the third factor, or it's interface conditions, etc.

means 'generative'[2]). In this sense, the challenge of generative grammar at first is to construct an explanatory theory of grammar, which 'explicitly' describes what native speaker unconsciously know about language. Another aspect of this approach paves the way to take grammar as the biological mechanism which is implemented in our brain in some sort of form. The orientation of the biological approach is not a recent idea, as Chomsky (2012: 12) says in the interview with James McGilvray[JM].

> "[Noam Chomsky: NC] Every since this business [generative grammar: YS] in the early fifties – two or three students, Eric Lenneberg, me, Morris Halle, apparently nobody else – the topic we were interested in was, how could you work this into biology? The idea was so exotic, no one else talked about it. Part of the reason was that ethology was just…"
> [JM] "Excuse me; was that [putting the theory of language into biology: JM] a motivation from the beginning?"
> [NC] "Absolutely: we were starting to read ethology, Lorenz, Tinbergen, comparative psychology; that stuff was just becoming known in the United States."

We can pose problems about language once we take it as a biological/internally implementation since the externalized object (which is called language) is the phenomena itself. The problems of the internalized system can be summarized as below.

(1) a. What constitutes knowledge of language?
b. How is knowledge of language acquire?
c. How is knowledge of language put to use?
d. How did these properties of the mind/brain evolve in the species?
e. How are these properties realized in mechanism of the brain?

First of all, to study language, we have to know what it is (1a). Then we can ask the next question how can we acquire the language or the use of language. The most of the time was spent to elucidate the internal system of language in the generative grammar. In the advent of the principles and parameters approach, we rose to the level what the UG looked like. The sets of principles, which is based on the UG component is not one of the possible format but invariant and language just grows in the language environment with the parameters setting. The next step is how to explain the UG itself? Assuming that internal system of language is

[2] See Chomsky (1965) for the details. To leave the detail, the notion of "generative" comes from the Turing's mathematics, in which the generative procedure is processed step by step, analogous to proof construction and gets the theorem, which becomes the input. In this sense, recursion, the one of the fundamental properties of human language, can be explained. Notice that generative grammar is intensional procedure that does not show infinite array of expression by itself.

implemented in our brain, the biological structure of the human being must evolve; the question is when and how this evolve? This problem arises only if we assume the system are implemented in an organic system. Notice that we should not mix the idea of the computation and implementation (e.g., We cannot observe Merge when we see our brains visually). The seminal work of Marr's (1982) computational theory shows us that there are three level of representations for seeking the explanatory theory. One is the computational level (2a). This is what Chomsky pursuits and Marr himself also emphasized the importance of this level.

(2) a. Computational theory:
What is the goal of the computation, why is it appropriate, and what is the logic of the strategy by which it can be carried out?
b. Representation and algorithm:
How can this computational theory be implemented? In particular, what is the representation for the input and output, and what is the algorithm for the transformation?
c. Hardware implementation:
How can the representation and algorithm be realized physically?

The next level "representation and algorithm" involves the time scale that interacts with the other cognitive systems and the other factors to perform language (2b). Implementation is, in this case, the neural-based model, which seems not to be clear so far (but see Friederici et al. 2017). To sum, generative grammar aims to construct a theory of syntax, which is explanatory, a biological system and it is a computational level, not descriptive and algorithm or implementational level.

3. On grammaticality/acceptability

As Chomsky's (1965) work mentioned, the notion of competence and performance has been divided and it has been emphasized that the theory of grammar put an attention to competence. [3] Performance, however, shows interesting phenomena, "acceptability." Acceptability is defined in Chomsky (1965) as follows.

> "to refer to utterances that are perfectly natural and immediately comprehensible without paper-and-pencil analysis, and in no way bizarre or outlandish. Obviously, acceptability a matter of degree, along various dimensions." (Chomsky 1965: 10)

He also notes on the difference between "grammatical" and "acceptability."

[3] See Miller and Chomsky (1963) for the detail of the language use model for performance.

> "The notion acceptable is not to be confused with 'grammatical'. Acceptability is a concept that belongs to the study of performance, whereas grammaticalness belongs to the study of competence…Like acceptability, grammaticalness is, no doubt, a matter of degree, but the scales of grammaticalness and acceptability do not coincide."
>
> (Chomsky 1965: 11)

There are four possible strings which we can predict are, as illustrated in (3).

(3) (i) grammatical and acceptable strings
(ii) grammatical, but unacceptable strings
(iii) ungrammatical, but acceptable strings
(iv) ungrammatical and acceptable strings

(3i) and (3iv) are not so surprising though as Chomsky (1965) pointed out there is no coincidence between grammaticalness and acceptability. As for (3ii), it is observed that self-embedding structure is difficult to parse, rather than nesting structures.[4,5]

(4) Anyone$_1$ who feels that if$_2$ so-many$_3$ more$_4$ students$_5$ whom we$_6$ haven't$_6$ actually admitted are$_5$ sitting in on the course than$_4$ ones we have that$_3$ the room had to be changed, then$_2$ probably auditors will have to be excluded, is$_1$ likely to agree that the curriculum needs revision. (Chomsky and Miller 1963: 286)

As for (3iv), there are interesting issues which are mainly explored at the field of psycholinguistics. In the examples in (5), *the deer* is temporally ambiguous whether it is the object of the verb *stalk* or the subject of the main clause (one of the *garden path effects*). Once it is interpreted through the whole sentence, it is expected to regard *the deer* as the subject of the main clause though subjects in an experiment regard *the deer* as the object of the verb, which results in ungrammatical, according to Traxler (2012: 176).

(5) a. While the hunter was stalking the deer frank from the puddle.
b. While the hunter was stalking the deer in the zoo drank from the puddle

Another case is below. (6a) and (6b) are bizarre since in common sense, foods do not eat animals. In this case, 'good-enough' approach (Ferrreira and Patson 2007) suggests that our

[4] See also Chomsky (1965: 13-14): "…it is known that an optimal perceptual device, even with a bounded memory, can accept unbounded left-branching and right-branching structures, though nested (hence ultimately self-embedding) structures go beyond its memory capacity. Notice that this effect is some aspects of performance not competence."

[5] As for the degree of grammaticalness, see Chomsky (1965) § 4.1.1. and Epstein (1990) for details.

interpretation depends on lexico-semantic information, only syntax or parsers and if there is a mismatch between syntactic level and the lexico-semantic level, the lexico-semantic information is selected. In this case, (6a) and (6b) are interpreted to (6c), which is plausible to think.

(6) a. The mouse was eaten by the cheese.
b. The cheese ate the mouse.
c. The mouse ate the cheese.

In this section, we discussed the distinction between grammaticality and acceptability. The grammaticality depends on the theory of grammar, while acceptability is based on speaker's intuition.

4. Toward the simplest system of narrow syntax

As we saw in the previous section, there is a distinction between grammaticalness and acceptability. My claim is here: If syntax, the sub-component of the theory of grammar, concerns about the selection, it also concerns about acceptability, which results in crash-proof grammar. However, under the minimalist program, crash-proof approach is undesirable since it brings a lot of constraint in the syntax, particularly constraints on Merge.

Selection has been discussed in the (pre-)government and binding theory(GB)'s literature that (i) it is not subsumed in the government relation, (ii) it is controversial whether subcaterization/c(ategorial)-selection is necessary or not. As for the first point, given that X'-theory forms the D-structure, I assume tentatively that the definition of the government as below.[6]

(7) Government
A governs B iff A and B mutually m-command each other.
(8) m-command
A m-commands B iff
a. A does not dominate B and B does not dominate A and
b. The min max projection that dominates A also dominates B.

For example, the example (9) shows that the verb 'think' selects a complementizer 'that'. The verb 'think' has the subcategorization frame such as 'think [__that]', which is not a government relation since 'that' does not m-command the verb 'think' because of the min max projection CP.

[6] This formulation is much more informal and I won't go into the detail of the notion of government or barriers.

(9) I [$_{VP}$ think [$_{CP}$ that John will win.

In general, the thematic roles, such as AGENT or PATIENT are realized as NP. This means that we do not need to specify the category of the thematic role. Thus, it is argued that we need s(emantic)-selection, not c(ategorial)-selection since we can predict the category from s-selection. In this case (9), the verb specifies not only the category CP, but also the type of sentence such as 'that-clause'. C-selection is not enough to capture the other cases: c-selection can select 'whether' or 'if', which results in undesirable(i.e., we need to see the lexical item in the maximal projection, which is head to head relation) and as we saw above, s-selection is predictable from c-selection (canonical structural realization (CSR)). However, consider (10)-(11).

(10) a. I asked what the time was
b. I inquired what the time was
(11) a. I asked the time
b. *I inquired the time

According to Grimshaw (1981), s-selection properties show that an s-selector selects such as 'question'(Q) and it must be realized as CP, not NP. Thus, (10a-b), (11b) is predictable from CSR. However, (11a) cannot be explained by CSR, which leads to the conclusion that c-selection should be remained as Grimshaw (1981) argues.

However, Pesetsky's (1982) argues that we can eliminate c-selection once we assume Case theory. In the example of (11), CSR of 'Q' and exclamation 'E' are both CP and NP, but only the verb that can assign case to its complement can licenses the case of NP. Put differently, in the example of (11), 'ask' can assign accusative case to the NP, but 'inquire' does not have the ability to case-assigning. Thus, the interaction of the CSR and case theory explain the example (11).

In the advent of the minimalist program, the early model of MP shows that the operation applies to local relation driven by features. The satisfaction of the requirement of uninterpretable features must be as soon as possible. The delay of the satisfaction of the strong-feature (or 'virus' features) must induce the derivation crash, so it must be done at syntax. The idea behind this is to derive economical derivation and optimal derivation once we try to explain language as a perfect system (i.e., Strong minimalist thesis: Language is an optimal solution to interfaces.).[7] Derivations are compared with each other and the fewer steps and the most economical derivation was grammatical sentences. This kind of approach

[7] See Chomsky (2000, 2010) for relevant discussions.

also leads to the idea of 'crash-proof', which forms only grammatical sentences. In Chomsky (2000: 132), it is argued that derivation must be satisfied the condition below.

(12) Properties of the probe /selector α must be satisfied before new elements of the lexical subarray are accessed to drive further operations.

In the same line, Collins (2002: 46) proposes the Locus Principle as below.

(13) Let X be a lexical item that has one or more probe/ selectors. Suppose X is chosen from the lexical array and introduced into the derivation. Then the probe /selectors of X must be satisfied before any new unsaturated lexical items are chosen from the lexical array. Let us call X the locus of the derivation.

Under Bare Phrase Structure(BPS), the structure building operation is Merge which takes two syntactic objects are combines into a set. In Chomsky (1995), it is formulated as Merge (α, β) = K, where K is {γ, {α, β}}, γ is the label of the set {α, β} and it is either α or β. Collins (2002) takes seriously that if we merge, say, α and β, there is no projection in it ({α, β}). Suppose that we cannot appeal to the label, selection also has to satisfy without seeing the labels. Furthermore, he proposes that selection is subject to the Chomsky's (2000) AGREE system, which he calls *Subcat*.

(14) Let X (a lexical item) be in SO1, and let Y (a lexical item or feature) be in SO2, where SO1 and SO2 are two independent syntactic objects (lexical items or phrases). If X and Y enter into a Subcat relation, then Merge (SO1, SO2).

The same line of this approach can be seen in Pesetsky and Torrego's (2006) Vehicle Requirement on Merge (VRM), Wurmbrand's (2014) Merge Condition, which are constraints on Merge. Namely, these are all the same line with 'crash-proof' syntax.

However, Chomsky (2004:132) notes that,

> "Elimination of s-selection has a number of consequences: it undermines at least the simplest way of predicting labels (as in MI), requiring a restatement in terms of SEM(H). It also entails that derivations cannot be failure-proof ("crash-free"), thus undermining the strong version of I[nterface]C[ondition] mentioned earlier. But these cannot be reasons for insisting on s-selection, features, plainly. We have to find other ways to determine labels, or show that they are dispensable. And IC must be weakened."

Assuming that narrow syntax does not depend on semantic properties and θ-roles are determined at the interface (such as the Hale-Keyser's theta theory), Chomsky (2004) argues that free-Merge generates whatever it does. In this, sense, grammar cannot be 'crash-proof'. Merge itself is free from features (selection-features, φ-features, edge-feature, etc), and there is no constraint on Merge. Move is reinterpreted as an instance of Merge (namely, internal Merge), which means that movement is also not driven by any features such as edge features. As for selection, Chomsky (2004) illustrates the example of external argument, which shows that if we have some feature which is relevant to θ-role, say, feature F, this must be satisfied at spec-head relation. This spec-head relation causes local problem since it's government relation, which is unformulable in minimalist syntax. Furthermore, external argument is generated on the edge of the phase, so the derivation must be move on before it is external merged in to the edge.

Going back to the Collins' (2002) Locus principle, the notion does not refer to the labels, but it refers to head and it refers to the semantic properties which is interpreted at interface. My claim is here: if Merge itself is not sensitive to what we Merge (because we can merge anything), there should not be rules such as referring to the features on the syntactic objects which are the candidate for the Merge. The targets for Merge is only syntactic object ('heads' or 'phrases'), that is, Merge can form either {H, XP} or {XP, YP} structures, that is, there is no constraint on Merge. The next question is how about agreement? The agreement system is a search algorithm that applies to the output of the Merge, so that it's not problematic with respect to Merge. And this process does not refer to the label, but refers to the features of the heads (e.g., features of gender/person/number). At this point, contra to Collins' (2002) Locus principle, based on Agree operation, the *Subcat* and AGREE are different and *Subcat* cannot be applied at narrow syntax because Merge comes for free. AGREE system refers only to the output of Merge.

The further question is, "can AGREE system apply within narrow syntax?" According to Chomsky (2013), there is a labelling algorithm (an instance of minimal search) which minimally detects the head of its syntactic objects. The search procedure does not form or attach "label" to the objects, but finds the prominent element, so that interfaces can read off what the syntactic objects are. AGREE system might be subsumed in this system, which means that agreement might be done by minimal search. How about valuation process and case assigning? This must be needed to be examined in the further researches.

5. Conclusion

In this paper, I reviewed the notion of grammar, the distinction between grammaticality and acceptability. I also argued that narrow syntax should not be 'crash-proof' and has to be free from selection. The next question should be on AGREE system, or in more general, syntactic relations in NS in that how can we simplify the system under the only Merge hypothesis.

Bibliography

Chomsky, N. (1956). Three models for the description of language. *IRE Trans. Inform. Theory*. 113-124.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

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

Chomsky, N. (2004). Beyond explanatory adequacy. In A, Belletti (Ed.), *Structures and beyond: The cartography of syntactic structures, vol.3*, 104-131. Oxford: Oxford University Press.

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

Chomsky, N. (2012). *The science of language*. Cambridge University Press.

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

Chomsky, N. (2017a). Generative Linguistics in the 21st Century: The Evidence and the Rhetoric. Talk given at University of Reading, UK, May 11.

Available at (https://www.facebook.com/theuniversityofreading/)

Chomsky, N. (2017b). The pre-history of simplicity of grammar. Talk given at MIT workshop on simplicity in generative learning, Cambridge, MA: MIT, September 23.

Available at (https://simplicity2017.mit.edu/program)

Collins, C. (2002). Eliminating Labels. In S. D. Epstein & T. D. Seely. (Eds.), *Derivation and explanation in the minimalist program*, Blackwell Publisher. 42-64.

Embick, D &David, P. (2015). Towards a computational(ist) neurobiology of language: correlational, integrated and explanatory neurolinguistics. *Language, cognition and neuroscience*, *30*(4), 357-366.

Epstein, S. D. (1990). Differentiation and reduction in syntactic theory: A case study. *Natural Language and Linguistic Theory*, *8*, 313-323.

Epstein, S. D., Kitahara, H. & Seely, T. D. (2015). *Explorations in maximizing syntactic minimization*. Routledge.

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

Friederici, A. D., Chomsky, N., Berwick, R. C., Moro, A. & Bolhuis, J. J. (2017). Language, mind and brain. *Nature human behaviour* (DOI: 10.1038/s41562-017-0184-4).

Grimshaw, J. (1979). Complement selection and the lexicon. *Linguistic inquiry, 10*(2), 279-326.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of Visual information*. New York: Freeman.

Miller, G. (2003). The cognitive revolution: a historical perspective. *TRENDS in Cognitive Sciences, 7*(3), 141-144.

Miller, G & Chomsky, N. (1963). Finitary models of language user. In R.D. Luce, R. Bush & E. Galanter (Eds.), *Handbook of Mathematical Psychology*, *II*, New York: Wiley. 419-492.

Ott, D. (2010). Grammaticality, interfaces, and UG. In M. T. Putnam. (Eds.), *Exploring crash-proof grammar*, John Benjamins publishing company. 89-104.

Pesetsky, D. & Torrego, E. (2006) Probes, goals and syntactic categories. In Y, Otsu (Ed.), *Proceedings of the 7th Tokyo Conference on Psycholinguistic*, Tokyo: Hituzi Syobo. 25-60.

Seely, T. D. (2006). Merge, derivational c-command, and the subcategorization in a label-free syntax. In C, Boeckx (Ed.), *Minimalist essays*, John Benjamins Publishing company. 182-217.

Traxler, J. M. (2012). *Introduction to psycholinguistics*. Wiley-Blackwell.

Watumull, J., Hauser, M. D., Roberts, I. G., & Hornstein, N. (2014). On recursion. *Frontiers in psychology* (DOI: 10.3389/fpsyg.2013.01017).