

HOW TO TRIGGER ELEMENTS OF I-LANGUAGES

DAVID W. LIGHTFOOT
Georgetown University

1 Global Approaches to Acquisition¹

In *Syntactic structures* Chomsky viewed a grammar as a device to generate the sentences of a language (1957: 11). Linguists could establish what the grammar might be through a metric EVALUATING its success in generating a corpus in an optimal way rather than by following procedures to DISCOVER a grammar or to DECIDE whether a grammar was correct or not for a given corpus (1957: ch. 6). In this respect Chomsky was reflecting the anti-positivist moves of the 1950's (e.g. Popper 1959), showing that there is no predetermined path to scientific truth and that the best we can do is say which analyses are better for certain data. He argued against formal approaches that could not capture properties of natural languages (Markov processes and pure phrase structure grammars) and in favor of derivations relating different levels of abstraction (transformations).

Syntactic structures contains nothing about cognition, psychology, or the acquisition of language by children, nor does the long, influential review by Robert Lees (1957), apart from the mystified optimism of his final section on 'learning theory.' We know that as Chomsky wrote *Syntactic structures*, he was preparing to write his 1959 attack on Skinner's behaviorist approach to analyzing the human language capacity and he was interacting with Eric Lenneberg, who was to pursue the biological approach to language earliest and furthest (Lenneberg 1967). Several years later, in the introduction to the published version of his dissertation, he wrote that he had viewed his approach to linguists justifying their grammars through an evaluation metric as having 'a psychological analog' in what a child does in acquiring a grammar in the first few years of life (1975: 11, 12, 36, etc.): children also use an evaluation metric, but subconsciously. However, he had viewed the psychological analog and discussion of cognition and biology as 'too audacious' for the time (1975: 35). For discussion, see my introduction to the 2002 second edition of *Syntactic structures*.

¹ Many thanks for helpful comments on an earlier draft from some of my favorite readers: Cedric Boeckx, Elan Dresher, Tony Kroch, Anne Lobeck, Betty Tuller, and Marit Westergaard.

By 1965, either Chomsky had become more audacious or the field had shifted sufficiently to allow him to discuss the cognitive and biological consequences of *Syntactic structures*. Given that Chomsky had been thinking along these lines from the beginning, there was an easy translation of the methodological claims of the earliest work into substantive claims about human cognition and biology. The famous first chapter of *Aspects* lays it out carefully and makes matters of children's acquisition central for linguists constructing their grammars. If grammars are learnable by children and meet other biological requirements, then there are great empirical demands that must be met, narrowing the range of viable hypotheses and thereby helping linguists to converge on plausible hypotheses and children to attain their mature system. In particular, we must show how children can "project" from the Primary Linguistic Data (PLD) that they experience to the mature grammars that they eventually attain, solving the profound "poverty-of-stimulus" problems. Chomsky viewed children as following an evaluation metric preferring certain grammars over others with respect to a given corpus of sentences.

Here I aim to tease apart the methodological goal of justifying analyses and the psychological goal of characterizing how children acquire their system and to suggest a different approach to childhood acquisition, which became viable after another major methodological move made by Chomsky twenty years after *Aspects*.

In *Aspects*, Chomsky wrote:

Certain problems of linguistic theory have been formulated as questions about the construction of a hypothetical language-acquisition device ... We may think of the theorist as given an empirical pairing of collections of primary linguistic data associated with grammars that are constructed by the device on the basis of such data. Much information can be obtained about both the primary data that constitute the input and the grammar that is the "output" of such a device, and the theorist has the problem of determining the intrinsic properties of a device capable of mediating this input-output relation. (Chomsky 1965: 47)

Chomsky writes here of 'collections of PLD' being 'associated with grammars' and a common approach to children evaluating their grammars has been GLOBAL: children evaluate postulated grammars as wholes against the whole corpus of PLD that a child encounters, checking which grammars generate which data.

For example, Robin Clark's genetic algorithm (1992) employs his Fitness Metric to measure very precisely the fitness of grammars in generating sets of sentences. The key idea is that certain grammars yield an understanding of certain sentences and not others; put differently, they generate certain sentences and not others. The Fitness Metric quantifies the failure of grammars to parse sentences by counting the "violations," sentences experienced that cannot be generated by the grammar being evaluated. There are two other factors involved in his Fitness equation, a superset penalty and an elegance measure, but those factors are subject to a scaling condition and play a minor role, which I ignore here. The Fitness Metric remains the most sophisticated and fully worked out evaluation measure that I know. It is a global measure of success, assigning indices to whole, fully formed grammars against a whole corpus of sentences.²

² It is worth noting that that the metric is technically and conceptually flawed insofar as it is based on an assumption that grammars with a greater number of parameters set correctly will be the fittest, the most successful in parsing/generating incoming data but Dresher (1999: 54-58) demonstrated that this assumption is false, that there is

Gibson & Wexler (1994) take a different approach but their children also effectively evaluate whole grammars against whole sets of sentences, although they react to local particularities; their children are ‘error-driven.’ They acquire a mature grammar eventually by using a hypothetical grammar (a collection of parameter settings) and revising it when they encounter a sentence that their current grammar cannot generate, an “error.” In that event, children follow the Triggering Learning Algorithm to pick another parameter setting and continue until they converge on a grammar for which there are no unparseable PLD, no errors. Gibson & Wexler used a toy system of three binary parameters that define eight possible grammars, each of which generates a set of sentence types. The child converges on a whole grammar in light of the complete set of sentence types experienced, essentially testing the generative capacity of grammars against a corpus of sentences.

There are major feasibility problems for both these global, evaluation-based approaches to language acquisition, evaluating whole grammars against comprehensive sets of sentences experienced (Lightfoot 2006: 76f). If there are thirty points of binary choices (a conservative estimate), there are over a billion grammars to evaluate against the data set; if there are forty points of variation, then over a trillion grammars to evaluate; if there are fifty points of variation, then the numbers become astronomical. And bear in mind that each grammar generates an infinite set of sentences and structural descriptions. In addition, in order to check whether the generative capacity of a grammar matches what the child has heard, s/he will need to retain a memory in some fashion of everything that has been heard. Such an approach to language acquisition faces great difficulties when one thinks beyond toy systems with just a few parameter settings.

Alongside the feasibility issues, one’s theory must accommodate change in I-languages from one generation to another. Seeking to explain changes through language acquisition requires a different approach to language acquisition. If children attain grammars by evaluating them against a corpus of data, then they would need to be confronted with the data generated by the new grammar in order to first select the grammar that generates them. This introduces problems of circularity: what comes first, the new grammar to generate the new data or new data that require the child to select the new grammar?

2 Children Discovering Elements of I-languages

A different approach would separate the methodological and psychological components. One would acknowledge the validity of Chomsky’s 1957 arguments that linguists need to evaluate the success of their hypothesized grammars rather than to seek a discovery procedure leading to correct hypotheses or a means to decide whether a hypothesis is correct or not. However, one might view children differently, acquiring their mature grammar by following a kind of subconscious discovery procedure, albeit not the kind of discovery procedure contemplated by Chomsky (1957) but one incorporating a rich theory of Universal Grammar (UG).

In the early work, Chomsky had taken children to be acquiring grammars that generated socially defined languages like English, Turkish, or Warlpiri and he postulated ‘an ideal speaker-

no smooth correlation between accuracy in what is parsed/generated and the number of parameters set correctly. See Lightfoot (2006: 74) for discussion.

listener, in a completely homogeneous speech-community, who knows its language perfectly' (1965: 3), idealizing away from individual variation.

In 1986, he made a major methodological move, abandoning the idealized speaker-listener and the idea of socially defined languages as having any kind of psychological reality. He embraced language variation and distinguished between external E-language and internal, individual, intensional I-languages as the proper focus of linguistic theory. In effect, he adopted Wilhelm von Humboldt's (1836) distinction between the languages of nations and those of individuals. Hermann Paul (1877: 325) emphasized the individual and biological view of language, noting in an early work 'dass die reelle Sprache nur im Individuum existiert' ('real language exists only in individuals'). Later, Paul (1880: 31) attacked the group psychology of Lazarus and Steinthal and wrote that 'Wir müssen eigentlich so viele Sprachen unterscheiden als es Individuen gibt' ('we must in fact distinguish as many languages as there are individuals').

Chomsky (1986) followed von Humboldt and Paul and distinguished external E-language and internal, individual I-languages. E-language refers to language out there in the world, the kind of thing that a child might be exposed to, an amorphous, mass concept. I-language, on the other hand, refers to a biological SYSTEM that grows in a child's mind/brain in the first few years of life and characterizes that individual's linguistic capacity. It consists of structures, categories, morphemes, phonemes, and features, and computational operations that copy items, delete them, assign indices to them, etc. One's I-language is a private object that permits thought and partial communication with certain other speakers, although it may differ from the I-languages of those speakers.

Everything was now individualized and linguists were free to postulate different systems for different individuals, allowing the possibility of billions of grammars or, now, "I-languages." Indeed, a little later, Kroch (1989) postulated the idea that individuals may use multiple co-existing grammars, which turned out to be very productive for work on syntactic change and opened the possibility of billions and billions of grammars, which could be studied through the methods of population biology.

These ideas of E-language and I-languages suffice for the purposes of accounting for language acquisition and we do not need the conventional, socially defined English or Estonian. One can view children as acquiring their individual I-language on exposure to external E-language. Rather than evaluating systems against a set of data, children can be viewed as paying no attention to what any I-language or grammar generates but instead growing an I-language by identifying and acquiring its elements (Lightfoot 2006). Children parse the E-language they hear and discover the categories and structures needed to understand what they hear, thereby accumulating the elements of their I-language (for this "cue-based acquisition," see Dresher 1999, Lightfoot 1999, and Sakas & Fodor 2001).³

UG provides children with the structures that they might need in order to understand and parse their ambient E-language; robust E-language triggers them. Children are born to parse and at a certain stage of development, after they know that *cat* is a noun referring to a domestic feline and *sit* is an intransitive verb, they may hear an expression *The cat sat on the mat* and recognize that it contains a Determiner Phrase (DP) consisting of a determiner *the* and a noun *cat* and a Verb Phrase (VP) containing an inflected verb *sat* (V+I) followed by a Preposition Phrase (PP) *on the mat*, i.e. $_{VP}[V+I PP]$. The child makes use of the structures needed to parse what is heard (i.e. the structures "expressed" by the E-language experienced) and, once a structure is used, it is incorporated into the emerging I-language. In this way, a child discovers and accumulates the

³ The "cues" terminology has given rise to confusions and misunderstanding and I drop it here.

elements of its I-language, which are required to parse the E-language around them; children acquire elements of their I-language in piecemeal fashion.

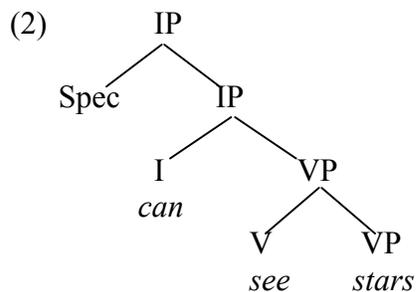
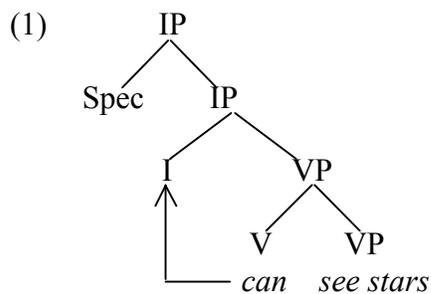
At no stage does the child calculate what its current I-language can generate; rather s/he simply accumulates the necessary structures and the resulting I-language generates what it generates. Furthermore, if UG makes available a thousand possible structures for children to draw from, that raises no intractable feasibility problems comparable to those facing a child evaluating the generative capacity of grammars with thirty possible parameter settings, checking the grammars against what has been heard. It involves no elaborate calculations. Children developing some form of English I-language learn without apparent difficulty irregular past tense and plural forms for a few hundred verbs and nouns. Learning that there is a structure $v_P[V+I PP]$ (see previous paragraph) seems to be broadly a similar kind of learning, similar to acquiring the irregular past tense forms, although much remains to be said.⁴

Under this approach to language acquisition, children pay no attention to the generative capacity of their emerging I-language but they attain particular elements of it step-by-step (see Dresher 1999 on the resulting ‘learning path’) and they do that by experiencing particular elements of E-language that they must attempt to parse. Work seeking to explain diachronic changes through acquisition has enabled us to link changes in E-language to changes in particular elements of I-languages, giving us a clear idea of what triggers what in some cases. I will illustrate this by sketching briefly a phase transition in the history of English that is now well understood.

3 Identifying Triggers

There is good reason to believe that English I-languages underwent a change whereby words like *can*, *could*, *must*, *may*, *might*, *will*, *would*, *shall*, *should* and *do* were once categorized as more or less normal verbs but then were re-categorized as Inflection elements in all known grammars of English speakers after the time of Sir Thomas More in the early sixteenth century. For More and his forebears, verbs like *can* moved to a higher Inflection position, as in (1), and after More they were generated directly as Inflection elements and occurred in structures like (2), a single shift in the system, which was manifested by the simultaneous loss of several phenomena, the phase transition. Early speakers had used expressions like *He has could see stars*, *Canning see stars ...*, *She wanted to can see stars*, *She will can see stars*, and *She can music*. Such sentences may be generated by an I-language with structures like (1), and not by I-languages with structures like (2). The SINGULARITY of the change in I-languages explains the PARALLELISM in the loss of these phenomena (for details, see Lightfoot 2006, in press a, b).

⁴ These structures may affect many sentences or only a few. Tony Kroch notes that in modern English we have remnants of verb-second in *Kim left and so did Jim*. Such examples look like irregular verbs, residues of once productive processes.



A critical property of this change is that it consisted entirely in the LOSS of phenomena and there were no new forms emerging. Since children converge on their I-language in response to ambient simple expressions and not in response to “negative data” about what does not occur, the new, less diverse data need to be explained by a new abstract system that fails to generate the obsolescent phenomena. There were no new forms in which the modal auxiliaries began to occur and so the trigger for the new system must lie elsewhere. In this case, the new PLD cannot be the new output of the new grammars, not only because that would be circular but because, in any case, there are no new forms. Changes like this, which consist only in the loss of expressions, make up a kind of poverty-of-stimulus argument for diachrony: there appear to be no new forms in the PLD that directly trigger the loss of those expressions.

If we ask why this or any other I-language change happened, there can only be one answer under an approach explaining changes through acquisition: children came to have different PLD as a result of a prior change in EXTERNAL language. We have a good hypothesis about what the prior E-language change was in this case.

Early English had complex morphological properties. For example, we find *fremme*, *fremst*, *fremþ*, *fremmaþ* in the present tense and *fremed*, *fremedest*, *fremede*, *fremedon* in the past tense of ‘do;’ *sēo*, *siehst*, *siehþ*, *sēoþ* in the present tense for ‘see;’ *rīde*, *rītst*, *rītt*, *rīdaþ* for the present tense of ‘ride,’ and *rād*, *ride*, *rād* and *ridon* for the past tense. There was a massive loss of verb morphology in Middle English, beginning in the north of England and due to intimate contact with Scandinavian speakers, leaving the third person singular *-s* ending as the only surviving element of present tense verb morphology. Again I skip interesting details but the crucial factor is that external language that children heard changed such that the modern modal auxiliaries like *can*, *shall*, etc came to be morphologically distinct from other verbs; as members of the small preterite-present class, they always lacked the one surviving feature of present tense verb morphology, the *-s* ending of the third person singular. Furthermore, their “past tense” forms (*could*, *would*, *might*, etc) had meanings that were not past time, reflecting old subjunctive uses (3). The evidence indicates that these modal verbs were re-categorized in people’s internal

systems, because they had become formally distinct from other verbs as a result of the radical simplification of morphology.⁵ So we see domino effects: changes in what children heard, the newly reduced verb morphology, led to a different categorization of certain verbs, which yielded systems (2) that were not compatible with the obsolescent data.

(3) They might/could/would leave tomorrow.

More was the last known speaker with the old system. For a period, both systems co-existed: some speakers had (1) and others had (2), the former becoming rarer over time, the latter more numerous. A large literature is now devoted to this kind of sociological variation, changing over time, and the spread of the new system can be studied through the methods of population biology (Lightfoot in press, b). As some speakers began to have new I-languages, that affected the ambient E-language, making it more likely that members of a speech community would acquire the new I-language. Partha Niyogi (2006) provided interesting computational simulations of this approach to language acquisition and diachronic change.

An old idea holds that one can sometimes discover properties of systems by seeing how they change. The major contributions of diachronic work in syntax lie in explaining one kind of variation, due to co-existing I-languages, and in revealing what the E-language trigger might be for any particular property of I-languages. In particular, by looking at change one can sometimes generate productive hypotheses about which PLD trigger which elements of I-languages, associating particular PLD drawn from E-language with particular elements of emerging I-languages. From a certain perspective, one may paraphrase Theodosius Dobzhansky: nothing in (language-specific) syntax makes sense except in the context of change – that is, ideally for any language-specific syntactic property, one would show how it might have emerged, just as ‘nothing in biology makes sense except in the light of evolution’ (Dobzhansky 1973).

It is important to recognize that the same elements of I-language may be triggered by different PLD in different languages. For example, we have seen that the new Inflection category in English I-languages was triggered by new properties of verb morphology. O’Shannessy (2013) reports a new language also with an Inflection category but triggered by quite different features of E-language having to do with irrealis properties. Similarly, children developing an English I-language could learn that VP’s have V-complement order from a simple sentence like *Bent visited Oslo*; since verbs do not move in English I-languages and Inflection lowers on to verbs, *visited Oslo* can only be analyzed as $_{VP}[\text{visit}+I \text{ Oslo}]$. A Norwegian child, however, could not draw the same conclusion from the word-for-word translation *Bent besøkte Oslo*, which does not reveal the underlying position of the verb. Norwegian I-languages are verb-second: finite verbs move to a high, “second” position (presumably in the C projection), yielding simple structures like $_{CP}[\text{Bent } c\text{besøkte } _{VP}[\text{besøkte Oslo}]]$. The verb-second analysis is required by synonymous sentences like *Oslo besøkte Bent*, where the finite verb surfaces to the left of the subject DP and requires a structure $_{CP}[\text{Oslo } c\text{besøkte } _{IP}[\text{Bent } _{VP}[\text{besøkte Oslo}]]]$, as well as “topicalized” expressions like $_{CP}[\text{Søndager } c\text{besøkte } _{IP}[\text{Bent } _{VP}[\text{besøkte Oslo}]]]$ ‘On Sundays Bent visited Oslo.’ Therefore, *Bent besøkte Oslo* does not reveal the structure of the VP and a more complex sentence like *Bent kan besøke Oslo* ‘Bent can visit Oslo’ is needed to express the $_{VP}[\text{V DP}]$ structure, *Bent kan _{VP}[\text{besøke Oslo}]*. Similarly, the complex *Bent kann Oslo besuchen*

⁵ The SYNTACTIC evidence for the verbal status of the modals is not directly affected by the new morphology. Therefore children must have ignored some of the input data if they re-categorized the modals; those data simply became obsolete.

reveals the ${}_{VP}[DP V]$ structure of German, another verb-second language but with complement-verb order.

4 Minimalism and Acquisition

Lightfoot (1999) postulated that UG provide the structures ${}_I V$, ${}_{VP}[V DP]$, ${}_{VP}[DP V]$, and ${}_{SpecCP}[XP]$, available to children if their PLD triggered them. In addition, Lightfoot (2006: 82-86) hypothesized the eleven structures of (4) to capture a wide range of properties involving various syntactic structures, results of computational operations, binding relations, and intonation patterns. They presupposed a substantial contribution from UG (including a condition on deletion, a binding theory, and a bottom-up analysis of intonation structure, and more) and they were each triggered by the sentences on the right, which expressed or required the relevant structure (for details, see Lightfoot 2006).

(4) a. ${}_{CE}$	Peter said Kay left
b. ${}_{CP}[wh-]$	Who did Jay see?
c. ${}_{VE}$	Jay saw Ray and Jim Kim
d. ${}_{NP+I}$	Kim's happy
e. ${}_{VPE}$	Kim didn't
f. ${}_{VP}[VP \# DP]$	Jay gave to Ray his favorite racket
g. ${}_{DP}[DP D NP]$	Kim's hat
h. ${}_{IP}[DP_i V_{DP[-self]_i}]$	Kim hurt herself
i. ${}_{DP_i V_{DP}[DP'S_{NP}[NP \text{pronoun}_i]]]$	Kim _i saw Bill's pictures of her _i
j. ${}_{N}[X-1 N-2]$	BLACKboard
k. ${}_{XP}[2 1]$	black BOARD

Postulating such UG structures enabled us to solve a wide range of poverty-of-stimulus problems. The structures constituted rich and structurally complex information that would have evolved through genetic mutation. That is part of a tradition noted by Chomsky (2007: 2): 'At the time of the 1974 discussions, it seemed that [the faculty of language] must be rich, highly structured, and substantially unique.' There is nothing inherently untenable about that, given that genetic mutations hold at a biochemical level and the structures at a cognitive level and that we know next to nothing about how such levels comport. Indeed, if (4) is what is required to solve poverty-of-stimulus problems, then one celebrates the explanations achieved.

However, one always seeks better alternatives. Since *Aspects*, work in generative syntax has gone through a number of distinct paradigms. Beginning with *Aspects* and flourishing through the 1970's we had conditions on the expressive power of rules, then a model where principles of Government and Binding were central, followed soon by the Principles-and-Parameters approach focusing on language variation and on how the variation might be attained by children exposed only to simple, primary data. In fact, considerations of acquisition were central from the time of *Aspects* through the Principles-and-Parameters period.

Now we are in a paradigm of Minimalism, which is driven by quite different forces, seeking to minimize the information of the linguistic genotype and seeking simpler alternatives to the structures of (4). There is little substantive discussion of variation and even less of acquisition and, as Boeckx (2011: 207) puts it, 'central Minimalist tenets clash with the traditional

Principles-and-Parameters approach.’ Chomsky (2007: 4) writes

Throughout the modern history of generative grammar, the problem of determining the character of [the faculty of language] has been approached “from the top down”: How much must be attributed to UG to account for language acquisition? The [Minimalist Program] seeks to approach the problem “from the bottom up”: How little can be attributed to UG while still accounting for the variety of I-languages attained?

The goal of Minimalism is to see what information that has been apparently needed at UG can be explained in terms of more general principles or be shown to be unnecessary.

Proposals within the Minimalist Program provide severe, general, and sometimes elegant limits on the structures available for narrow syntax. Boeckx (2011: 210) advocates a ‘Strong Uniformity Thesis,’ whereby ‘principles of narrow syntax are not subject to parameterization’ and variation ‘would be confined to the margins of narrow syntax, especially the morpho-phonological component (PF).’ Along those lines, Gianollo, Guardiano, & Longobardi (2008) took variation seriously and explored the possibility of a theory of syntactic variation beyond narrow syntax in terms of the properties of functional features, which may be grammaticalized, checked by a lexical category, ‘spread on’ a lexical category, or checked by a strong feature. And, of course, there are many other proposals. By embracing the radical limits imposed by current Minimalist work, perhaps we could find how those limited structures are expressed in what children hear and then determine which structures are expressed robustly enough to be incorporated into children’s I-languages. We would abandon evaluation metrics and checking the generative capacity of I-languages against what is observed and instead seek a means by which children may DISCOVER the structural elements that are EXPRESSED (i.e. required) by the E-language that they are exposed to, tweaking *Aspects*’ global approach to acquisition to the local approach advocated here. The research agenda would be to seek to establish what particular aspects of PLD might trigger particular structural elements hypothesized by Minimalist work; the challenge would be to show that there are plausible PLD that could have the necessary triggering effects. In this way we could seek to resolve Boeckx’s ‘clash’ between Minimalist tenets and the traditional concern with acquisition.

In a similar spirit, Yang (2010) argues for the complementarity of structures provided by UG and statistical generalization. Structures need to be expressed by E-language robustly. Lightfoot (1999) sought to explore some effects in changes in robustness of expressions triggering elements of I-languages but it is important to remember that there is unlikely to be a uniform definition of robustness: some I-language elements are expressed with some frequency and others may be triggered by single events (such as the learning of certain vocabulary items).

I have argued here that one can sometimes identify aspects of E-language that trigger specific elements of I-languages. When such triggers can be identified, they may cast light on how the changing distribution of E-language properties trigger structural elements of new I-languages. Rather than abandoning any concern with explaining language acquisition as if it were too mysterious to explain, we might link the “local” approach to acquisition, seeking E-language triggers for specific elements of I-languages, with the range of structures allowed by our favorite Minimalist proposals. What is NOT a rational approach is to assert that pre-Minimalist ideas about UG are too complex to have evolved through genetic mutation and therefore must be discarded. Given the limitations on our understanding of the evolution of the language faculty (Hauser et al. 2014) and the richness of what we have learned about the acquisition of language,

that would be foolish until we have some kind of alternative offered.

References

- Boeckx, Cedric. 2011. Approaching parameters from below. In *The biolinguistic enterprise: New perspectives on the evolution and nature of the human language faculty*, ed. A.-M. Di Sciullo & C. Boeckx. Oxford: Oxford University Press, 205-221.
- Chomsky, Noam. 1957. *Syntactic structures* (2nd ed. 2002). Den Haag: Mouton.
- Chomsky, Noam. 1959. Review of B. F. Skinner *Verbal behavior*. *Language* 35.1: 26-57.
- Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 1975. *The logical structure of linguistic theory*. New York: Plenum.
- Chomsky, Noam. 1986. *Knowledge of language: Its nature, origin, and use*. New York: Praeger.
- Chomsky, Noam. 2007. Approaching UG from below. In *Interfaces + Recursion = Language? Chomsky's minimalism and the view from syntax-semantics*, ed. U. Sauerland & M. Gaertner. Berlin: Mouton de Gruyter, 1-30.
- Clark, Robin. 1992. The selection of syntactic knowledge. *Language Acquisition* 2: 83-149.
- Dobzhansky, Theodosius. (1973). Nothing makes sense in biology except in the light of evolution. *American Biology Teacher* 35: 125-129.
- Dresher, B. Elan. 1999. Charting the learning path: Cues to parameter setting. *Linguistic Inquiry* 30.1: 27-67.
- Gianollo, Chiara, Cristina Guardiano, & Giuseppe Longobardi. 2008. Three fundamental issues in parametric linguistics. In *The limits of syntactic variation*, ed. T. Biberauer. Amsterdam: Benjamins, 109-142.
- Gibson, Edward & Kenneth Wexler 1994. Triggers. *Linguistic Inquiry* 25.3: 407-454.
- Hauser, Marc D., Charles Yang, Robert C. Berwick, Ian Tattershall, Michael J. Ryan, Jeffrey Watumull, Noam Chomsky, & Richard C. Lewontin. 2014. The mystery of language evolution. *Frontiers of Psychology* 5: 401.
- Humboldt, Wilhelm von 1836. *Über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluss auf die geistige Entwicklung des Menschengeschlechts*. Royal Academy of Sciences of Berlin. [*Linguistic variability and intellectual development*, transl. G. C. Buck and F. A. Raven 1971. Philadelphia: University of Pennsylvania Press.]
- Kroch, Anthony. 1989. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1: 199-244.
- Ledgway, Adam & Ian G. Roberts (eds.). in press. *Cambridge handbook of historical syntax*. Cambridge: Cambridge University Press.
- Lees, Robert. 1957. Review of *Syntactic structures*. *Language* 33.3: 375-408.
- Lenneberg, Eric. 1967. *The biological foundations of language*. New York: John Wiley.
- Lightfoot, David W. 1999. *The development of language: Acquisition, change, and evolution*. Oxford: Blackwell.
- Lightfoot, David W. 2002. Introduction to second edition of *Syntactic structures*. Berlin: Mouton de Gruyter.
- Lightfoot, David W. 2006. *How new languages emerge*. Cambridge: Cambridge University Press.
- Lightfoot, David W. in press a. Acquisition and learnability. In Ledgway & Roberts in press.

- Lightfoot, David W. in press b. Imperfect transmission and discontinuity. In Ledgway & Roberts in press.
- Niyogi, Partha. 2006. *The computational nature of language learning and evolution*. Cambridge, MA: MIT Press.
- O'Shannessy, Carmel. 2013. The role of multiple sources in the formation of an innovative auxiliary category in Light Warlpiri, a new Australian mixed language. *Language* 89.3: 328-353.
- Paul, Hermann. 1877. Die Vocale der Flexions- und Ableitungssilben in den ältesten germanischen Dialecten. *Beiträge zur Geschichte der deutschen Sprache und Literatur* 4: 314-475.
- Paul, Hermann. 1880. *Prinzipien der Sprachgeschichte*. Tübingen: Niemeyer.
- Popper, Karl. 1959. *The logic of scientific discovery*. London: Hutchinson.
- Sakas, William & Janet D. Fodor 2001. The structural triggers learner. In *Language acquisition and learnability*, ed. S. Bertolo. Cambridge: Cambridge University Press, 172-233.
- Yang, Charles. 2010. Three factors in language variation. *Lingua* 120.5: 1160-1177.

