



UNIVERSITAT DE
BARCELONA

Grau d'Estudis Anglesos

Treball de Fi de Grau

Curs 2018-2019

IS CHOMSKYAN LINGUISTICS TENABLE?

A WAR OF ATTRITION

NOM DE L'ESTUDIANT: Gonzalo Bermejo Miranda

NOM DEL TUTOR: Roger Gilabert

Barcelona, 11 de juny de 2019

Acknowledgements

Whatever else is unsure in this stinking dunghill of a world a mother's love is not. Your mother brings you into the world, carries you first in her body. What do we know about how she feels? But whatever she feels, it, at least, must be real.

James Joyce

I remember, the Players have often mentioned it as an honour to Shakespeare, that in his writing, (whatsoever he penn'd) hee never blotted out line. My answer hath beene, would he had blotted a thousand.

Ben Jonson

Abstract

Since the late 1950's and up to the last ten years, Chomskyan or Generativist linguistics and its view of language as driven by abstract syntactic mechanisms—*i.e.* Universal Grammar—has dominated the field of Linguistics with an iron fist. But, are those syntactic abstractions real? Is the empirical evidence behind them uncontroversial? This essay answers in the negative to both those questions. It is a tour that will move in space—from Harvard to Tokyo to the Amazonian rainforest—in time—from Old English to Modern English—as well as through different species—from humans to birds and apes. I will review evidence from first and second language acquisition, I will look at the implications of the latest research program in Generativism: the Minimalist Program, and it will end with a humbling trip into the language skills of our nearest primate cousins.

Keywords: Generativism, syntax, rules, recursion.

Desde finales de los años 50 y hasta los últimos diez años, la lingüística Chomskiana o generativista y su visión de un lenguaje controlado por mecanismos sintácticos abstractos, englobados en la Gramática Universal, ha dominado el campo de la lingüística con mano de hierro. Pero, ¿son esos mecanismos abstractos reales? ¿están demostrados empíricamente de forma irrefutable? Este ensayo responde con un no ambas preguntas. Es un viaje que nos llevará en el espacio desde Harvard a Tokio e incluso a la selva amazónica, en el tiempo desde el inglés antiguo hasta el moderno, así como a través de distintas especies. Examinaremos estudios sobre la adquisición de la lengua materna y de las segundas lenguas, también las implicaciones de la última tendencia generativista, el programa minimalista, y acabaremos con una breve incursión en las capacidades lingüísticas de los primates no humanos.

Palabras clave: Generativismo, sintaxis, reglas, recursividad.

Table of Contents

INTRODUCTION: Syntax.exe?	1
CHAPTER 1 Overruled	
1.1 The original Wug Test.....	3
1.2 The Japanese Wug Test.....	6
CHAPTER 2 Found in Translation	
2.1 What do you mean by <i>he</i> ?	9
2.2 Gulliver’s transfers.....	11
CHAPTER III Drought or Flood	
3.1 The Poverty of the Stimulus.....	15
3.2 Secondary Roads.....	19
3.3 Philosophical and epistemological obstacles.....	21
CHAPTER IV UG loses weight	
4.1 The Minimalist Program.....	23
4.2 Without recursion.....	25
4.3 Recursion reloaded.....	27
4.3.1 Recursion is in the eye of the beholder	27
4.3.2 English walks. Pirahã hops.....	28
4.3.3 Recursion on the fly.....	28
CHAPTER V Out of the Blue?	
5.1 In the beginning was the word.....	30
5.2 Not Only Us.....	32
5.2.1 Matata and Kanzi	33
5.2.2 Sherman and Austin.....	36
ENVOY	38
WORKS CITED	39

Introduction: Syntax.exe?

I think that we are forced to conclude that grammar is autonomous and independent of meaning

Noam Chomsky, *Syntactic Structures*.

Chomskyan generative linguistics bestows a position of privilege on syntax. Syntax is considered autonomous from all other levels of language such as phonology or pragmatics, and of course, meaning. Syntax is, in Hamlet's words, *unmixed with baser matter*.

This position is held even though “empirical results culled from psycholinguistic and corpus research suggest that syntax, semantics, suprasegmental phonology, and the lexicon are closely linked to one another” (Hilferty, 2003, p. 26).

A syntacto-centric¹ view of language constrains a Chomskyan researcher to:

- a) a mind-as-a-computer approach towards language development in infants
- b) a sudden appearance of the linguistic capacity in modern humans, and
- c) the impossibility of there being any antecedents of language—i.e. protolanguage—even in our genetically nearest cousin, the bonobo.

Syntax is not, however, the Rosetta stone of language. My hope is that, to quote Sir Arthur Conan Doyle in *The Sign of the Four*: “when you [linguists] have eliminated the impossible, whatever remains, however improbable, must be the truth”.

The aim of this paper is to go over the main evidence Chomskyan scholars typically adduce in favour of the autonomy-of-syntax thesis, and Sherlock-like, eliminate the impossible.

In the first section, I will look closely into the well-known Wug Test and how children, both English and Japanese, do *not* really pass it, as it is traditionally assumed. In the second, I will look at evidence from the field of Second Language Acquisition. Next, I will inspect the argument from the poverty of the stimulus, how there is actually no such poverty and why it is posited. In the fourth section, I will deal with the consequences of reducing the language faculty to mere recursion. In the fifth chapter, I will review Chomsky's take on the evolution of language. Finally, I will look at some groundbreaking results in Ape Language Research. Of course, one could simply take aim at the premise of the independence of syntax from meaning² by simply posing examples illustrating the unfeasibility of such an independence.

¹ Expression coined by Ray Jackendoff in his 1997 book *The Architecture of the Language Faculty*.

² For a detailed treatment of the entanglement between syntax and other linguistic dimensions, see Hilferty (2003).

For example, these two sentences from Randy Allen Harris's book, *The Linguistics Wars*:

- (1) Everyone on Cormorant Island speaks two languages
- (2) Two languages are spoken by everyone on Cormorant Island

Should not these two sentences mean the same?

From a Chomskyan point of view, they should, since they share the same deep structure³; unfortunately, they do not necessarily share the same meaning. Sentence (1) favours the interpretation that everybody on the island is (at least) bilingual but does not say anything about what those two languages are. Its passive counterpart in (2), on the other hand, seems to suggest that the same two languages are spoken by everyone on the island. Each sentence has a different theme and focus—*everyone* vs. *two languages*—, and consequently, different entailments. Even an innocent syntactic transformation like *passive* gets tainted by semantics. Now, let us look at another instance of the interplay of syntax and semantics:

- (3) John is as stubborn as the Rock of Gibraltar.
- (4) *John is as stubborn as the Rock of Gibraltar is.

If syntax and meaning were not entangled, sentence (4) should partake of the grammaticality of (3) since, from the point of view of syntax, they are equivalent. Instead, making the copula overt forces upon the sentence an impossible prosopopoeia of the promontory, which renders the sentence ungrammatical.

We can see the connection between prosody and syntax in incredulity constructions⁴:

·Them, get married? ·Me, finish my TFG?

For a sentence like this to be grammatical, one has to utter it with the appropriate intonation. A robotic monotonic delivery would make the sentence unacceptable and even unintelligible. Many such examples could be given but the approach of this paper is, as stated above, to look at the implications of the premise rather than at the premise itself and through the concinnity of the different arguments achieve a synergic effect that makes Chomskyan linguistics untenable.

So, let us begin.

³ The deep structure of a linguistic expression is a generativist construct that connects related structures through an underlying shared form. For example, active sentences and their passive counterparts.

⁴ “This construction is very odd [...] because the subject is in the accusative case and the verb is non-finite” (Tomasello, 2000, p. 236).

Chapter 1 Overruled

“Remember not only to say the right thing in the right place, but far more difficult still, to leave unsaid the wrong thing at the tempting moment.”

— Benjamin Franklin

1.1 The original Wug Test

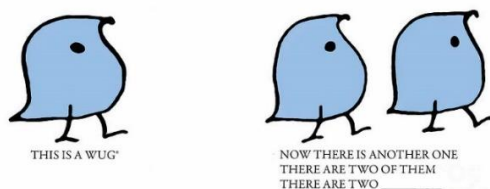
According to Chomskyan linguistics, children develop abstract rules, which they can apply to an independent lexicon, which carries on its shoulders all the semantic responsibility.

In 1958, Jean Berko Gleason published a paper in which morphosyntax did seem to evolve on a track parallel to the rest of language, thus supporting the Generativist view. According to her study, children seemed to have acquired certain abstract rules that they could apply to new vocabulary (specifically non-words) just like a computer programme would. The participants were 28 pre-schoolers and 28 first-graders with ages 4-7 from Harvard Preschool and Michael Driscoll School respectively. The testing proceeded with the aid of drawings and prompting by the experimenter-cum-interviewer.

The morphosyntactic features examined were: “the plural and the two possessives of the noun, the third person singular of the verb, the progressive and the past tense, and the comparative and superlative of the adjective” (Gleason, 1958, p.151).

1.1.1 Plurals

In English there are three allomorphs for forming plurals: /-s/ for words ending in a non-sibilant voiceless sound, /-z/ for words ending in a non-sibilant voiced sound, and /- ɪz/ for words ending in a sibilant: (-/s, z, ʃ, ʒ, tʃ, dʒ/). Participants were tested on each one of them. For example, for the case of the plural marking /-z/, children would be shown a drawing of a bird, and then told it was called a *wug*. Immediately afterwards, they would look at a picture of two of those beings. When prompted to name the two creatures, they — allegedly consistently— would appropriately say: *two wug-s*. *Quod erat demonstrandum*. Case closed.



(Gleason, 1958, p.165)

For decades, these results have been considered irrefutable proof of the Chomskyan premise⁵, often without looking at the actual study and without considering it worthwhile to replicate it in a language different from English. I will look in some detail at Gleason's study and at an attempt to replicate Gleason's study in Japanese that fails to show a mind exclusively based on rule application, but rather more inclined to work by analogy.

In her study, Gleason claims that "if the subject [sic] can supply the correct plural ending, for instance to a noun that we have made up [nonce word], he has internalized a working system of the plural allomorphs in English and is able to generalize to new cases and select the right form. If a child knows that the plural of *witch* is *witches*, he may simply have memorized the plural form. If, however he tells us that the plural of *gutch* is *gutches*, we have evidence that he actually knows, albeit unconsciously, one of those rules which the descriptive linguist, too, would set forth in his grammar" (150). However, Gleason never states what expressions like *can supply* or *is able to generalize* exactly mean to her. If a child supplies the correct answer 70 % of the time, does that mean he or she has acquired a rule? If he or she hits the target form half the time? Is the learner in that case half way through his acquisition of the putative rule? These key epistemological issues are never addressed⁶.

Moreover, even if the linguistic behaviour of a child could indeed be modelled using just words and rules, would that be proof that the model is a transparent reflection of the internal mechanisms producing that output? Taking this correspondence for granted is falling for what Hilferty *et al.* (1998) call *The Projection of Complexity Fallacy*; "in this fallacy, analysts tend to project the properties of their descriptions onto the objects they describe" (Hilferty *et al.*, 1998, p.3). If Newton had fallen prey to the projection fallacy⁷, he would have looked for a law as complicated as the planetary orbits looked to him. Fortunately, however, he did not, and his law of gravitation is pleasingly simple; just a force which decreases with the square of the distance and increases with the mass that creates it. A seasoned thinker, Newton knew better than carrying complexity from observables over to explanations.

⁵ Steven Pinker, in his 1999 book *Words and Rules: The Ingredients of Language* states that "when children are old enough to sit still in experiments, they pass the wug test: After hearing that a man knows how to *rick* or *bing*, they say that yesterday he *ricked* or *binged*." Such unqualified statements are typical in popular science books.

⁶ Another problem is that the study deals with percentages and it is impossible to calculate a p-value just from percentages, and without a p-value one cannot know whether one's results are significant or not

⁷ As Plato and Ptolomy did with their celestial spheres cosmological models. The Minimalist Program is, in part, an attempt to escape the fallacy but it is not without problems of its own as I shall see below.

When one looks at the actual numbers in the study, the prospects for the proponents of the autonomy-of-syntax thesis are not any brighter. When tested for plurals, pre-schoolers' percentage of correct answers only got above the 70% line for two items: *glasses* and *wugs*. Ironically, the item cited by Gleason at the beginning of her paper as illustrating the successful acquisition of the plural rule, *gutch/gutches*, got a meagre 28% for pre-schoolers and a not much better 38% for first-graders. If the passing mark had been 65/100, half the items would have been fails for pre-schoolers whereas first-graders would have failed in four of the 10 plurals⁸. The fact that 99% of the first-graders produced *glasses* does not show much since *glasses* is both the plural of *glass* and a pair of lenses. If it was in fact proof that the rule for forming plurals with the allomorph /- ɪz / had been effectively acquired, why do they perform so poorly when attempting to apply it? An alternative explanation is that children are just proceeding by analogy with the most common forms in the language rather than applying content-less abstract algebraic rules as highly specialized operators in an assembly line.

Now we turn our attention to how children did with verbs.

1.1.2 Verbs

In her study, Gleason sees no interest in examining the formation of the past participle, but the reason she gives for such an omission is odd to say the least. She says: “the past participle of regular or weak verbs in English is identical with the past tense, and since the regular forms were our primary interest, no attempt was made to test for the past participle” (151). Her eschewing the past participle is then a consequence of the functioning of the rule whose existence she set out to prove. That rationale is viciously circular and not scientifically *kosher*. One of the features children were tested on was the gerund ending *-ing*. The possibility that analogical reasoning may be driving kids in their production of new forms fits well with the fact that they are so successful when generalizing the present participle to new verbs for two reasons. First, all verbs in the lexicon take the same present participle ending: *-ing*, and second, there are so many verbs for them to analogize on. The progressive tense rates high both in semantic and phonological schematicity⁹ (no matter how different in shape and

⁸ Of course, results showing a 100% accuracy would have been statistically irrelevant due to ceiling effects, it would have only meant that the task was too easy. Nevertheless, the results are far away from such scores.

⁹ “Schematicity refers to the degree of dissimilarity of the members of a class. Highly schematic classes cover a wide range of instantiations” (Klafehn, 2013, p.175).

meaning verbs are, they all form the progressive adding *-ing*) and in type frequency (even very young children know many verbs); these are the dimensions that estimate the analogical productiveness of patterns¹⁰: abundance and diversity.

When children were tested on their ability to apply rules to new verbs Gleason points out that almost no children used the ablaut pattern *ring/rang/rung*, implying that analogy is not what children use. However, she later concedes: “not one preschool child knew *rang*” (Gleason, 1958, p.165) and that “With *ring* they do not have the actual past *rang*, and therefore no model for generalization” (Gleason, 1958, p.166). It is hard then to rule analogy out.

On the face of all this, the study could at least be qualified as non-conclusive. Instead, Gleason’s envoy is very different in tone: “The answers were not always right so far as English is concerned; but they were always consistent and orderly answers, and they demonstrated that there can be no doubt that children in this age range operate with clearly delimited morphological rules” (171). A more hedged conclusion would have been in order.

1.2 The Japanese Wug Test

Now I turn to Klafehn (2013), which offers both a criticism of Gleason’s article—citing John Taylor’s referring to “conclusions about productivity based on the wug test as The Great wug Hoax” (Klafehn, 2013, p.171) — and a replication/adaptation of her study for children acquiring Japanese as their first language. Since the criticism runs along the same lines as the one presented in this paper I will focus on the Japanese version of the Wug Test.

Klafehn (2013) begins her paper quoting American linguist Joan Bybee to clarify the difference between two kinds of linguistic generalization, rule and analogy:

“A rule (combination of root and suffix) and analogy (use of a novel item in an existing pattern based on stored exemplars)” (Bybee, 2010, p.57)

It will be important to bear this distinction in mind as we walk the reader through the Japanese adaptation of the Wug Test, which—given the typological distance between English and Japanese—may be a little bit hard to follow.

¹⁰ A way to understand how analogy works in language is to think that analogy is to language as peer pressure is to in-group behaviour.

1.2.1 A Japanese preamble

Japanese has two verb classes, regular and irregular. The irregulars are *kuru* (come) and *suru* (do). Nonetheless, the verb *suru* is “highly productive in so-called *light verb constructions* where a noun and *suru* are combined. In these constructions *suru* is used mainly to contribute its inflectional meaning” (Klafehn, 2013, p.175)¹¹. Similar cases happen when, in English, we speak of deer hunting or window-shopping, but still we do not say deer-hunt-doing.

The regular verbs class is further divided into two groups, those with their roots ending in a vowel, which show no productivity in modern Japanese; and those with their roots ending in a consonant; among the latter, those with their roots ending in *-r* do show some productivity¹². These facts already run afoul of a rule-based approach to language in Gleason’s sense. The rule-based approach to language would predict that the irregular pattern should not be productive—no place for light verb constructions—, and that the regular pattern should be equally productive across the board. As it turns out, none of those predictions was fulfilled. Mere regularity does not necessarily predict productiveness, but high *schematicity* does.

The regular verbs with their root ending in */r/* constitute more than 20% of the verbs. Besides, in the nonpast both the */r/* consonant-root verbs and the vowel root-verbs end in */-ru/*, which gives this group an additional boost in frequency. An analogy-based approach to language predicts that final *r*-root verbs will be the most productive; the study confirmed this.

1.2.2 The Study

The participants were 21 five-and six-year old monolingual speakers of Japanese children. The study also included 34 older children and adults as well as two undergraduate students from Tokyo University for purposes of comparison and control.

The training stage presented the participants with seven real verbs to make sure that they understood the task. Three verbs had vowel-ending roots: *mi-ru* (watch), *ne-ru* (sleep), *ki-ru* (wear) and four had consonant-ending roots: *yom-u* (read), *kir-u* (cut) and *ker-u* (kick).

¹¹ “The sudden and unanticipated resignation of Prime Minister Abe in 2007 led to the coining of the term Abe-suru with the meaning to abandon one’s responsibility” (Klafehn, p.174).

¹² “The only regular verbal paradigm that shows any productivity is the root-final */r/* paradigm. Well attested examples are *sabor-u* ‘cut class’; *dabur-u* ‘double’; and *gugur-u* ‘do a Google search’. While it has been suggested that these root-final */r/* paradigms are limited to loan words, the very recent *geror-u* ‘vomit, hurl’; appears to be a native creation. Another recent native example, attributed to the Asahi newspaper, is the verb *Asahir-u* ‘to fabricate or invent something, or to bully’” (Klafehn, p.176).

The investigators made sure that all the participants knew all the inflected forms of the model verbs. After the training, the task proper began. The participants were introduced to six 6 nonce-verbs which they could analogize to one of the verbs presented in the training task, these were: *hom-u* analogous to *yom-u*, *ri-ru* analogous to *mi-ru*, *mur-u* analogous to *kir-u*, *me-ru* analogous to *ne-ru*, *mer-u* analogous to *kir-u*, *rir-u* analogous to *kir-u*, and *hok-u* analogous to *kak-u* or *ik-u*.

Aided by a stationary or a moving picture, interviewers then prompted the participants to produce inflected forms, so the design of the experiment was very similar to Gleason's one. If analogy, rather than rules, is what guides speakers towards new forms, participants will be drawn to forms based on the *kir-u* paradigm—the one with more exemplars—, but they are not expected to be consistent or very accurate across the paradigm, a consistency a rule would provide. That is exactly what was found.

The verbs *mur-u*, *mer-u* and *rir-u* were relatively more successful than the rest, although not very successful in absolute terms. One can safely say that Japanese children, like English children, *failed* the wug test. Rather than rule-applying machines, “The children were very conservative, often preferring to use known vocabulary rather than novel verbs.” For example, “rather than inflect the verb *mur-u*, they described the action they saw with *asi o ugokasite iru* (=she is moving her legs)” (Klafehn, 2013, p.180). The attraction towards the most productive paradigm is best exemplified in children's producing the incorrect form “**hom-u-tte iru* rather than *hon-de iru*” (181). This means that they are taking the nonpast *homu* (a form they are comfortable with) and transforming it into *homur-u* forcing it into the root-final /r/ paradigm. Even older participants showed a “very strong tendency to inflect *ri-ru* and *me-ru* as the consonant-final verbs *rir-u* and *mer-u*” (181).

The conclusion is that children, at least Japanese and English children, do *not* compute abstract rules as if language worked like a software programme in a computer¹³.

A computer-like mind in the generativist sense has proved elusive to identify in mother tongue studies but maybe it will be easier to spot with speakers of a second language.

¹³ One can always argue that children that young may mostly be trying to give the right answer and probably just be confused by the experiment. This is a real possibility and a fair objection by generativists and I make no bones about it. Our main goal is to show that Wug tests are far from conclusive.

Chapter 2 Found in translation

“Never knew before what eternity was made for. It is to give some of us a chance to learn German.”
— Mark Twain

2.1 Who do you mean by *he*?

Support for UG has also come from the SLA field by scholars like Roger Hawkins. Hawkins (2007) cites various studies purportedly proving the existence of innate linguistic knowledge in the form of abstract constraints. I will now look at those studies and his interpretation.

The first study is a contrastive analysis between English and Spanish of the Overt Pronoun Constraint (OPC) by Pérez-Leroux and Glass. This is how Hawkins explains the constraint: “Where there is an alternation in a language between an overt and a null pronoun, only the null pronoun can take a quantified expression as antecedent” (Hawkins, 2007, p.466).

Hawkins illustrates this principle using some Spanish and English sentences:

(1) *Nadie cree que Ø ganará el premio.* [Nobody thinks that (he) will win the prize.]

Ø can refer to *nadie* or to someone else.

(2) *Nadie cree que él ganará el premio.* [Nobody thinks that he will win the prize.]

él cannot refer to *nadie*, it can only refer to someone else, from a previous context.

In English, there is no such constraint, since in English an overt subject is mandatory.

(3) *Nobody thinks that he will win the prize.*

he can refer to nobody.

Moreover, when there is no alternation, Hawkins explains that the overt pronoun must refer to the quantified antecedent as in: *Todo el mundo dice que el presidente habla de él.* In this sentence, *él* has, according to Hawkins, *Todo el mundo* as its antecedent.

The study revolved around participants—a group of native speakers of Spanish and a group of English L2 speakers of Spanish—performing a translation task from English into Spanish in which this constraint played a crucial role.

The sentence to be translated was: *But no journalist said that he is guilty.* The context strongly primed the participants to interpret *he* as referring alternatively to either the quantified antecedent *no journalist* or to a referential antecedent from the context.

If OPC guides native speakers in their choices the expected results for them would be, for the quantified antecedent: no pronoun: *Ningún periodista dijo que fuera culpable.* [No

journalist said he himself was guilty.] In the case of the referential antecedent there is no prediction since OPC would give freedom of choice¹⁴.

As it is often the case when nativist researchers set out to prove the existence of universals, they end up showing just the opposite.

TABLE 5.1 Proportion of overt and null pronouns used (based on Pérez-Leroux and Glass, 1999: 234, Table 1)

Co-referent	Pronoun	L2 Spanish (n = 18)	Native Spanish (n = 20)
Referential	él	32%	68%
	∅	58%	31%
Quantified	él	0%	14%
	∅	93%	85%

L2 speakers followed the OPC more rigidly than native speakers did: in the quantified antecedent situation, they never produced an overt pronoun, whereas native speakers, supposedly driven by this UG constraint, used an overt pronoun 14% of the cases, violating the principle. Hawkins explains this unexpected result thus: “responses from informants in performance tests, including responses from native speakers, are rarely categorical [ironically, the L2 group is categorical, never supplying overt *él* with a quantified antecedent]¹⁵. Extraneous task variables such as inattention, misinterpretation of the context and other factors may intervene. The important point is whether subjects are making the relevant contrast and not whether their judgments are categorical” (Hawkins, 2007, p.468). When performance does not mirror the supposed internal competence, nativists say that performance is faulty and unreliable, but if it does, they never say coincidence is due to accidental extraneous factors. Non-nativists can never win. Data never seems to matter much. Moreover, how could L2 speakers have learnt the OPC principle so well (93% vs. 0%)? They cannot have deduced it from the input since this knowledge “does not appear to be available from experience and must therefore be innate” (466). Moreover, Hawkins adds; “there is little evidence that this phenomenon is taught in classrooms, or that Spanish language teachers are generally even aware of it” (467).

An alternative explanation Hawkins considers is that it is linear distance what is determining the participants’ choices; maybe when the antecedent is far, an overt pronoun is chosen, and when it is close, the null pronoun is preferred.

¹⁴ As a native speaker of Peninsular Spanish my tendency is to always drop the subject in sentences like those.

¹⁵ An objection to this may be that teachers have been very insistent with their students to drop the pronoun. But then, why did they keep 32% of the time for the referential antecedent?

To falsify this possibility, one would need an experiment in which speakers choose the null pronoun when the antecedent is quantified, even if it is far. A 1997 study by Kazue Kano with Japanese native speakers and English learners of Japanese set out to test this.

The study showed a strong tendency (83% for natives and 79% for non-natives) for choosing the null pronoun when the antecedent was quantified, even if it was far in the sentence.

The study however does not rule out, it could not¹⁶, the possibility that linear distance between the pronoun (overt or null) and its antecedent was the criterion Spanish speakers were using in Pérez-Leroux study, but it does rule out the possibility that that is the motivation in Japanese. Still a 17% of native speakers of Japanese decided against a putative OPC criterion that the null pronoun referred to a quantified antecedent.

OPC as a universal constraint does not hold water. What the solution to this problem is is far from established, but one should at least discard these sorts of hypotheses.

2.2 Gulliver's transfers

The last section of Hawkins's study deals with morphosyntax, specifically the challenge it represents for native speakers of Mandarin to pronounce the consonant clusters that appear when inflecting regular verbs with stems ending in a consonant such as *walk* or *phone*. Hawkins cites Lardiere (1998), a longitudinal case study of a bilingual speaker of Mandarin and Hokkien in which "results from naturalistic production data collected over eight years are reported" (Lardiere, 1998, p.1). Specifically, the study looked at her verb inflection.

The results show that overall, irregular verbs were "inflected in 46% of cases, while the regulars in only 6%" (Hawkins 474) *Prima facie*, these results show "that the phenomenon is as compatible with an emergentist account as it is with a nativist one". To clarify what this means, "in an emergentist theory, outcomes can arise for reasons that are not obvious or predictable from any of the individual inputs to the problem" (Bates and Goodman, 1997, p.3). Thus, language is a property appearing out of the interaction of general cognition mechanisms and a meaningful context in which to use them. In my view, these results are especially compatible with the emergentist perspective; irregular verbs being by far more frequent than most regular ones. The five most frequent verbs in Modern English are *to be*, *to have*, *to do*, *to say*, and *to go* correspond to Old English *bēon/wesan* (both irregular),

¹⁶ To assume that I would need to assume the universality of the principle under examination, which would be circular.

habban (weak class 3¹⁷), *dōn* (irregular), *secgan* (weak class 3) and *gān* (suppletive irregular). The reason is that after centuries of analogical ironing of irregularities, only the most frequent among the irregular Anglo-Saxon verbs were used frequently enough for their idiosyncratic inflection to be remembered faithfully, unaltered by successive generations.

Hawkins, however, is not satisfied with an emergentist outlook and suggests instead that “there are reasons to think that under-determined principles of UG are also implicated. One prediction of a simple ban on word-final consonant clusters is that mono-morphemic words like *fact* /-kt/, *lense* /-nz/ should be affected as much as inflected forms like *walked* /-kt/ and *phones* /-nz/. This is not always found. Goad *et al.* (2003) looked specifically at the suppliance of subject-verb agreement –s by 12 L1 speakers of Mandarin at intermediate and low-advanced levels of English proficiency. None had problems in producing final consonant clusters in mono-morphemes” (Hawkins 474).

That is what Hawkins claims, is it true though?

In Goad’s study, participants turned out to behave linguistically according to two patterns: on the one hand, some deleted inflection across the board—this group she calls *ATB*— on the other hand, there was the *Variable deletion group*, which deleted or not according to some UG constraints that I will briefly talk about below. Towards the end of the article, Goad gives us a table displaying the participants’ accuracy for “word-final clusters in monomorphemic forms”: 57% for the *ATB* group and 68% for the *Variable deletion group*. Both Hawkins and Goad consider these numbers tantamount to accuracy, but I beg to differ. From my point of view, general phonotactics looms large in these participants’ poor performance, and the reason why they do better for monomorphemic forms has all to do with semantics.

Final consonant clusters can discriminate between words in two different ways. In monomorphemic forms, they give rise to what I call *cross-lexemic cluster-minimal pairs*. *Cross-lexemic cluster-minimal pairs* are those like *went* vs. *when*, *mask* vs. *mass*, *diss* vs. *disk*, or *pain* vs. *paint*. Missing the consonant cluster means pronouncing a completely different word. Inevitably, more attention¹⁸ is paid to its pronunciation. On the other hand, the final consonant clusters that appear when inflecting regular verbs give rise to *intra-*

¹⁷ This class looks “rather like a mixture of class 1 and class 2 [...] The result of this, when combined with the fact that all four verbs are of very high frequency is that there is a great deal of variation in form” (Richard Hogg, *An Introduction to Old English*).

¹⁸ In Stephen Krashen’s terminology, they become *monitor over-users* (Krashen, 1982, p.19) for that particular aspect. It would not surprise me if accuracy in hitting the target consonant cluster slowed down their production.

lexemic cluster-minimal pairs like *arrive* vs. *arrived*. The semantics stakes are in this case much lower, speakers let themselves go and their L1 phonotactics take over. Of course, nativists would flatly reject such a holistic explanation, because it infringes the autonomy-of-syntax thesis. Still, in language as in life, context is well-nigh everything.

Goad's study itself has other conceptual weaknesses, which I proceed to talk about briefly. As I have explained above, Goad *et al.* explain this phenomenon from a UG point of view. Their starting point is the establishment of a prosodic abstract hierarchy. They then establish two constraining principles: *exhaustivity* and *nonrecursivity*. Afterwards, Goad *et al.* proceed to explain how and when they are violated in the two languages in question. "In English, both EXHAUST and NONREC can be simultaneously violated in inflected forms (...) although both constraints can be independently violated in Mandarin, they cannot be simultaneously violated" (Goad *et al.*, 2003, p.247-248). All this, together with yet another constraint—BINARITY—which in English limits syllable rhymes to a maximum of two mores, accounts for the difficulties Mandarin speakers suffer when it comes to pronounce words like *walked* ['wɔ:kt]. However, this mathematical explanation is faulty even for English when a diachronic perspective is taken; we will now see why that is.

In a nutshell, Goad *et al.* explain that *width* is pronounced ['widθ] because if '-th' was directly added to *wide* ['waɪd] it would *give* ['waɪd·θ], whose first segment ['waɪ] would violate binarity (the correct form violates EXHAUST and NONREC but that seems to pose no problem)¹⁹. To avoid this violation [aɪ] shortens to [ɪ]. The problem with this explanation is that it misses the forest for the trees. Close syllable shortening (CSS) is a process that evolved gradually, and it was only complete by the end of the Middle English Period. In Old English, the past indicative of the word *cēpan* was *cēpte* [ke:pte]. Progressively the final /e/ reduced to schwa and eventually disappeared, that produced a closed syllable [ke:pt] which gradually shortened its nucleus to /e/, giving the present form [kept]. However, diachronic processes do not happen overnight, they take many generations to complete. The Great English Vowel Shift extended over four centuries: from the 14th to the 18th century. If generativists want this well-established fact to fit with their approach based on constraints

¹⁹ It is surprising that when discussing phonotactics, generativists do not make any reference to physiology or to the dynamics of opposite forces constituted by lazy speakers and demanding hearers. For more information on the latter topic please check *The Unfolding of Language* by Guy Deutscher.

and its violations, they would have to shelter constraints and violations that fluctuate with time, but that would make no sense within the paradigm.

Jonathan Swift left us testimony to the reality of the gradual and slow character of these kinds of changes. In his book, *A Proposal for Correcting, Improving and Ascertaining the English Tongue*, Swift bemoaned²⁰ the evolving pronunciation of past suffix clusters: “By leaving out a vowel to save a syllable, we form so jarring a sound, and so difficult to utter, that I have often wondered how it could ever obtain”.

Goad *et al.* state that *[ə'ri:vɪd] is “unattested” (249); the same way that an *avant la lettre* Generativist grammarian writing in Chaucer’s time could have said that /kept/ is unattested. However high the Saussurian wall between diachronic and synchronic linguistics may appear, language is in a state of constant flux and a snapshot explanation based on generativist inflectional phonotactics can only take us so far.

Apart from these conceptual issues—, which one cannot simply seep under the rug—Goad’s study suffers from a methodological problem which is nowhere clarified in the article. For reasons that for lack of space I cannot go into here, Goad *et al.* make certain predictions depending on whether a certain word can reorganize its prosodic structure through resyllabification when in contact with other words. Logically then one would be interested in contrasting how participants pronounced words in isolation vs. in connected speech. However “oral production data were elicited by having subjects describe two sets of pictures which illustrated sequences of events” (255). That is to say, no such contrast was investigated or considered worthy of special attention, and all items were uttered in connected speech.

I think it is safe to conclude that, to the very least, neither Pérez-Leroux nor Goad’s study offer conclusive evidence for the psychological reality of UG.

The evidence for the existence of a universal grammar has proved as flimsy in studies with second language speakers as it was for first language speakers. Still, there may be an uncontroversial logical necessity for its existence regardless of how difficult it is to put our finger on it. That will be our next topic.

²⁰ “In 1712, the English language, according to satirist Jonathan Swift, was in chaos. He outlined his complaints in a public letter to Robert Harley, leader of the government, proposing the appointment of experts to advice on English use. The model was to be based on that of the Académie Française, which had been regulating French since 1634. His proposal, like all the others, came to nothing. To this day, no official regulation of the English language exists” (British Library Website: <http://www.bl.uk/learning/timeline/item126681.html>).

Chapter 3 Drought or Flood.

“nothing that is worth knowing can be taught”

— Oscar Wilde

3.1 Poverty of the Stimulus

The term *argument from poverty of the stimulus* (PoS) as such was first used in Chomsky’s book *Rules and Representations* in 1980. Roughly speaking, what PoS boils down to is that there is not enough data in the linguistic environment for an infant to learn certain linguistic facts of its target language. I have to state the argument vaguely because, as Pullum himself plaintively denounces, “no one²¹ attempts to *state* the argument” (Pullum, 2002, p.11).

The linguistic environment is either unrewarding, finite, idiosyncratic, incomplete, degenerate or lacking in negative feedback—the infant is not informed about the ungrammaticality of an incorrect utterance. Pullum and Scholz (2002) explain how different nativist linguists (Ramsey, Stich, Garfield, Seidenberg, Haegeman, among others) have each highlighted one or several of those aspects. If the evidence is indeed too degenerate or too incomplete or too idiosyncratic (according to whoever makes the claim) to reliably help the infant to acquire its language, this would prove that certain linguistic knowledge, UG, must be innate. This is not a misrepresentation of the nativist position, as recently as 2002 Chomsky wrote, “A child is exposed to only a small proportion of the possible sentences in its language, thus limiting its database for constructing a more general version of that language in its own mind/brain. This point has logical implications for any system that attempts to acquire a natural language on the basis of limited data. It is immediately obvious that given a finite array of data, there are infinitely many theories consistent with it but inconsistent with one another” (Hauser, Chomsky and Fitch, 2002, p.1577).

Although the PoS pretty much opens every Chomskyan textbook today, as pointed in Thomas (2002) this runs contrary to its historical development and the logical order that deduces its necessity²². Let us take a very brief look of its gradual appearance from a historical perspective. In *Syntactic Structures* Chomsky pointed out that “the speaker, on the basis of a finite and accidental experience with Language, can produce or understand an indefinite number of sentences” (Chomsky, 1957, p.15). Later, in his 1959 review of B.F. Skinner’s *Verbal Behaviour*, he drew attention to the learner’s internal structure, to what he brings to

²¹ No-one in the Generativist camp.

²² PoS is the best example in Linguistics of putting the cart before the horse.

the table, to the detriment of the external factors that so much interested Skinner²³. Nevertheless, according to Thomas (2002) “the review itself is not a major contribution to the development of the concept” (54). In 1965, in *Aspects of the Theory of Syntax*, Chomsky goes as far as to say, “the primary linguistic data that he [the child] uses as a basis for this act of theory construction may (...) be deficient in various respects” (Chomsky, 1965, p.201). Finally, in 1980, as mentioned above, he took the definite step and overtly stated a PoS, drawing inspiration from the Platonic dialogue *Meno*; in it, using maieutics, Socrates retrieves from an ignorant slave knowledge of geometry that no one could have taught him and consequently had to have been there from the beginning²⁴. Plato attributed it to reincarnation; Chomsky attributes it to our genetic blueprint.

The reason Chomsky thinks there is a paucity of data stems from the autonomy-of-syntax thesis. If syntax is meaningless, then it is just like algebra, and children cannot really do algebra—and neither can most adults. This by itself raises the bar for the amount and quality of the data children would need to learn all the grammar facts of their language. Unaided by meaning, the autonomy of syntax makes learning syntax all but impossible. The only two ways out of the conundrum are: a) syntax is not autonomous from the rest of language and there is more and more usable data than Chomsky is willing to concede or b) there is something making up for the scarcity of data, namely UG (PoS *ergo* UG)

The problem for generativists when it comes to inspecting the first solution is that Noam Chomsky does not like data, particularly raw data one draws from looking into the actual linguistic environment, in other words, he discounts corpus linguistics. He mostly relies on the intuition about language native speakers possess. This faith in intuition does not falter even when explicitly proved wrong. As a token of this steadfastness, look at this exchange between Chomsky and late American linguist Anna Granville Hatcher:

Chomsky: The word *perform* cannot be used with mass-word objects: one can perform a task, but one cannot perform *labour*.

Hatcher: How do you know if you don't use a corpus and have not studied the verb *perform*?

Chomsky: How do I know? Because I am a native speaker

²³ As long as that structure is not exclusive to language, I could not agree more.

²⁴ “We do not learn, and that what we call learning is only a process of recollection.” Plato, *Meno*

of the English language. (Harris, 1993, p.97)

Hatcher came up, on the spot, with a counter-example: *perform magic*.²⁵

Of course, Chomsky could have responded saying that Hatcher's counter-example came from her intuition, so one can never win: it is a Popperian nightmare.

As for corpus linguistic, here is how he feels about it, this is from a recent interview with University of Pecs Professor Jozsef Andor in 2004:

“Corpus linguistics doesn't mean anything. It's like saying suppose a physicist decides, (...) what they're going to do is take videotapes of things happening in the world and they'll collect huge videotapes of everything that's happening and from that maybe they'll come up with some generalizations or insights. Well, you know, sciences don't do this. But maybe they're wrong. Maybe the sciences should just collect lots and lots of data and try to develop the results from them. Well if someone wants to try that, fine. (...), if results come from study of massive data, rather like videotaping what's happening outside the window²⁶, fine-look at the results. I don't pay much attention to it” (p. 97 in Andor, Jozsef. (2004) ‘The Master and his Performance. An interview with Noam Chomsky’, in *intercultural pragmatics* 1.1:93-111)

In any case, Chomsky would only consider corpora which were exhaustive and exclusive to one individual: “in order to demonstrate that there is no relevant experience with respect to some property of language, we would have to have a complete record of a person's experience; there is no empirical problem in getting this information, but nobody in his right mind would try to do it” (Piattelli-Palmarini,1983, p.113).

Therefore, corpora are either possible to get but irrelevant or impossible to collect if relevant. Turning one's back to corpora makes it very hard for anyone to get to know the reality (or at least an approximation to it) of what children actually hear and what they can infer from it through inductive reasoning. Nonchalantly, Chomsky forfeits this empirical imperative.

Pullum and Scholz (2002) did use corpus linguistics — they used the Wall Street Journal 1987-1989 (Linguistic Data Consortium 1993), abbreviated as WSJ. They inspect four candidates to unlearnable-from-the-data linguistic knowledge: plurals in noun-compounding, auxiliary sequences, anaphoric one and auxiliary initial clauses. I will focus on the last one,

²⁵ This magic bullet can be dodged by considering it an aberration since it is not productive. Still the question of the infallibility of intuition remains.

²⁶ This is just an intellectually lazy cop-out covering for empirical data aversion. Scientists do not look out the window, but they do look through telescopes.

which they refer to as “The apparent strongest case of alleged learning from crucially inadequate evidence discussed in the literature, and certainly the most celebrated” (Pullum, 2002, p.36). Essentially, it refers to the acquisition by children of the structure-dependent rule for making yes/no questions like (2) out of declarative sentences containing a complex noun phrase like (1) and their avoiding ungrammatical forms like (3):

(1) *The man who is tall is in the kitchen.*

(2) *Is the man who is tall in the kitchen?*

(3) **Is the man who tall in the kitchen?*

If the infant’s procedure to acquire language were a brute force process of trial and error of hypotheses, a possible null hypothesis would be to think that the question is formed simply by fronting the first auxiliary rather than the main clause auxiliary that affects the whole noun phrase. The problem for Chomskyan linguists is that there is not enough data to rule this hypothesis out so. If children end up with the right abstract knowledge, it is because this is preinstalled (UG), not because they proceed by ruling out hypotheses²⁷: “A person might go through much or all of his life without ever having been exposed” and “you can go over a vast amount of data without ever finding such a case²⁸” (Piattelli-Palmarini, 1980, p.115)

However, Pullum points out, if Chomsky was right and only once in a blue moon were sentences like this uttered, there could perfectly be speakers “who have acquired an incorrect structure-independent generalization instead, but who are never detected because of the rarity of the crucial situations in which they would give themselves away” (Pullum, 2002, p.40). Incompetence would roam free. Chomskyans can’t have their cake and eat it too.

Going back to the question of scarcity, Pullum cites Sampson (1989) offering anecdotal evidence of children encountering the pattern in question, noting that William Blake’s poem *Tiger*²⁹ contains the line: *Did he who made the lamb make thee?* which comes from the declarative: *He who made the lamb make thee.* Fronting the first auxiliary would have given: **Did he who make the lamb made thee?* Sentences as typical for children to hear as *will those who are coming raise their hands?* or *could a tyrannosaurus that was sick kill a triceratops?*

²⁷ It would be the unsolvable Gavagai problem carried over to syntax. I agree trial and error of hypotheses does not seem to be what children do. Statistical patterns and general cognitive biases seem to be more interesting paths to explore.

²⁸ Chomsky constantly leapfrogs criticism by jumping from a theoretical logical impossibility to empirical insufficiency.

²⁹ "The Tyger" is a poem by the English poet William Blake published in 1794 as part of the *Songs of Experience* collection. Literary critic Alfred Kazin calls it "the most famous of his poems, and *The Cambridge Companion to William Blake* says it is "the most anthologized poem in English" It is one of Blake's most reinterpreted and arranged works.

also contain the pattern in question. More scientifically, only 500 questions in the Washington Street Journal (WSJ) they come across examples like:

Is a young professional who lives in a bachelor condo as much a part of the middle class as a family in the suburbs, whose declarative counterpart is A young professional who lives in a bachelor condo is as much a part of the middle class as a family in the suburbs.

The objection could be raised that sentences like that are not typically addressed to children but Berman (1990) “notes that speech addressed to the learner is not the only evidence about the language that the learner has; learners learn from at least some utterances that are not addressed to them. And indeed, Weddell and Copeland (1997) found that even children between ages 2 and 5 pick and understand much more from the language they hear on television than their parents typically realize” (Pullum 2002:23).

Even excluding do-support from the search, the 180th interrogative they find in the WSJ corpus is of someone wondering *is what I’m doing in the shareholders’ best interest?*

Summing up, children are very likely to have access to sentences that show that it is the main clause verb the one that is fronted³⁰ to form the yes/no question.³¹

3.2 Secondary Roads

There are two other ways to circumvent the necessity of positing a UG to fill the gap between supposedly impoverished data and final native competence. One is to think of complex noun phrases like *the man who is tall* or *what I’m doing* as an intonation unit, “Croft (1995) has shown that intonation units have been shown to correspond very closely to grammatical units” (Hilferty, 2003, p.70). This simplifies sentences and ties a noun phrase together tighter than any generativist upside down tree nonterminal node would. The fact that units such as *the man who is tall* are themselves conceptual units lends support to this approach.

Another one is to consider transitional probabilities; the probability of finding an adjective next to *who* without a prosodic pause is practically nil. *Who* can be heard adjacent to an adjective separated by a prosodic pause in sentences like: *look at John who, short as he is, plays in the NBA*. But children, who are finely tuned to these nuances, do not confuse the two cases. Of course, none of those arguments would fly for a Chomskyan because syntax is autonomous.

³⁰ Even fronting itself could be regarded as an unnecessary assumption.

³¹ If indeed they are using structure-dependency rules, which cannot be taken for granted across the board.

One of the take-home messages of Chomskyanism is that language is all about structure dependency and that linear dependency plays no role at all: “Hierarchical representation is at the heart of human language”; “hierarchical representations are omnipresent in human language syntax” (Berwick and Chomsky, 2016, p.115). This is because, generativists say, the language faculty cannot count: “‘Unnatural’ phonotactic rules, such as one that would require every fifth sound in language to be a consonant of some type—a so-called ‘counting’ language. ‘Counting’ languages aren’t natural languages” (126). I agree that the phonotactic rule given as an example of a *counting* language is indeed far-fetched and absurd, but I disagree with the fact that languages are alien to counting. The claim is merely an untruth. In OE sentences beginning with *bā* (then), the verb must appear in second position³². That is a linear dependency rule and it is legitimate to suppose that it was successfully acquired by infants in the 9th century. My point here is simply to dispel the myth that linear dependency is irrelevant. Chomsky (2016) insists on the irrelevance of any rule that is not hierarchical³³:

(1) *He said Max ordered sushi.*

(2) *Max said he ordered sushi.*

(3) *While he was holding pasta, Max ordered sushi.*

In sentence (1), Chomsky explains that *Max* and *he* cannot refer to the same person. The reason is that in the corresponding phrase structure analysis *he* governs *Max* and apparently that is all there is to it.

But here is the catch: these sentences, as always in generativist analyses, are out of context, particularly out of their communicative context. The first two sentences are actually reported speech; so sentence (1) is equivalent to (1’) He (someone) said: “Max ordered sushi”

Thus, for *he* to refer to *Max* would mean that Max referred to himself in the third person and we do not normally do that³⁴. However, it is easy to imagine a very commonplace situation in which that takes place. Imagine a double date: John and Jane, and Max and Maxime. Max and Maxime are fighting, Maxime wants John and Jane to take her side and sighs: “I never know what Max wants”, Max, irate retorts: “Max wants you to shut up”. Jane herself missed Max’s response because she was in the bathroom washing her hands. After a long night, John

³² In syntax, verb-second (V2) word order places the finite verb of a clause or sentence in second position with a single constituent preceding it, which functions as the clause topic.

³³ The chapter I refer to here is tellingly called *Triangles in the Brain*. It should be clear that for Chomskians phrase structure trees are not a descriptive model; they are supposed to be a psychological reality.

³⁴ Royalty often do though.

and Jane get home and John relates the episode to Jane: —John: [...] and then *he said Max wants you to shut up*. Are *he* and *Max* the same person? Aye. *Voilà*, we just overruled structure-dependency.

3.3 Philosophical and epistemological obstacles

Moreover, PoS arguments are not without philosophic-epistemological problems of their own. These have been addressed most notably by Geoffrey Sampson in his 1980 article *Popperian Language-Acquisition Undefeated*, published as part of a debate with Stephen P. Stich in *The British Journal for the Philosophy of Science*. The argument is like this: if linguists cannot extract the rules of grammar from the linguistic behaviour of the speakers —because there is a PoS—, how do they know the child has in fact acquired the right grammar. It could perfectly be a different grammar altogether if its effects cannot be perceived in the actual utterances. Generativists, like Stich and Chomsky, could respond that the adult speaker can introspect into intuitive knowledge to confirm or disconfirm whether the grammar the child has developed does indeed coincide with that of the native speakers surrounding him. But “that would be circular” says Sampson because it presupposes what one is trying to elucidate, namely whether grammatical competence can be gleaned from the data or not: “Only if Stich and Chomsky are right in their anti-empiricist approach to language-acquisition that the special category of data available to a scientist who chooses to investigate his own language community includes truths which might not be discoverable by a rational scientist working exclusively with the kind of behavioural data available to the child. Stich is not entitled to assume the correctness of this view in the course of an argument purporting to show that it is correct” (Sampson, 1980, p.65). Innate knowledge is then unfalsifiable and has to be accepted by *fiat*. Stich responded to Sampson a year later in a paper called *Can Popperians learn to talk?* Stich claims that the difficulties posed by Sampson can be avoided if a gradual scale of similarity is posited between grammars, since he implicitly concedes that unsurmountable obstacles arise when the equivalence between grammars is laid out in binary polar terms. Stich argues that the infant acquires a grammar which is “roughly input-output equivalent” (Stich 163) to that of the adult speakers surrounding it. According to Sampson, this category is reached when the grammars are close to generate the same class of sentences. If the linguistic behaviour is similar, the grammar generating it must be so too. The problem with this tactic is that it fits ill with a hard-core

Chomskyan approach which is interested in the internal machinery and a clear and solid distinction between competence and performance. What would the meaning be of two rules being roughly equivalent is anybody's guess.

Nevertheless, if one maintains that syntax is a sort of abstract algebra devoid of meaning and context there is indeed an impossibility of acquisition caused by a PoS that can only be solved by a putative UG. The way out of this maze is to turn to a view of acquisition that is piecemeal, conservative and meaning-based. Children take very conservative baby steps in their acquisition of language. They progress through what Michael Tomasello refers to as verb-islands, which are constructions that pivot on a certain lexical item tied to specific meaningful situations. For example [*X draw Y*], or [*cut Z*]. The important thing here is that the patterns with which a certain verb can be used are not automatically generalized and transferred to other verbs. Each verb is an island constructed upon a certain usage framework, but at the early stages of language development, these patterns cannot swim between islands to be shared amongst the different verb islands, and are for the time being marooned. Children “use some of their verbs in the transitive construction—namely, the ones they have heard used in that construction—but they do not use other of their verbs in the transitive construction—namely, the ones they have not heard in that construction (Tomasello, 2000, p.222). Tomasello also argues for an abandonment and replacement of the “mathematical view of language” (247), which is the one advocated by generativists. Along this same line of thought, that most of grammar acquisition stands on the shoulders of the lexicon, Bates and Goodman (1997) showed that there is a “strong relationship between grammar and lexical development during the early stages of language learning” (Bates and Goodman, 1997, p.4). Looking at “recent evidence on the relationship between lexical development and the emergence of grammar in normally developing children” they prove that the “emergence and elaboration of grammar are highly dependent upon vocabulary size throughout this period” (5). For example, the best predictor of grammatical complexity at 28 months is “total vocabulary size at 20 months” (6). Bates and Goodman conclude that there is no evidence for a “hard dissociation between grammar and the lexicon” (14). This dovetails perfectly with the verb-island hypothesis. It seems that the situation is like a game of whack-a-mole for Chomskyans, no sooner a problem is dealt with than a new one appears. An attempt to deal solve all the difficulties simultaneously has been to reduce UG to a minimum.

Chapter 4 UG loses weight

4.1 The Minimalist Program

"I enjoy acronyms. Recursive Acronyms Crablake "RACRECR" Especially Create Infinite Regress"

— Douglas R. Hofstadter

The latest trend in generativist linguistics — the Minimalist Program (MP) — has been to try to reduce UG to a conceptual minimum: the capacity for recursion: “recursion is the only uniquely human component of the faculty of language” (Hauser, Chomsky and Fitch 2002 1569) A rule is called recursive when it can be repeatedly applied to its own output. For example, in X-bar theory we have: $X' \rightarrow \text{Adjunct}, X'$. This derivative rule can be indefinitely applied, as when piling up adjectives on a noun, like in the expression *big bad wolf* we have: $\bar{N} \rightarrow \text{Adj}_1 \bar{N}$: bad wolf

We can apply the rule again and obtain:

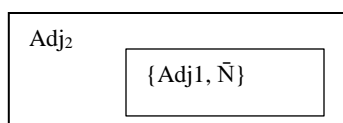
$\bar{N} \rightarrow \text{Adj}_2 \text{Adj}_1 \bar{N}$: big bad wolf.³⁵

If we write it this way: $\{\text{Adj}_2 \{\text{Adj}_1 \bar{N}\}\}$ this nomenclature gives us the inkling of modelling recursion by means of an algebraic operator that takes two arguments and simply makes a pair out of them, like this: $M[x, y] = \{x, y\}$. This operator is called Merge³⁶ (M) in the MP. In the example above, it would work like this:

$M[\text{Adj}_1, \bar{N}] = \{\text{Adj}_1, \bar{N}\}$

$M[\text{Adj}_2, \{\text{Adj}_1, \bar{N}\}] = \{\text{Adj}_2, \{\text{Adj}_1, \bar{N}\}\}$

Recursion works just like Matryoshka dolls:



Recursion allows English to construct sentences like³⁷:

[Colourless [green ideas]] sleep furiously

This is [the rat [that ate [the malt [that lay [in [the house [that Jack built]]]]]]]

[Buffalo buffalo [Buffalo buffalo buffalo]] buffalo Buffalo buffalo

³⁵ Since Merge is purely abstract, it has no way of knowing that **bad big wolf* is not correct so the merge-only approach is already a false start.

³⁶ Chomsky describes merge simply like this: “An operation that takes two objects already constructed, call them X and Y, and forms from them a new object that consists of the two unchanged, hence simply the set with X and Y as members” (Berwick and Chomsky, 2016, p.70).

³⁷ English was not born with recursion. Sentences like today’s *I know that the king is dead* have their origin in *I know that. The king is dead.* The process that transforms one structure into the other is called *grammaticalisation*. For more information check the book *Grammaticalization* by Elizabeth Traugott.

With this reduction, generativists have increased the plausibility of three things:

(1) UG having popped up in a single and relatively recent, in evolutionary terms—around 10^5 years ago— miraculous mutation: “It seemed necessary to attribute great complexity to UG in order to capture the empirical phenomena of languages and their apparent variety. It was always understood, however, that this cannot be correct. UG must meet the condition of evolvability, and the more complex its assumed character, the greater the burden on some future account of how it might have evolved—a very heavy burden” (Berwick and Chomsky, 2016, p.93) Miraculous mutations can only go so far though. I will look at this below.

(2) UG being codified in our genes. Genes can hardly be credited with the codification of the tracking of empty categories like traces, OPC, subjacency conditions, move α , wh-Island Constraint, locality principle A...the list goes on. It is asking for a suspension of disbelief that, paraphrasing Hamlet speech to the struggling actors, would out-Coleridge Coleridge.

(3) a causal relationship between universal tendencies in human languages and UG. If UG appeared relatively recently—before the out-of-Africa exodus—, it would have had little time for evolving and it would be that and not the universality of human experience and human bodies³⁸ what would account for universal tendencies in languages.

However, they are, like Claudius, hoist with their own petard, because the transition from a bonny UG to a bony one does not run smooth. Many problems arise because of this conceptual U-turn. If UG is just recursion, then:

(1) it does not seem to be much of a leg-up for children acquiring unlearnable linguistic facts. The has-to-be-learned periphery easily dwarfs the innate core.

(2) how can there be at least one extant language³⁹ — and who knows how many other there may have existed and become extinct? — that seems to function perfectly well without?

(3) is it not surprising that recursion is present in other cognitive domains and other species?

I will concentrate in problems (2) and (3).

³⁸For an extensive treatment of the effects of embodiment in language see *Louder Than Words: The New Science of How the Mind Makes Meaning* by Benjamin K. Bergen.

³⁹ Pirahã.

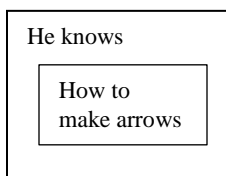
4.2 Without recursion

In 2005, Dan Everett published *Cultural Constraints on Grammar and Cognition in Pirahã*. In that paper, Everett drew attention to some aspects of “the only surviving member of the Muran language family” (Everett, 2005, p.622) such as the lack of numerals, the lack of colour terms, or the simplicity of its pronoun system. However, what really spurred an outcry in the generativist camp was what Everett referred to as a *lack of embedding*. As a bonus, Everett argued for an intimate relationship between culture and language that would account for “these apparently disjointed facts about the Pirahã language (...) ultimately derive from a single cultural constraint in Pirahã, namely, the restriction of communication to the immediate experience of the interlocutors. Grammar and other ways of living are restricted to concrete, immediate experience (where an experience is immediate if it has been seen or recounted as seen by a person alive at the time of telling), and immediacy of experience is reflected in immediacy of information encoding—one event per utterance” (622).

If a language with no recursion was not bad enough, this absence of recursion was caused by culture. This was a real thorn in generativists’ side.

Now I will inspect how Everett infers the lack of embedding for Pirahã.

Everett illustrates this syntactic fact through the sentence *He knows how to make arrows well*, (Everett, 2005, p.629). In the generativist tradition, this is analysed using recursion as: [*He knows [how to make arrows well]*].



Translating into Pirahã, the result is:

(1) *hi ob -a' a'a' i' kahai kai -sai*

To facilitate the following of Everett’s argument, I will codify the different elements of the Pirahã sentence thus: $a_1a_2a_3b_1b_2$, where

$a_1 = hi = he$

$a_2 = ob = sees/knows$

$a_3 = -a' a'a' i' = well/attractive$

$b_1 = kahai = arrow$

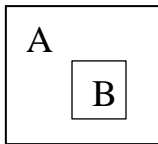
$b_2 = kai-sai = \text{making}$

$a_1 a_2 a_3 = A$

$b_1 b_2 = B$

Sentence (1) is then AB

Another grammatical order for this sentence would be $b_1 b_2 a_1 a_2 a_3$, which is BA. An ungrammatical order would be $*a_1 b_1 b_2 a_2 a_3$, which can be graphically represented as:



This would roughly correspond to: [he [arrow making] knows well].

If the latter ordering was grammatical, it would suggest embedding, but the sentence is incorrect in Pirahã. The grammaticality of AB and BA argues for “the paratactic conjoining of the noun phrase ‘arrow-making’ and the clause ‘he sees well’” (Everett, 2005, p.629), that is, linear dependency. Another plausible analysis, Everett concedes, is embedding or structure dependency, that is to consider B in the order AB as an embedded object of A, what we would traditionally call an X-Complement. Everett rules out this possibility because “as an object, the phrase “arrow-making” should appear before the verb, whereas here it follows it”⁴⁰ (629), this corresponds to the ungrammaticality of $*a_1 b_1 b_2 a_2 a_3$.

“Further, although the order of “complement” and “matrix” clauses can be reversed, the “embedded” clause can never appear in direct-object position” (629). AB and BA are then the only acceptable orders.

hi go' 'igi' -ai kai -sai hi 'ob -a' a' a' i'

How this relates to a cultural constraint is more difficult to assess. Everett argues that it “follows from the principle of immediacy of information encoding” (631), because “embedding increases information flow beyond the threshold of the principle” (631). In other words, it would betray *the one event per utterance* principle stated above.

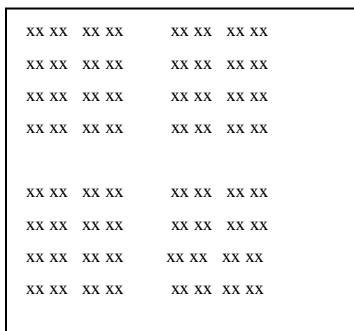
⁴⁰ Pirahã is predominantly SOV.

In an essay edited by British philosopher Nigel Warburton⁴¹, Everett says that “perhaps the Indonesian language Riau could lack recursion” too. Difficulties pile up on Chomsky’s desk. The other drawback of reducing UG to merge that I will inspect is the presence of recursion in other cognitive domains and in other species.

4.3 Recursion reloaded

4.3.1 Recursion is in the eye of the beholder

In 2005, Ray Jackendoff and Steven Pinker published *The nature of the language faculty and its implications for evolution of language*. The abstract announces: “We show that recursion is found in visual cognition, hence it cannot be the sole evolutionary development that granted language to humans”. This is claimed in contraposition to the claim that “there are no unambiguous demonstrations of recursion in other human cognitive domains, with the only clear exceptions (mathematical formulas, computer programming) being clearly dependent on language” (Hauser, Chomsky and Fitch, 2005, p.179). The way Jackendoff and Pinker give the lie to such a claim is simple and elegant. They show this fractal-like image:



Jackendoff and Pinker explain the way in which we perceive this display: “This display is perceived as being built recursively out of discrete elements which combine to form larger discrete constituents: pairs of x’s, clusters of two pairs, squares of eight clusters, arrays of four squares, arrays of four arrays, and so on. One could further combine four of these super-arrays into a still larger array and continue the process indefinitely” (217).

Recursion is then undeniably present in the visual apparatus.

⁴¹ <https://aeon.co/essays/why-language-is-not-everything-that-noam-chomsky-said-it-is>.

4.3.2 English walks, Pirahã hops.

Another non-linguistic domain in which our brain makes use of recursion is the motor programme for walking. How do we walk? Let us call the position of our right foot in a time i R_i and ditto for the left foot. Every step consists in leaving one foot still, let us say the right foot, and moving the other foot forward twice the distance of separation between the two feet. Every time, the position of the feet depends on where they were in the previous moment. We can code this mathematically as $(R_i, L_i) = (R_{i-1}, L_{i-1} + 2(R_{i-1} - L_{i-1}))$

Hopping does not work like this, when we hop, the position of the right and left foot is the same and it is simply the number of hops multiplied by how much we hop. In this case we have $(R_i, L_i) = N \times \text{distance}$. There is no recursive dependency.

4.3.3 Recursion on the fly

I will now look at the presence of recursion in other species, in this case songbirds. Gentner *et al.* (2006), in a study with eleven European starlings, it was shown that “the European starlings accurately recognize acoustic patterns defined by a recursive, self-embedding grammar (...) thus the capacity to classify sequences from recursive, centre-embedded grammars is not uniquely human” (Gentner, 2006, p.1204).

Starlings sing songs which contain warbles and rattles, which in the article are called *motifs*. The main idea was to use those motifs to build sequences that use either iteration—which creates adjacent dependency—or recursive self-embedding—which creates hierarchical dependency—(see Figure 1) and see if after training birds respond differently to the different sorts of construction. Gentner *et al.* chose eight rattles (a_i) and eight warbles (b_i) to create motif sequences of these two kinds:

ITERATION $(AB)^2$: $a_i b_j a_k b_l$

RECURSION $A^2 B^2$: $a_i a_j b_k b_l$ with $i, j, k, l = 1, \dots, 8$ ⁴²

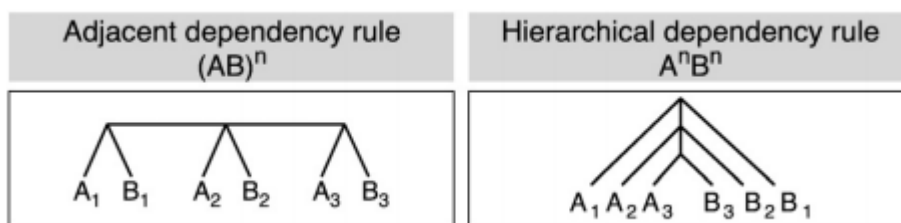


Figure 1.

⁴² This gives us $8^4=4096$ possible sentences for each of the two languages.

Out of the 4096 possible sentences, they picked 16 as the baseline training stimuli. Once the birds managed to successfully distinguish between the two patterns, several possibilities other than their being capable of detecting recursion were investigated and ruled out. One such possibility was that the birds might have just rote memorized the specific sequences present in the training. However, Gentner *et al.* explain that when transferred abruptly from the 16 baseline training stimuli to 16 new sequences from the same two grammars (A^2B^2 and $(AB)^2$) the starlings did significantly well: “significantly better than chance performance” (Gentner, 2006, p.1204)

Another one was that maybe the starlings were able to recognize just the iterative pattern — the easiest one— and could only tell that the recursive one was *not* the one they knew. This was discarded by having each pattern contrasted with different types of sequences belonging to none of the patterns birds knew, and which were neither iterative nor recursive. They were successful at every task. The cherry on the cake was that starlings proved capable of generalizing their discriminatory behaviour for A^3B^3 , $(AB)^3$, A^4B^4 , and $(AB)^4$.

Gentner *et al.* conclude, “at least a simple level of recursive syntactic pattern is therefore shared with other animals” (Gentner, 2006, p.1206)

Summing up, recursion seems to be neither universal across human languages nor even exclusive to language or humans.

But how did language come into existence? Is there anything known at all? Can we at least discard something?

Chapter 5 Out of the Blue?

La Société de Linguistique de Paris, fondée en 1866, doit une partie de sa célébrité à l'article II de ses premiers statuts: art. 2 - La société n'admet aucune communication concernant soit l'origine du langage, soit la création d'une langue universelle

5.1 In the beginning was the word

As for the phylogenetic appearance of language, Chomsky's stance can be summed up thus: "Within some small group from which we are descended, a rewiring of the brain took place in some individual, call him Prometheus⁴³, yielding the operation of unbounded Merge, applying to concepts with intricate (and little understood) properties [...] Prometheus's language provides him with an infinite array of structured expressions" (Chomsky, 2010)

"One fact that does appear to be well established is, as I have already mentioned, that the faculty of language is a true species property, invariant among human groups — and furthermore, unique to humans in its essential properties. It follows that there has been little or no evolution of the faculty since human groups separated from one another. Recent genomic studies place this date not very long after the appearance of anatomically modern humans about 200,000 years ago, perhaps some 50,000 years later, when the San group in Africa separated from other humans. There is some evidence that it might have been even earlier. There is no evidence of anything like human language, or symbolic activities altogether, before the emergence of modern humans, *Homo Sapiens Sapiens*. That leads us to expect that the faculty of language emerged along with modern humans or not long after — a very brief moment in evolutionary time. It follows, then, that the Basic Property should indeed be very simple. The conclusion conforms to what has been discovered in recent years about the nature of language — a welcome convergence" (Chomsky, 2016).⁴⁴

The autonomy-of-syntax thesis implies that it *had* to originate in a single individual "Prometheus", it is hard to imagine such an abstract computational system" the operation of unbounded Merge" appearing in more than one person at the same time. Lightning does not strike twice. Language can emerge gradually in a group of individuals who share the same communicative goals and intentions, but a content-less abstract grammar cannot.

English linguist Vyvyan Evans drives this point home in his review of *Why Only Us* published in *New Scientist* magazine⁴⁵: "The alternative sees language as an evolutionary

⁴³ This mythical terminology should already warn us of the implausibility of the argument.

⁴⁴ <https://truthout.org/articles/noam-chomsky-on-the-evolution-of-language-a-biolinguistic-perspective/>

⁴⁵ <https://www.newscientist.com/article/2078294-why-only-us-the-language-paradox/>

outcome of a shift in cognitive strategy among ancestral humans, fuelled by bipedalism, tool use and meat-eating. This new bio-cultural niche required a different cognitive strategy to encourage greater cooperation between early humans. Building on the rudimentary social-interactional nous of other great apes, an instinct for cooperation does seem to have emerged in ancestral humans. And this would have inexorably led to complex communicative systems, of which language is the most complete example”.

If, as seen above, the autonomy of syntax implies the existence of UG. The presence of universals in language in the generativist sense forces the appearance of such a system to be relatively recent in evolutionary time, around 150k years ago. The existence of such universals, *pace* Chomsky, is far from being universally accepted by linguists, as illustrated in this famous quote by Martin Joos “languages can differ from each other without limit and in unpredictable ways” (Joos, 1957, p.96).

If the mutation had happened much earlier, natural selection would have had enough time to do bricolage with the newly acquired faculty and then so much for language universals.

The question is, what is the use of an internal capacity for language if you cannot communicate with anyone else? Both because nobody else has the mutation—until your offspring is born—and because the capacity for externalization is not necessarily a concomitant event, in Chomsky’s own words: “language evolved as an instrument of internal thought, with externalization a secondary process” (2016, p.74). Without externalization, the adaptation advantage of an *inner capacity for language* escapes us. Again, in Evans’s words: “The reader is asked to swallow the following unlikely implication of their logic: language didn’t evolve for communication, but rather for internal thought. If language did evolve as a chance mutation, without precedent, then it first emerged in one individual. And what is the value of language as a communicative tool when there is no one else to talk to?”

But assisting thought for what? They were already hunting and gathering very successfully. Is not hunting and gathering what animals—by definition without language—do? The fate of language evolution is left to chance. In fact, Berwick and Chomsky make no bones about it: “one really ought to move from a *gene’s-eye view* to a *gambler’s-eye view*” in order to be a “modern evolutionary theorist” (Berwick and Chomsky, 2016, p.23). We can only echo Einstein’s misgivings about Quantum Mechanics: “God does not play dice”

Again, if syntax were the consequence of a lucky mutation that essentially created the species, it would make sense to affirm that: “There is no evidence of anything like human language, or symbolic activities altogether, before the emergence of modern humans, *Homo Sapiens Sapiens*” (Chomsky 2016). Unfortunately, for Chomskyans this fits ill with recently discovered evidence that “Neanderthals were fully articulate beings” (Dediu, 2018, p.49). This sort of evidence is always indirect because sounds do not fossilize, so we have to rely on proxies for language as indications of it. Dediu and Levinson (2013) showed that there is fossil evidence indicating that “Neanderthals had the modern vocal apparatus, the breathing control, and acoustic sensitivity involved in modern speech” (Dediu and Levinson, 2018, p.52) Dediu pushes back the origin of modern language to at least half a million years, claiming that it is very likely that *Homo heidelbergensis*, the common ancestor of Neanderthals and modern humans already had the capacity. So much for language being “unique to humans in its essential properties”.

The two quotes opening this chapter contain many factual assumptions given as if they were well-established facts, which nonetheless are far from widespread acceptance in the scientific community. My intention was to review what dissident opinions have to say about them.

5.2 Not Only Us

If our language capacity is the result of a miraculous mutation which befell one lucky individual *Prometheus*, about a hundred thousand years ago, who transmitted it to his offspring, then it is a truism that animals, not even apes, cannot speak. The mutation did not happen to them because lightning does not strike twice. The difference between them and us is qualitative, not quantitative.

If, on the other hand, language is understood as the result of a gradual evolution of general cognitive and cultural capacities which fall along a continuum, then it makes sense to look for antecedents in our closest cousin *Pan paniscus*, who may possess those same capacities to a lesser extent. The difference is, in this case, quantitative. Ape Language Research (ALR) looks for the traces of language in apes, particularly bonobos like Kanzi or Matata, and chimpanzees like Sherman and Austin. I will talk about all four of them.

One of the goals of ALR is to help severely retarded children learn a system they can use to communicate with others. The fact that this goal is successfully achieved, is in itself proof enough that ape capacities and human capacities are linked by an Ariadne’s thread.

Still, this does not seem to convince critics of ALR of apes' linguistic abilities. They based their scepticism on failures such as project Nim. Nim Chimpsky was a chimpanzee that was raised pretty much like a human child and taught American Sign Language (ASL). Initially, project Nim appeared to have been a success as Nim was thought to be producing grammatical strings and syntax was supposed to be the yardstick by which linguistic achievement should be measured. However, appearances deceive, and under close inspection Nim's signing was shown to have but the trappings and the suits of language but was not language. Nim was simply mimicking his instructors, who enthusiastically overintepeted Nim's utterances. What a falling-off was there! ALR was all but over. Fortunately, one small village in Armorica—read The Yerkes National Primate Research Centre in Atlanta, Georgia⁴⁶—resisted the dismantling and Susan Savage-Rumbaugh amongst others could continue their research.

Bonobos are much more social than chimpanzees; they are more collaborative, their faces much more expressive, their demeanour sweeter. Their societies are matriarchal⁴⁷, their sexual relations much freer and less possessive than ours⁴⁸. It is hard not to feel their closeness to us. If any ape is going to break the language barrier, the bonobo is that ape.

5.2.1 Matata and Kanzi

When Matata, a female bonobo, arrived at the centre, she was eagerly taught the artificial language Yerkish⁴⁹ but Matata's progress was disappointing to say the least. However, in Matata's classes, an unregistered incognito student was absorbing the contents of the lessons like a sponge. Explicit teaching failed with mature Matata and her already shaped brain; however, implicit learning triumphed with young Kanzi and his still mouldable brain. Age and plasticity did the trick. The third element was fostering Kanzi's talents by sharing with him communicatively relevant situations. Researchers used Yerkish and spoken English combined when addressing Kanzi and results could not have been better.

⁴⁶ Curiously enough, the state bears the foretelling name of a polyglot king, he spoke French, German, English and Italian.

⁴⁷ They are either matriarchal or one must abstain from characterizing them as for gender dominance.

⁴⁸ They seem to ignore the causal relationship between intercourse and offspring.

⁴⁹ Yerkish was developed by German philosopher Ernst von Glasersfeld in the 1970's. It employs a keyboard whose keys contain lexigrams, symbols corresponding to objects or ideas.

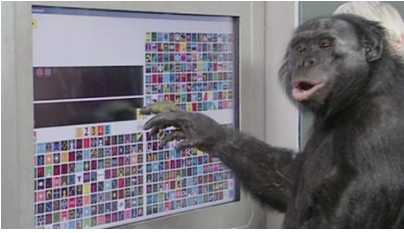


Figure 1 (Kanzi *speaking* Yerkish)

A few years later, it one could safely said that Kanzi had acquired language.

However, sceptics do not rely on testimony; testimony does not count as scientific. Sceptics will only fold under the weight of data.

Sue Savage-Rumbaugh set out to look for that data. She compared the comprehension of spoken English of a two-year old human child, Alia with Kanzi's, who was eight years old at the time of the experiment. They were both exposed to around 600 sentences of seven types, all types were imperatives except for the fourth type.

Type 1: Put object X in/on transportable object Y. *Put the ball on the pine needles.*

Type 2A: Give (or show) or show object X to animate A *Give the lighter to Rose*

Type 2B: Give object X and object Y to animate A *Give the peas and the potatoes to Kelly.*

Type 2C: (Do) action A on animate A *Give Rose a hug.*

Type 2D: (Do) action A on animate A with object X *Get Rose with the snake.*

Type 3: (Do) action A on object X (with object Y) *Knife the sweet potato.*

Type 4: Announce information

Type 5A: Take object X to location Y. *Take the snake outdoors.*

Type 5B: Go to location Y and get object X. *Go to the refrigerator and get a banana.*

Type 5C: Go get object X that's in location Y. *Go get the carrot that's in the microwave.*

Type 6. Make pretend animate A do action A on recipient Y. *Make the doggie bite the snake.*

Type 7. All other sentence types. (from Savage-Rumbaugh 1993)

Many of the instructions were odd so they could not be accused of simply making (un) educated guesses as to what was logical to do with the objects mentioned. To the surprise of sceptics, although not of his caretakers, Kanzi did very well⁵⁰. See Figure 2 with the results.

⁵⁰ In order to avoid Clever Hans effects, Savage-Rumbaugh donned a mask that made visual cues impossible, as shown in figure 3.

TABLE 5
OVERALL PROPORTIONS OF DIFFERENT RESPONSE CODES OBTAINED BY KANZI AND ALIA

RESPONSE	KANZI				ALIA			
	All Trials		Blind Trials		All Trials		Blind Trials	
	N	%	N	%	N	%	N	%
C	369	57	246	59	319	54	220	54
C1	11	2	10	2	20	3	18	4
C2	36	6	16	4	32	6	16	4
C3	34	5	21	5	13	2	9	2
C4	8	1	7	2	7	1	4	1
C5	9	1	7	2	0	0	0	0
Total correct	467	72	307	74	391	66	267	65
FC	153	23	87	21	124	21	84	21
OE	6	1	5	1	20	3	17	4
I	9	1	8	2	9	2	7	2
W	8	1	4	1	32	6	24	6
NR	10	2	4	1	11	2	8	2
Total wrong	186	28	108	26	196	34	140	35
Overall total	653	100	415	100	587	100	407	100

NOTE.—Kanzi had seven mistrials, and Alia had two mistrials. Percentages are rounded to the nearest whole number.

Figure 2 (from

Savage-Rumbaugh, 1993)



Figure 3

Both participants did well overall, scoring around 70 % right, although Kanzi did slightly better than Alia. “The clear outcome from the study is that two normal individuals of different ages and different genera (Homo and Pan) were remarkably closely matched in their ability to understand language” (98). Despite her coming in second, nobody, me included, would deny Alia’s possession of language. Kanzi’s achievement, on the other hand, did not receive with the same recognition.

Perhaps most surprising of all; Kanzi did especially well in type 5C sentences, those including embedding—the generativist touchstone. “The manner in which Kanzi responded

to the type 5C sentence format was most impressive” (89). Thus Kanzi, “was able to comprehend the syntactic relations among word units, not just the units themselves” (90) However, Kanzi’s performance was very weak (33% correct) in type 2B sentences. In coordinated pairs, Kanzi usually missed the second element. Savage-Rumbaugh attributes these mistakes to a shorter attention span or a worse working memory⁵¹, “Kanzi’s difficulty was perhaps due more to short-term memory limitations than to processing limitations” (85) The overwhelming majority of Kanzi’s mistakes were not due to faulty syntax but to a semantic mix-up, like grabbing plaster instead of paint. In a way, he subverted the expectations nativist linguists would have predicted about his performance. He was clearly not acting on rote memorization or on conditioning. He understood words as linguistic symbols and was able to combine them in grammatical constructions.

Ironically enough, nativists (amongst others) —who make so much of the distinction between competence vs. performance, and between I-language vs. E-language (or externalization)—were only too quick to dispose of the possibility of Kanzi real understanding of a natural human language because he was not producing syntactic patterns. Another objection that could be posed to the Kanzi study was that communication was being established between ape and human, not between apes. But that has also happened in Georgia.

5.2.2 Sherman and Austin

Two chimpanzees, Sherman and Austin, also from the Yerkes Centre, learnt to communicate with one another. Not only did they communicate but communicated cooperatively to achieve a goal that was beneficial for both. In this case getting pieces of food.

Sherman and Austin were placed in two separate adjacent rooms from which they could communicate with each other using Yerkish lexigrams. One of them had access to a series of rooms one of which was alternatively baited with food. Each of the rooms could be opened using a specific tool but the chimpanzee who had access to the rooms did not have access to the tools and vice-versa. They had to cooperate to get the food—which once obtained they had to share—and they had to communicate to be able to cooperate. They could communicate very effectively because they had language: they both knew Yerkish.

⁵¹ Apes’ working memory and the so-called cognitive trade-off is currently under research. See for example Matsuzawa (2013).

When this experiment was proved a success, critics said that neither ape had any communicative internal drive; they were just all out for the food and used simple association or conditioning to get it. They were not using lexigrams as semantic symbols but as convenient tactics. To rule this possibility out, they were deprived of the possibility of using Yerkish by turning off the keyboard containing the Yerkish ideograms: “we felt that if they attempted to communicate at all without it, we would have to conclude that they recognized that the reasons underlying the communicative process had to do with differing states of knowledge and the consequent need to share information” (1994, p.86). Experimenters scattered through the floor labels corresponding to the different types of foods. If it were a communicative spirit what drove them, they would use these labels. If they were simply going through the motions, mindlessly following a routine, they would not.

The former was the case; they successfully used the labels and were all the happier, and fuller for it.

In their process of learning Yerkish symbols Sherman and Austin showed to have, whether instinctively or as something learnt, the word learning constraint known as *Mutual Exclusivity*, which amounts to one word per object. What happened was that when presented with a new food for which they had learned no symbol, Sherman used an unassigned symbol in the keyboard, and Austin followed suit. This by itself is a remarkable achievement.

Summing up, apes can learn to communicate using language and treat words as standing for symbolic conceptual categories, rather than a mere behaviouristic association of ideas. Nobody claims that apes can communicate at a human level. Humans play in a different league, but they certainly play the same sport. Now it is up to the nativist linguist to explain away Kanzi’s performance on a par with a human infant or Sherman and Austin’s communicative intentions for the achievement of a mutually beneficial goal.

Why has such an achievement gone practically unacknowledged by the scientific community? The reason is that they were not focused on syntax, as Savage-Rumbaugh points out, “because I had eschewed syntax as a goal in my project, Sherman and Austin received not even as much as a footnote” (Savage-Rumbaugh, 60). Again, the placing of syntax as the autonomous centre of language diminishes cases like this, which was “the first documented case of interindividual communication” (Savage-Rumbaugh, 1994, p.78).

ENVOY

Chomsky's theories of language were irrelevant.

Marvin Minsky.

This dissertation was not about offering new answers, but rather on giving the lie to answers that have been the official truth for decades, despite being plainly false.

We do not have a complete understanding of what language is, but we do know what it is not. Language is not a mathematical formula or constraint that is activated or triggered at a certain age at the infant's brain and is then applied to individual meaningful items. I wish it was that neat, that clean and elegant, but language is messy, fluid and complex.

It cannot be said that children have clear abstract rules even for such constructions as plurals as has been confirmed for the case of adult speakers in foreign language learning contexts. Children need, and have, huge amounts of meaningful data at their disposal to learn to speak, in a process in which words and grammar go hand in hand rather than their separate ways.

The supposed uniquely human essence of language, recursion, is found in other domains and in other species and it is hard to affirm that it came from a miraculous mutation.

We have even seen how apes have, if not language, at least some kind of protolanguage.

Syntax, semantics, phonology, context, all dimensions are connected and as long as they are kept watertight from one another it will be very difficult to make clear progress.

A rich and complex UG has proved to be as problematic as a simple UG.

It is time to explore other avenues, more interdisciplinary and more promising.

WORKS CITED

- Aitchison, J., & Jackendoff, R. (1998). The Architecture of the Language Faculty. *Language*, 74(4), 850.
- Andor, J. (2004). The master and his performance: An interview with Noam Chomsky. *Intercultural Pragmatics*, 1(1).
- Bates, E., & Goodman, J. (1997). On the inseparability of grammar and the lexicon: Evidence from acquisition, aphasia and real-time processing. In G. Altmann (Ed.), *Special issue on the lexicon, Language and Cognitive Processes*, 12(5/6), 507-586.
- Berko, Jean (1958). "The Child's Learning of English Morphology". *WORD*. 14 (2-3): 150-177.
- Berwick, R. C., & Chomsky, N. (2017). *Why only us: Language and evolution*. Cambridge, MA: The MIT Press.
- Bybee, Joan. 2010. *Language, Usage and Cognition*. New York, NY: Cambridge University Press.
- Chomsky, N. (1957). *Syntactic structures*. S-Gravenhage: Mouton.
- Chomsky, N. (2010). Some simple evo devo theses: How true might they be for language? *The Evolution of Human Language*, 45-62.
- Chomsky, N. (2015). *Aspects of the theory of syntax*. Cambridge, MA: The MIT Press.
- Croft, W. (1995). Intonation units and grammatical structure. *Linguistics*, 33(5).
- Dediu D and Levinson SC (2013) On the antiquity of language: the reinterpretation of Neandertal linguistic capacities and its consequences. *Front. Psychol.* 4:397.
- Dediu, D., & Levinson, S. C. (2018). Neanderthal language revisited: Not only us. *Current Opinion in Behavioral Sciences*, 21, 49-55.
- Everett, D. (2005). Cultural Constraints on Grammar and Cognition in Pirahã. *Current Anthropology*, 46(4), 621-646.
- Gentner, T. Q., Fenn, K. M., Margoliash, D., & Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature*, 440(7088), 1204-1207.
- Goad, H., White, L., Steele, J., 2003. Missing inflection in L2 acquisition: defective syntax or L1-constrained prosodic representations? *Canadian Journal of Linguistics* 48, 243-263.

- Harris, R. A. (1995). *The linguistics wars*. New York: Oxford University Press.
- Hauser, M. D. (2002). The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science*, 298(5598), 1569-1579. doi:10.1126/science.298.5598.1569
- Hawkins, R. (2008). The nativist perspective on second language acquisition. *Lingua*, 118(4), 465-477.
- Hilferty, J., (2003). In Defence of Grammatical Constructions
- Hilferty et al. (1998)
- Hogg, R. M., & Alcorn, R. (2012). *An introduction to Old English*. Edinburgh: Edinburgh University Press.
- Jackendoff, R., & Pinker, S. (2005). The nature of the language faculty and its implications for evolution of language (Reply to Fitch, Hauser, and Chomsky). *Cognition*, 97(2), 211-225.
- Joos, M., & Hamp, E. P. (1957). *Readings in linguistics*. Washington.
- Klafehn, T. (2013). Myth of the Wug Test: Japanese Speakers Can't Pass it and English-speaking Children Can't Pass it Either. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society*, 37
- Krashen, S. D. (1995). *Principles and practice in second language acquisition*. New York: Phoenix Elt.
- Lardiere, D., 1998. Case and tense in the 'fossilized' steady state. *Second Language Research* 14, 1–26.
- Piattelli-Palmarini, M. (1983). *Language and learning: The debate between Jean Piaget and Noam Chomsky*. Cambridge, MA: Harvard University Press.
- Pinker, S. (2015). *Words and rules the ingredients of language*. New York: Basic Books.
- Pullum, G. K., & Scholz, B. C. (2002). Empirical assessment of stimulus poverty arguments. *The Linguistic Review*, 18(1-2). doi:10.1515/tlir.19.1-2.9
- Sampson, G. (1980). Popperian Language-Acquisition Undefeated. *The British Journal for the Philosophy of Science*, 31(1), 63-67.
- Savage-Rumbaugh, E. S., Murphy, J., Sevcik, R. A., Brakke, K. E., Williams, S. L., Rumbaugh, D. M., & Bates, E. (1993). *Language comprehension in ape and child*. Chicago: The University of Chicago Press.

- Savage-Rumbaugh, E. S., & Lewin, R. (1994). *Kanzi: The ape at the brink of the human mind*. New York: John Wiley & Sons.
- Stich, S. P. (1981). Can Popperians Learn To Talk? *The British Journal for the Philosophy of Science*, 32(2), 157-164.
- Thomas, M. (2008). Development of the concept of “the poverty of the stimulus”. *The Linguistic Review*, 18(1-2), pp. 51-71.
- Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition*, 74(3), 209-253.