

Syntax Evolution: Problems and Recursion

Ramón Casares

ORCID: [0000-0003-4973-3128](https://orcid.org/0000-0003-4973-3128)

To investigate the evolution of syntax, we need to ascertain the evolutionary rôle of syntax and, before that, the very nature of syntax. Here, we will assume that syntax is computing. And then, since we are computationally Turing complete, we meet an evolutionary anomaly, the anomaly of syntax: we are syntactically too competent for syntax. Assuming that problem solving is computing, and realizing that the evolutionary advantage of Turing completeness is full problem solving and not syntactic proficiency, we explain the anomaly of syntax by postulating that syntax and problem solving co-evolved in humans towards Turing completeness. Examining the requirements that full problem solving impose on language, we find firstly that semantics is not sufficient and that syntax is necessary to represent problems. Our final conclusion is that full problem solving requires a functional semantics on an infinite tree-structured syntax. Besides these results, the introduction of Turing completeness and problem solving to explain the evolution of syntax should help us to fit the evolution of language within the evolution of cognition, giving us some new clues to understand the elusive relation between language and thinking.

Keywords: *syntax evolution, problem solving, Turing completeness, recursion, language and thinking.*

Syntax Evolution: Problems and Recursion

§1 Introduction	3
§1.1 The Anomaly of Syntax	3
§1.2 The Evolution of Syntax	4
§2 Syntax	5
§2.1 Grammar	5
§2.2 Chomsky's Identity	6
§2.3 Decidability	6
§3 Problem Solving	7
§3.1 Turing's Identity	7
§3.2 Turing Completeness	7
§4 Recursion	8
§4.1 Versions of Recursion	8
§4.2 Properties of Recursion	9
§4.3 Recursive Language	10
§4.4 Universal Grammar Is Universal	11
§4.5 The Key	12
§4.6 Human Uniqueness	13
§4.7 Our Hypothesis	14
§4.8 Merge Is Not Recursion	15
§4.9 Chomsky on Syntax Evolution	16
§5 Evolution	17
§5.1 Requirements	17
§5.2 Variables	18
§5.3 Sentence	19
§5.4 Problem	20
§5.5 Routine	21
§5.6 Trial	21
§5.7 Analogy	23
§5.8 Resolution	24
§5.9 Recursivity	25
§5.10 Quotation	26
§5.11 Lisp	26
§5.12 Summary	28
§6 Discussion	29
§6.1 Problems and Evolution	29
§6.2 Syntax and Problems	30
§6.3 Syntax and Evolution	31
§6.4 Beyond the Singularity	32
§7 Conclusion	33
Acknowledgements	34
References	34

§1 Introduction

§1.1 The Anomaly of Syntax

¶1 · Why did only we humans evolve Turing completeness? That is the question which drives my investigations. It is a technical question that assumes that the members of our own species *Homo sapiens* are Turing complete, and that only our species is Turing complete, and then it demands some previous clarifications about Turing completeness and about the assumptions.

¶2 · Turing completeness is the maximum computing power, and computing is a syntactic activity, because computing consists in transforming strings of symbols following syntactic rules. The rules are syntactic because they do not take into account the symbols meanings, nor the computer intentions, but only the symbols themselves and their positions in the strings. So we will start in Section §2 with syntax, taking the main ideas in Chomsky (1959), namely that syntax is computing and that then we should locate natural language syntax in a computing hierarchy. Natural language syntax seems to be mildly context sensitive, but for us here it is enough to acknowledge that it is decidable, as it should be to guarantee that its generation and parsing are free of infinite loops. Summary of §2: syntax is computing, and natural language syntax is decidable.

¶3 · Computing was founded by Turing (1936) to serve as a mathematical model of problem solving. The model is successful because it abstracts away the limitations of a device in memory and speed from its computing capacity, and because we humans exhibit that computing capacity completely. So in Section §3 we will see that problem solving is computing, that Turing completeness is the maximum computing power, and then it is the maximum problem solving power, and that we humans are Turing complete because we can calculate whatever any Turing machine can compute.

¶4 · Computing is just one version of recursion, so in Section §4 we will deal with recursive devices, languages and functions. For us here, only Turing complete devices are recursive, and only the languages used to program Turing complete devices are recursive, while any computable function is recursive. A mathematical fact is that every Turing complete device is undecidable, but we show that the syntax of a recursive language can be decidable. We also show that, to compute any recursive function, the recursive language of the recursive device needs a functional semantics, which is a semantics that can provide the meaning of any recursive function without exceptions. And the characteristic property of Turing complete devices, its full programmability, indicates us that we are the only Turing complete species, because only we can learn any natural or artificial language, as English or LISP.

¶5 · At this point an evolutionary anomaly, which we will call the anomaly of syntax, should be apparent, because it seems that we are computationally, which is to say syntactically, too competent for syntax: Why would evolution select and keep a capacity to generate and parse *any* language, when generating and parsing *one* language would be enough to survive? Why would evolution select and keep our *undecidable* Turing complete syntactic capability to perform a much simpler *decidable* syntax?

¶6 · If syntax does not require Turing completeness, then some other activity should require it. So let us take any possible activity, say vision. But our vision is not better than bonobos vision, and it is less demanding than eagles vision, and therefore vision cannot explain why only we evolved Turing completeness. Then, the activity requiring Turing completeness has to be something that no other species is doing, and speaking a

language with syntax is something that no other species is doing, but then we have to face the anomaly of syntax, because syntax does not require Turing completeness.

¶7 · The anomaly of syntax is a narrow formulation of Wallace’s problem; in his words, “Natural selection could only have endowed the savage with a brain a little superior to that of an ape whereas he possesses one very little inferior to that of an average member of our learned societies”, which we requote from [Bickerton \(2014\)](#), page 1. Most people disregard Wallace’s problem and the anomaly of syntax because otherwise a complex capacity as recursion, also known as Turing completeness, could not have an evolutionary cause. Only the most consequent scientists, as Chomsky, would follow the anomaly implications until its final consequences. As understanding Chomsky’s reasons always sheds light on deep problems, we will try to explain his position on the anomaly of syntax.

¶8 · In our case, we avoid the anomaly of syntax using a coincidence already suggested: syntax is computing, and problem solving is computing, too. Then our computing machinery would have evolved to satisfy requirements coming both from syntax and from problem solving; in other words, syntax and problem solving co-evolved in humans. And though syntax does not require Turing completeness, full problem solving does. Now, we can answer part of our original question: humans evolved Turing completeness to solve as many problems as possible.

§1.2 The Evolution of Syntax

¶1 · If solving as many problems as possible explains why we evolved Turing completeness, and it is so good, why did *only* we humans evolve Turing completeness? Our answer here to this why-*only*-we question will be more tentative and less informative than the answer to the why-we question, and we will address it by proposing how could have been our evolutionary road to recursion under the hypothesis that syntax and problem solving co-evolved in humans towards Turing completeness. So we will start to answer the why-*only*-we question in Section §5 by examining the requirements that full problem solving impose on language.

¶2 · Firstly, we will learn that to express problems we need variables, which are words free of meaning, and sentences, because separate words are not enough. Therefore, semantics is not sufficient and syntax is necessary to represent problems.

¶3 · Then, we will examine the two kinds of conditions that full problem solving impose on language: those relating to data structures, that require an infinite syntax able to deal with hierarchical tree structures, and those other relating to computing capacity, that require Turing completeness. Therefore, our conclusion will be that full problem solving requires a functional semantics on an infinite tree-structured syntax. This way, the full problem solver can represent any problem, and it can calculate any computable function without restrictions, so it can calculate any way of solving problems. A LISP interpreter will serve us to show how, just by meeting the requirements of full problem solving, the result is a recursive device.

¶4 · In the next section, §6, we will discuss the meaning of all that we have seen, so we can try to answer our question. As syntax uses hierarchical tree structures of words, then words are a prerequisite for syntax, and therefore we will assume that the starting point of the evolution of syntax towards Turing completeness was a protolanguage with words but without sentences.

Animal Communication \longrightarrow Protolanguage \longrightarrow Recursive Language

¶5 · Having seen that a full problem solver needs an infinite tree-structured syntax and a functional semantics, we will argue that we evolved syntax before evolving functional semantics. If this is true, then syntax was not the end of the evolutionary road, but the beginning of the second leg. From that point, the road went through all the components that are needed to build a functional semantics, until it finally reached recursion. Investigating the impact of recursion on evolution, we will find that creativity explodes when Turing completeness is achieved causing an evolutionary singularity. So we agree with Chomsky that syntax was instrumental in creating the hiatus that separates our own species from the rest, but for us the cause of the hiatus is Turing completeness.

¶6 · And now we can answer the remaining part of our question: Why did *only* we humans evolve Turing completeness? Because Turing completeness is far away, and only our ancestors took each and every step of the long and winding road from animal communication to recursion. This answer is not as tautological as it seems to be: in order to be Turing complete, a species has to invent the word, and then it has to invent the sentence, and yet it is not recursive if it cannot invent a functional semantics, too. And only a cooperative species needs a language, as Tomasello (2008) argues, so it seems that ours was the only species that was complex enough and cooperative enough to evolve a recursive language.

§2 Syntax

§2.1 Grammar

¶1 · Chomsky (1959) presents a hierarchy of grammars. A *grammar* of a language is a device that is capable of enumerating all the language sentences. And, in this context, *language* is the (usually infinite) set of all the valid syntactic sentences.

¶2 · At the end of SECTION 2 in that paper, page 143, we read: “A *type 0 grammar* (*language*) is one that is unrestricted.” ¶ Type 0 grammars are essentially Turing machines”. At the beginning of SECTION 3, same page, we find two theorems.

THEOREM 1. For both grammars and languages, type 0 \supseteq type 1 \supseteq type 2 \supseteq type 3.

THEOREM 2. Every recursively enumerable set of strings is a type 0 language (and conversely).

Then THEOREM 2 is explained: “That is, a grammar of type 0 is a device with the generative power of a Turing machine.”

¶3 · From the two theorems we can deduce three corollaries.

COROLLARY 1. The set of all type 0 grammars (languages) is equal to the set of all grammars (languages).

This is because, according to THEOREM 1, type 0 is the superset of all grammars (languages), and more generally because type 0 is unrestricted.

COROLLARY 2. For each Turing machine there is a type 0 grammar (and conversely).

This is equivalent to THEOREM 2, but in terms of grammars (devices) instead of languages (sets).

COROLLARY 3. For each Turing machine there is a grammar (and conversely).

This results by applying COROLLARY 1 to COROLLARY 2.

§2.2 Chomsky's Identity

¶1 · The third corollary shows that ‘Turing machine’ and ‘grammar’ are equivalent devices, by which ‘computing’ and ‘syntax’ are equivalent concepts. Computing is equivalent to syntax, rather than to language, because language in [Chomsky \(1959\)](#) refers only to syntax; it does not refer to semantics, because meanings are not considered, nor to pragmatics, because intentions are not considered. Summarizing it, we get Chomsky's identity between syntax and computing.

$$\text{Syntax} = \text{Computing}$$

¶2 · So now we will provide the definition of syntax that follows from [Chomsky \(1959\)](#) after adjusting for this deviation towards language: *syntax* consists of transformations of strings of symbols, irrespective of the symbols meanings, but according to a finite set of well defined rules; so well defined as a Turing Machine is. This definition of syntax is very general and includes natural languages syntax, just replacing symbols with words and strings with sentences, but it also includes natural languages morphology or phonology, taking then sub-units of words or sounds.

¶3 · As an aside, please note that this definition of syntax manifests, perhaps better than other definitions, what we will call the *syntax purpose paradox*: syntax, being literally meaningless, should be purposeless. It is easy to underestimate the value of some mechanical transformations of strings of symbols.

§2.3 Decidability

¶1 · Concerning decidable languages, [Chomsky \(1959\)](#) states the following theorem, which is also in page 143.

THEOREM 3. Each type 1 language is a decidable set of strings.

But not conversely, as note 7a adds. Therefore, $\{\mathcal{T}_1\} \subset \{\mathcal{L}_D\}$.

¶2 · The hierarchy of [Chomsky \(1959\)](#) can be extended to include mildly context sensitive languages in it. We follow [Stabler \(2014\)](#), see **Theorem 1** in page 167, to locate mildly context sensitive languages in the hierarchy.

Chomsky	Language	Device
Type 0	\mathcal{T}_0 Unrestricted	Turing machine
—	\mathcal{L}_D Decidable	—
Type 1	\mathcal{T}_1 Context sensitive	Linear bounded automaton
—	\mathcal{L}_M Mildly context sensitive	—
Type 2	\mathcal{T}_2 Context free	Push-down automaton
Type 3	\mathcal{T}_3 Regular	Finite state automaton

The extended hierarchy is then: $\{\mathcal{T}_3\} \subset \{\mathcal{T}_2\} \subset \{\mathcal{L}_M\} \subset \{\mathcal{T}_1\} \subset \{\mathcal{L}_D\} \subset \{\mathcal{T}_0\}$.

¶3 · The languages defined by minimalist grammars are mildly context sensitive, and then it seems that natural language is in this class, as argued by [Stabler \(2014\)](#). This would guarantee that natural language generation and parsing are free of infinite loops, as they should be, because mildly context sensitive languages are decidable, $\{\mathcal{L}_M\} \subset \{\mathcal{L}_D\}$.

§3 Problem Solving

§3.1 Turing's Identity

¶1 · The connection between computing and problem solving goes back to the very foundation of computing. **Turing (1936)** defines his machine in order to prove that the *Entscheidungsproblem*, which is the German word for ‘decision problem’, is unsolvable. After defining the Turing machine, he shows that there is not any Turing machine that can solve the problem. This proof is valid only under the assumption that the set of Turing machines exhausts the ways of solving problems, where each Turing machine is a way of solving, because then that no Turing machine solves a problem implies that there is no way of solving it. Summarizing it, we get Turing's identity between problem solving and computing.

$$\text{Problem Solving} = \text{Computing}$$

¶2 · It is even more explicit **Post (1936)**, who writes that he has in mind a *general problem*, and then he goes on to define a ‘problem solver’ which is very similar and equivalent to a Turing machine.

¶3 · From then, the identity between computing and problem solving holds, even if it is not expressed. For example, for **Simon & Newell (1971)** a theory of human problem solving gets credit only if a computer program simulates precisely the behavior of a person solving the same problems.

¶4 · In Subsection §2.2, we saw that syntax is computing, and now we see that problem solving is computing, too. That both syntax and problem solving are computing is a coincidence that asks for an explanation that we will postpone until §6.2.

$$\text{Syntax} = \text{Computing} = \text{Problem Solving}$$

§3.2 Turing Completeness

¶1 · After assimilating that problem solving is computing, the next step is to assume that solving more problems provides more survival possibilities, and therefore that the evolution of computing was driven by the evolution of problem solving, see **Casares (T)**. In addition, if problem solving is computing, then the maximum problem solving power is the maximum computing power, and therefore the last step in this evolutionary thread is to achieve the maximum computing power, which is to achieve Turing completeness.

¶2 · By definition, a device is *Turing complete* if it can compute anything that any Turing machine can compute. More informally, a device is Turing complete if it can be programmed to do any computation, and then ‘Turing completeness’ is a technical phrase for ‘full programmability’.

¶3 · The Turing machine, as it was presented by **Turing (1936)** himself, models the calculations done by a person with a finite internal memory who can access as much external memory as he needs, meaning that we persons can compute whatever a Turing machine can compute provided that we can access as much external memory as we need and that we have enough time to accomplish the computation. These two conditions refer to memory access and to available time, and they do not refer to computing capacity, and therefore we are Turing complete in computing capacity. In other words, we are Turing complete because we can compute any program, that is, because we can follow any finite set of well-defined rules.

¶4 · That we are Turing complete is the fundamental point made by Turing (1936). For example, according to Davis (1982), only Turing's model convinced Gödel that computing exhausts what is effectively calculable.

¶5 · Turing completeness is the maximum syntactic power. This follows directly from the identity between syntax and computing seen in Subsection §2.2. We will call a syntax generated by a Turing complete grammar a *complete syntax*.

¶6 · The prototype of Turing complete device is a universal Turing machine, also defined by Turing (1936), but there are others, each one associated with one version of recursion.

§4 Recursion

§4.1 Versions of Recursion

¶1 · There are three classical versions of recursion:

- Recursive functions, due to Gödel, Herbrand, and Kleene, where a recursive function can be defined without reference restrictions, and, in particular, the definition of a function can refer to itself. Its classical reference is Kleene (1936a).
- Lambda functions, due to Church, Kleene, Rosser, and Curry, which are higher-order functions, that is, functions on functions to functions. In lambda-calculus everything is a function, or a variable. Its classical reference is Church (1935).
- Computable functions, due to Turing, and Post, that are ruled transformations of strings of symbols. There is a universal computer for which any computer is a string of symbols. Its classical reference is Turing (1936).

There are now more versions of recursion, but here we will manage with these three.

¶2 · Although apparently very different to each other, the three are equivalent:

- Kleene (1936b) showed that every lambda function is recursive, and the converse. This means that for each lambda function there is a recursive function that performs the same calculation, and the converse, and therefore lambda functions are equivalent to recursive functions.
- Turing (1937) showed that every lambda function is computable, and that every computable function is recursive. This means that for each lambda function there is a Turing machine that performs the same calculation, and that for each Turing machine there is a recursive function that performs the same calculation, and therefore, together with Kleene (1936b), computable, lambda, and recursive functions are equivalent.

Computable Function = Lambda Function = Recursive Function

¶3 · These mathematical equivalences show that we can implement recursion in several ways. Each implementation has its own advantages and disadvantages, and then they are not completely equivalent from an engineering point of view. So let us now introduce the devices that implement these kinds of functions:

- The universal Turing machine, already seen in §3.2, is a Turing complete device because it can compute whatever a Turing machine can compute, and then it can calculate any computable function. And also, by the equivalences between functions, a universal Turing machine can calculate any lambda and any recursive function.
- By definition, a lambda-calculus interpreter can calculate any lambda function, and then, by the equivalences, it can calculate any computable and any recursive function.

- Finally, an ideal (error free and eternal) mathematician can calculate any recursive function, and then any lambda and any computable function. This implementation of genuine recursion can be called formal mathematics.
- ¶4 · The precise conclusion is that whatever a universal Turing machine can calculate, it can also be calculated by a lambda-calculus interpreter, and by a mathematician using recursive functions, and the converse of both. Therefore, we say that a universal Turing machine, a lambda-calculus interpreter, a mathematician calculating recursive functions, or any other mathematically equivalent calculating device, is a *universal computer*. Only universal computers implement recursion. Then ‘universal computer’ is synonymous with ‘Turing complete device’, and also with ‘recursive device’.

Universal Computer = Turing Complete Device = Recursive Device

§4.2 Properties of Recursion

- ¶1 · Recursive devices have some properties in common: incompleteness, extendability, and undecidability, all three summarized by [Post \(1944\)](#). The three properties are related, but here we are most interested in one: undecidability. A calculating device is *decidable* if it can resolve, usually with a yes-or-no answer, in every case, that is, for any finite input data. Otherwise, it is undecidable. Then an undecidable device can get stuck in an infinite loop, never ending a calculation. As we can program a Turing complete device to do any computation, we can program it to do an infinite loop, and therefore every recursive device is undecidable.
- ¶2 · Note that these properties are properties of devices, and not properties of functions. A recursive function can be decidable, but all recursive devices are undecidable. This can be a source of confusion, because only the most capable devices are recursive, while even the simplest functions are recursive. An example of simple recursive function is the identity function that just returns what it takes, and then it is a ‘no operation’ instruction that, in fact, does not compute anything. To prevent this confusion, except when we refer explicitly to recursive functions, we will always use recursion, or recursive, to refer to the property of devices.
- ¶3 · Warning note: We are using here the most exacting definition of recursion possible, equating recursion to Turing completeness. Though this severe definition of recursion prevents confusion and succeeds where other definitions fail, mainly in discriminating human versus non-human capacities, see [Casares \(R\)](#), less demanding definitions of recursion pervade linguistics, see [Stabler \(2014\)](#) and [Watumull et al. \(2014\)](#), for example. Different purposes demand different definitions, but this should not be a problem provided we are always aware of which definitions are ruling. And here, recursion is Turing completeness.
- ¶4 · The existence of different implementations of recursion can be another source of confusion. Although from a mathematical point of view every computable function is recursive, and the converse, it is sometimes said that a function is recursive when it was implemented using self-reference, but not when the same function was implemented using iteration, for example. Here, we will avoid that loose way of speaking, because, for example, the identity function, which does not refer to itself, is nevertheless genuinely recursive, see [Kleene \(1936a\)](#), page 729, where it is called U_1^1 .
- ¶5 · So recursive devices are undecidable, and then if you can prove that a device, whatever the arguments it takes, always completes the calculations in a finite time, then you have

proved that the device does not implement recursion, or, in other words, that it is not Turing complete. The converse is not true; that a computation does not complete does not imply that the computer is recursive.

§4.3 Recursive Language

¶1 · A recursive device can calculate any recursive function, but, to do it, it needs the expression of the function. Then, each universal computer understands a way to express any possible recursive function. We will call this language that every Turing complete device needs to express any recursive function a *recursive language*. In other words, the recursive language is the language used to program the universal computer. Note that, in Casares (H) and everywhere else, this kind of language is called a *Turing complete language*, or a *complete language* \mathcal{L}_C , for short.

¶2 · For example, a typical lambda-calculus interpreter understands a recursive language Λ with a syntax defined inductively this way:

$$\begin{aligned} x &\in V & V &\subset \Lambda \\ x &\in V \Rightarrow x' &\in V \\ x &\in V, M \in \Lambda \Rightarrow (\lambda x M) &\in \Lambda \\ M, N &\in \Lambda \Rightarrow (M N) &\in \Lambda \end{aligned}$$

This means that in lambda-calculus there are only: an infinity of variables built by priming, which belong to the subset V ; lambda abstractions, that is, function definitions, $(\lambda x M)$; and function applications, $(M N)$. These last two are ordered pairs that grow into binary trees by induction.

¶3 · Other universal computers use other syntaxes, so implementing binary trees is not a requirement for recursion. For example, a typical universal Turing machine understands the transition table of the Turing machine to emulate. That table is a list of conditionals: if the current internal state is q_i , and the symbol read is s_r (s_r can also be *blank*), then go to internal state q_n (q_n can be equal to q_i), write symbol s_w (s_w can be equal to s_r or not, even *blank*), and move to the *right*, or to the *left*, or *halt*. In fact, as explained in Casares (U), the only syntactic requirement that any recursive language has to fulfill is that the set of its syntactic objects has to be infinite enumerable or bigger, that is, any infinite set complies.

¶4 · While these Turing complete devices implement complete syntaxes, as defined in §3.2, they need also semantics. Syntax alone is not enough, and some semantics are needed. It is not enough that the expressions represent the functions, the Turing complete device has to understand what the expressions mean to produce the results that the represented functions produce. Here, we will call any semantics providing the meaning of recursive functions a *functional semantics*. And then, a functional semantics is needed to implement a complete syntax!

¶5 · For example, the typical lambda-calculus interpreter has to know how to perform β -reductions, see Curry & Feys (1958), §3D, as for example $((\lambda x x) x') \rightarrow x'$, in order to be able to calculate any lambda function. Then, β -reduction is part of the functional semantics that makes this interpreter a recursive device. Note that $(\lambda x x)$ is the identity function, $U_1^1 = (\lambda x x)$.

¶6 · Complete syntax is synonymous with recursion, and a somewhat paradoxical property of recursion is apparent when we call it complete syntax. Complete syntax is syntax,

as defined in §2.2, but it is also more than syntax, because complete syntax requires a functional semantics. To solve the paradox, we will consider functional semantics part of semantics and also part of syntax, because functional semantics only provides the meanings of functions, and functions are the core of syntax. This way, functional semantics is in the intersection of semantics and syntax.

§4.4 Universal Grammar Is Universal

¶1 · Decidability partitions the set of grammars, and the set of the corresponding languages, in two subsets. For languages we get: $\{\mathcal{L}_D\} \cap \{\mathcal{L}_{\bar{D}}\} = \emptyset$ and $\{\mathcal{L}_D\} \cup \{\mathcal{L}_{\bar{D}}\} = \{\mathcal{T}_0\}$. The hierarchy of Chomsky (1959) is in the decidable part of $\{\mathcal{T}_0\}$, see §2.3, while this other hierarchy is in the undecidable part of $\{\mathcal{T}_0\}$. In Casares (H), both parts are combined and extended into a complete hierarchy.

Chomsky Language Device

Type 0	\mathcal{T}_0	Unrestricted	Turing machine
—	$\mathcal{L}_{\bar{D}}$	Undecidable	—
—	\mathcal{L}_C	Recursive	Universal Turing machine

The undecidable hierarchy is then: $\{\mathcal{L}_C\} \subset \{\mathcal{L}_{\bar{D}}\} \subset \{\mathcal{T}_0\}$.

¶2 · But this hierarchy is misleading because of the peculiar nature of universal Turing machines. A single universal Turing machine is as powerful as the set of all Turing machines, but it needs a recursive language to program it. Then, we should distinguish three languages when dealing with universal computers: the recursive language $L_C[L_S]$ that uses a syntax L_S to describe the programmed language L_P . We can write this as $L_C[L_S](L_P)$, meaning that $L_C[L_S] \in \{\mathcal{L}_C\}$ is implemented in the universal computer hardware, and (L_P) , where $L_P \in \{\mathcal{T}_0\}$, is programmed software. Then, in our case, the universal grammar of Chomsky (1965) implements $L_C[L_S]$, and L_P is any language that we can learn, say English or LISP.

¶3 · The equation of the recursive languages (1) is a consequence of the Turing completeness of its grammar, and it states that we can produce any language with a recursive language. In mathematical terms,

$$(1) \quad \forall L_P \in \{\mathcal{T}_0\}, \forall L_C[L_S] \in \{\mathcal{L}_C\}, \quad L_C[L_S](L_P) = L_P.$$

¶4 · The equation of the recursive languages (1) explains why we can learn any natural or artificial language. And it also explains why the universal grammar does not set any specific language. So it is not because the universal grammar is underspecified, as proposed by Bickerton (2014), but because the universal grammar is universal, in the sense that it is a universal computer. Then, the universal grammar is overqualified, rather than underspecified. This subject is further elaborated in Casares (U).

¶5 · Let us now differentiate the two components L_C and L_S of the recursive language $L_C[L_S] \in \{\mathcal{L}_C\}$. First, we will see that L_S can be a decidable language, that is, that $L_S \in \{\mathcal{L}_D\}$ is possible. For example, the syntax of the recursive language of the typical lambda-calculus interpreter is a context free language that uses six terminals, $V_T = \{x, ', (, \lambda, \sqcup,)\}$, where \sqcup is a space, two non-terminals, $V_N = \{S, X\}$, and that obeys the following five generative rules.

$$\begin{aligned} S &\rightarrow X \mid (\lambda X \sqcup S) \mid (S \sqcup S) \\ X &\rightarrow x \mid X' \end{aligned}$$

Context free languages generate trees, but the syntax of a recursive language can be simpler. For example, the syntax of the recursive language of the typical universal Turing machine is a regular language, because any Turing machine can be expressed as a list of five-word clauses $QYQYZ$, where $Q = \{q_0, q_1, \dots, q_Q\}$ is the finite set of internal states, $Y = \{\text{blank}, s_1, \dots, s_Y\}$ is the finite set of symbols, and $Z = \{\text{left}, \text{right}, \text{halt}\}$ is the set of movements, and then its syntax is the Kleene closure of the five-word clauses, $(QYQYZ)^*$. Second, we know that any recursive language $L_C[L_S]$ is necessarily undecidable because it can be programmed to generate any language, including undecidable languages. This shows that

$$(2) \quad \exists L_S \in \{\mathcal{L}_D\}, \quad L_C[L_S] \in \{\mathcal{L}_C\} \subset \{\mathcal{L}_{\bar{D}}\}.$$

¶6 · In those cases where $L_S \in \{\mathcal{L}_D\}$, only functional semantics can provide the resources needed to achieve universality, so we will distinguish the proper syntax defined by L_S from the functional semantics that we ascribe to L_C . In other words, the recursive language $L_C[L_S]$ of the recursive device has a syntax L_S and a semantics L_C . Equation (1) when $L_S \in \{\mathcal{L}_D\}$ shows that it is possible to program any language L_P using a decidable syntax L_S and a suitable functional semantics L_C . In mathematics we say that from (1) and (2) we deduce (3):

$$(3) \quad \forall L_P \in \{\mathcal{T}_0\}, \quad \exists L_S \in \{\mathcal{L}_D\}, \quad L_C[L_S](L_P) = L_P.$$

The conclusion is that the syntax of a recursive language can be decidable. Then it should be decidable. And in our case, it is decidable, or so it seems as seen in §2.3.

§4.5 The Key

¶1 · The equation of the recursive languages (1) shows that to implement one specific language we can either build a recursive device and then program the specific language in its recursive language, or instead build the specific grammar for the language. To note that the second is much easier than the first is to rephrase the anomaly of syntax.

¶2 · On the other hand, once a recursive device is available, it would not be wise to limit it. Then every particular natural language should be a recursive language with its functional semantics. We are saying that, when the L_P of the equation (1) is a particular natural language, then $L_P \in \{\mathcal{L}_C\}$ should be the case. We know it is possible, because L_P can be any language, and failing L_P to be recursive would imply that we could not express and mean every possible recursive function in that particular natural language L_P .

¶3 · Is there any particular natural language that is not recursive? [Everett \(2008\)](#) claims that the Pirahã language lacks recursion. I don't know, but if it is true, then you cannot translate a manual of LISP from English to Pirahã, as for example [McCarthy et al. \(1962\)](#).

¶4 · Another simpler test concerns logical paradoxes. Logical paradoxes lay completely on functional semantics, not needing any other semantics. For example, take the English sentence (4).

(4) This sentence is false.

The syntactic analysis of (4) is straightforward, but when we look for its meaning using English functional semantics we get trapped in an infinite loop. Now, the test: If it is not possible to translate this sentence (4) to Pirahã, then Pirahã is not recursive. This

test is easier than that of translating a LISP manual, but it is not as conclusive, because if the translation of the sentence (4) were possible, then nothing would be settled.

¶5 · Three notes before leaving the Pirahã language.

- Although unexpected, a non-recursive particular natural language is possible, because the L_P of the equation of the recursive languages (1) can be any language.
- The syntax of a recursive language does not need to use tree data structures, as shown by the syntax of the typical universal Turing machine recursive language, which only uses lists of fixed length clauses, so the Pirahã language can still be recursive. The question is whether the functional semantics that complement the simple Pirahã syntax is enough or not.
- Even if the Pirahã language were not recursive, the Pirahã speakers could still learn English, and then they would understand sentence (4) and a LISP manual in English.

¶6 · Going back to our main business, we were seeing that, while any recursive language is undecidable, its syntax can be, and should be, decidable. This explains an otherwise paradoxical fact: our syntactic abilities are in excess to those needed to speak a natural language. We are Turing complete, which is the maximum syntactic capacity, as required for better problem solving, though the natural language syntax parsers are not Turing complete, to assure parsing completion. In other words, recursion is not a requirement of natural language syntax, and being recursion a syntactic property, we conclude that recursion is not a requirement of language.

¶7 · We will repeat it just because this is the key of the paper; once this is assimilated, everything else is pretty straightforward. Both syntax and problem solving are computing, but while recursion, which is the property that signals the maximum computing power, is detrimental to any syntax, even to the syntax of a recursive language, the very same recursion is the most desirable property for problem solving.

§4.6 Human Uniqueness

¶1 · Because parsing tree data structures does not require Turing completeness, then a species that is not Turing complete can use tree data structures. On the other hand, it seems that we are the only Turing complete species, that is, the only one implementing recursion.

¶2 · To prove that a device is Turing complete you have to show that it can compute any recursive function. Therefore, to be qualified as Turing complete, a device has to take the expression of any recursive function in some language, and any data, and then return the right result every time. This implies that a recursive species needs a recursive language. The only known species with a recursive language is ours, and then it seems that we are the only recursive species.

¶3 · In any case, the equation of recursive languages applies to any Turing complete device, and this implies that the members of any recursive species, whether they use a recursive language to communicate or not, can learn any language, as it happens to the Pirahã speakers independent of the recursivity of their language. We are saying that the members of any recursive species can learn English, LISP, and in fact any language, and that that capability is independent of the interfaces, a conclusion that explains why our spoken language can be written without any further evolutionary intervention.

¶4 · If we are in fact the only Turing complete species, then recursion is uniquely human. It is also true that any human organ is also uniquely human, that is, the human eye is not like any other, and an anthropologist can distinguish a human bone from any non-human

bone. But perhaps recursion is different, because it is a property that our species *Homo sapiens* has, and every other species has not, and because recursion is a singular property, as explained below in Subsection §6.1.

¶5 · If recursion is a requirement of problem solving, and by this we mean that recursion has some survival value for problem solving, which is very likely true because of the recursive nature of problems (see Section §3), then the hypothesis that ‘what is uniquely human and unique to the faculty of language —the faculty of language in the narrow sense (FLN)— is recursion’, by Hauser, Chomsky, and Fitch (2002), is only partly true. While our species seems to be the only Turing complete one, the recursive machinery is being used not only by language but also by problem solving.

¶6 · Here, instead of the complete and ambiguous formulation of the hypothesis in page 1573, “FLN comprises only the core computational mechanisms of recursion as they appear in narrow syntax and the mappings to the interfaces”, we are using the simpler formulation ‘FLN is recursion’ for three reasons:

- ‘FLN is recursion’ is shorter and it was used twice by Hauser, Chomsky, and Fitch (2002), in the abstract and in the very same paragraph where the longer one is.
- Language can be spoken, written, and signed, and then it should be independent of the interfaces, or else it had to evolve three times.
- The mappings to the interfaces are irrelevant here, because our point is that recursion is not used only by language, and this negates both formulations of the hypothesis.

§4.7 Our Hypothesis

¶1 · Even if the ‘FLN is recursion’ hypothesis is not completely true, it touches one of the most important questions about language and syntax: their relation to recursion. To fix that hypothesis, we dare to state a new one that takes into account problem solving. Our hypothesis is that ‘syntax and problem solving co-evolved in humans towards Turing completeness’.

¶2 · The argument goes like this: Solving more problems provides more survival opportunities, and then, because of the recursive nature of problems, this thread of evolution goes naturally towards Turing completeness (Section §3). And Turing completeness requires a recursive language in which to express and mean any recursive function (Subsection §4.3). Then our syntax should have evolved to satisfy the requirements of problem solving.

¶3 · To be clearer, we agree with Chomsky and his ‘FLN is recursion’ hypothesis in the following points:

- we are the unique species with a recursive language,
- we are the unique recursive species,
- recursion is a syntactic property, and
- recursion is not a requirement of syntax, so it is not a requirement of language.

But then Chomsky does not want to go further. In fact, he cannot go further because, if recursion is not a requirement of language, meaning that recursion has not survival value for language, and recursion is exclusively a language property, as stated by his ‘FLN is recursion’ hypothesis, then recursion cannot have an evolutionary explanation. In the next subsections, §4.8 and §4.9, we will try to understand why Chomsky reached this conclusion.

¶4 · Additionally, for us,

- recursion is a requirement of full problem solving,

and then recursion can have an evolutionary explanation, and we can state our hypothesis: ‘syntax and problem solving co-evolved in humans towards recursion’.

§4.8 Merge Is Not Recursion

¶1 · Chomsky (2007), page 20, is wrong when he writes: “If the lexicon is reduced to a single element, then unbounded Merge will easily yield arithmetic.” The offending word is ‘easily’. The definition of Merge, in the same page, is:

$$\text{Merge}(X, Y) = \{X, Y\} .$$

Chomsky (2006), page 184, states it more precisely: “The most restrictive case of Merge applies to a single object, forming a singleton set. Restriction to this case yields the successor function, from which the rest of the theory of natural numbers can be developed in familiar ways.” The theory of natural numbers is also known as arithmetic.

¶2 · Let us call the single element of the lexicon \emptyset . Then,

$$\text{Merge}(\emptyset) = \text{Merge}(\emptyset, \emptyset) = \{\emptyset, \emptyset\} = \{\emptyset\} ,$$

which is indeed a singleton set. Reiterating we obtain: $\text{Merge}(\text{Merge}(\emptyset)) = \text{Merge}(\{\emptyset\}) = \{\{\emptyset\}\}$, $\text{Merge}(\text{Merge}(\text{Merge}(\emptyset))) = \text{Merge}(\{\{\emptyset\}\}) = \{\{\{\emptyset\}\}\}$, and so on. In the construction of the natural numbers by Zermelo (1908): $0 = \emptyset$, $1 = \{\emptyset\}$, $2 = \{\{\emptyset\}\}$, $3 = \{\{\{\emptyset\}\}\}$, and so on. Now, rewriting, $\text{Merge}(0) = 1$, $\text{Merge}(1) = 2$, $\text{Merge}(2) = 3$, $\text{Merge}(n) = n + 1$, and then Merge is indeed the successor function. So far, so good.

¶3 · But the successor function is not enough to develop the theory of natural numbers. This question was answered by Kleene (1952) in his exhaustive investigation of the genuine recursion of Gödel. The successor function is only Schema I, the simplest of the six (I-VI) schemata that are needed to define any recursive function of natural numbers, see pages 219 and 279. For example, Schema III are the identity functions, U_i^n , which are also needed to implement arithmetic.

¶4 · The lambda version of recursion uses a version of Merge that we will call **cons**. As Merge, **cons** takes two arguments and returns an object that contains the arguments, and the only difference is that the result of **cons** is an ordered pair, instead of a set:

$$\text{cons}(X, Y) = (X_{\sqcup} Y) .$$

By iterating uses of **cons**, we can build binary trees with left and right branches, while using Merge we cannot distinguish branches. Again, **cons** alone is not enough to implement arithmetic in the lambda version of recursion, because β -reductions and variables are also needed.

¶5 · This is as it should be. Merge is decidable, that is, a device implementing only Merge will always complete its calculations in a finite time, and therefore a device implementing only Merge is not recursive. By the way, the same happens to **cons** and to the successor function. It is as it should be because, being decidable, Merge meets the requirement of syntax parsers, and then Merge is well suited to be the main operation of syntax, as proposed by the minimalist program.

¶6 · On the other hand, arithmetic is undecidable. This is arguably the most fundamental truth of mathematics: in any finitary formalization of arithmetic that is consistent there

are arithmetic statements that cannot be decided true or false within the formalization, proved by Gödel (1930) for genuinely recursive formalisms, by Church (1935) for lambda formalisms, and by Turing (1936) for computable formalisms, where a finitary formalization is an implementation. Recursion was the result of this paramount investigation on the foundations of mathematics.

¶7 · The chasm that separates the single successor function from a full implementation of recursion shows us the anomaly of syntax from a new perspective. The most we could say would be that Merge is required to implement recursion, but even this is not completely exact. A version of Merge, **cons**, is needed to implement the lambda version of recursion, and another version of Merge, the successor function, is needed to implement the genuine version recursion, but the computable version of recursion only requires the conditional, and some auxiliary operations not related to Merge, to implement its functional semantics. In summary: Merge is not recursion, it is only an element that can be used to implement recursion. And, in any case, the way from Merge to recursion is not easy; it takes some steps, and not all are easy.

§4.9 Chomsky on Syntax Evolution

¶1 · Now, we will try to understand Chomsky on the evolution of syntax and recursion, under the light of the anomaly of syntax.

¶2 · Presenting the ‘FLN is recursion’ hypothesis, Hauser, Chomsky, and Fitch (2002) wrote: “If FLN is indeed this restricted, this hypothesis has the interesting effect of nullifying the argument from design, and thus rendering the status of FLN as an adaptation open to question”, page 1573.

¶3 · Pinker & Jackendoff (2004) disagreed: “Our disagreement specifically centers on the hypothesis that recursion is the only aspect of language that is special to it, that it evolved for functions other than language, and that this nullifies ‘the argument from design’ that sees language as an adaptation”, page 205. Then, in pages 216–217, arguing “that unlike humans, tamarins cannot learn the simple recursive language $A^n B^n$ ”, Pinker & Jackendoff state the anomaly of syntax: “If the conclusion is that human syntactic competence consists only of an ability to learn recursive languages (which embrace all kinds of formal systems, including computer programming languages, mathematical notation, the set of all palindromes, and an infinity of others), the fact that actual human languages are a minuscule and well-defined subset of recursive languages is unexplained.”

¶4 · It was in fact unexplained by Pinker & Jackendoff, and also by Fitch, Hauser, and Chomsky (2005) in their answer, who on this point wrote: “Fitch & Hauser do not even mention recursion in the cited paper, and the generation of limited-depth hierarchical phrase structure was not confused with recursion in that paper (although it was by some commentators on the article)”, page 204. Of course, they are twice right: recursion is frequently confused, and even unbounded $A^n B^n$ is not a recursive language because it is not possible to calculate every recursive function in a language that has not a functional semantics. The salient point, however, is that they wrote nothing about the anomaly.

¶5 · But Chomsky (2006) himself was indeed aware of the anomaly of syntax. We copy here the complete paragraph between pages 184 and 185:

The most restrictive case of Merge applies to a single object, forming a singleton set. Restriction to this case yields the successor function, from which the rest of the theory of natural numbers can be developed in familiar ways. That suggests a possible answer to a problem that troubled Wallace in the late nineteenth century:

in his words, that the “gigantic development of the mathematical capacity is wholly unexplained by the theory of natural selection, and must be due to some altogether distinct cause,” if only because it remained unused. One possibility is that the natural numbers result from a simple constraint on the language faculty, hence not given by God, in accord with Kronecker’s famous aphorism, though the rest is created by man, as he continued. Speculations about the origin of the mathematical capacity as an abstraction from linguistic operations are not unfamiliar. There are apparent problems, including dissociation with lesions and diversity of localization, but the significance of such phenomena is unclear for many reasons (including the issue of possession vs. use of the capacity). There may be something to these speculations, perhaps along the lines just indicated.

Perhaps not, but it seems that for Chomsky our mathematical capacity, which includes recursion, “is wholly unexplained by the theory of natural selection, [...] if only because it remained unused.”

¶6 · Our conclusion is that Chomsky took the anomaly of syntax as a proof that recursion was not selected, and then he proceeded consequently. For Chomsky (2005a), not too long before about 50 000 years ago, a mutation rewiring the brain took place yielding the operation of Merge. Because of some causes not related to natural selection, also called third factor causes by Chomsky (2005b), Merge had the effect of producing *easily* the complete recursion; but remember §4.8. Then, according to Chomsky (2000), page 163, where he argues that selection is “[a] factor, not *the* factor”, recursion could be the emergent product of the spontaneous self-organization of complex systems under the strictures of physical law. In summary: Chomsky had to invoke third factor causes to explain recursion, instead of invoking evolutionary or first factor causes, because he took the anomaly of syntax at face value and he is consequent.

¶7 · Our position differs from that of Chomsky: we can explain the survival value of recursion, and then recursion is not an incidental result of language. Under our hypothesis, recursion was selected because of its problem solving advantages, despite its syntactic inconvenience. Then, Wallace problem is not a problem, and we were using recursion for solving problems since we achieve Turing completeness, though its mathematical formalization came only recently. And, for us, see §4.8, the way from Merge to recursion is not easy, it takes some steps, and not all are easy. Then, Merge could be much more ancient, and it is possibly shared with other species, because the binary tree data structure is simple and useful.

§5 Evolution

§5.1 Requirements

¶1 · It is true that we can express any recursive function in our natural language, and that by reading the expression we can understand how the function works, to the point of producing the right results. This is just a different way to state that our natural language is recursive. On the other hand, problem solving achieves its maximum power when Turing completeness is achieved, and Turing completeness is mathematically equivalent to recursion. For our hypothesis, this happens because problem solving and syntax are both computing, and therefore computing was shaped by taking evolutionary requirements from both, and then while Turing completeness is not a requirement of syntax, our

natural language is nevertheless recursive because Turing completeness is a requirement of full problem solving. In other words, some traits of language were selected, totally or partly, because they enhanced our problem solving abilities increasing our survival fitness.

¶2 · To better assess our hypothesis, in this Section §5 we will analyze those requirements. Our guide to problem solving in this investigation will be the programming language LISP, as defined originally by McCarthy (1960). We have chosen LISP for three reasons:

- LISP has stood the test of time. LISP is still active in the field of Artificial Intelligence, where it makes easy the resolution of many problems.
- LISP syntax is minimalist, using only binary trees. The LISP syntax parser is the simplest two-way parser.
- LISP is based on lambda-calculus, but it fuses the functional semantics of the three classical versions of recursion.

In summary, LISP is a language well suited for solving problems that uses the simplest syntax and implements all versions of recursion.

¶3 · Although the election of LISP is not arbitrary, it is neither completely neutral, and some findings could be an artifact of choosing LISP. To counter this possibility as much as possible, we will argue the need of every feature that full problem solving requires, following Casares (P), and we will explain how each feature is implemented in LISP. In summary, we will use LISP to guide our investigation on the relation between natural language and problem solving. You can judge our election at the end of this Section §5.

§5.2 Variables

¶1 · To express a problem we have to refer to its freedom and to its condition. To name the freedom we have to use a word that does not refer to anything, that is, it has to be a word free of meaning. For example, if the problem is that we do not know what to do, then its more direct expression in English is ‘what to do?’. In this sentence, the word ‘what’ does not refer to anything specifically, but it is a word free of meaning and purely syntactical, that is, a function word. In particular, if the answer to this question were ‘hide!’, then the meaning of the word ‘what’ in the question would be ‘hiding’, but if the answer were ‘run!’, then it would mean ‘running’.

¶2 · Frequently we answer a question with a single word and no additional syntax, which is an indication that in these cases a bi-sentential syntactic analysis is needed because the pair question-answer is a grammatical whole, see Krifka (2011). For example, the meaning of the pair:

(Q) — Who is reading ‘War and Peace’?

(A) — Bill.

reduces to

(M) — Bill is reading ‘War and Peace’.

where ‘Bill’ is the meaning of ‘Who’.

¶3 · In mathematics, the typical way to express the freedom of a problem is to use the unknown x , which works the same way as the interrogative pronoun ‘what’. For example, if we want to know which number is the same when it is doubled as when it is squared, we would write:

$$x? [2x = x^2].$$

The condition is the equality $[2x = x^2]$, in this case. Equality is a valid condition because it can be satisfied, or not.

¶4 · The x is just a way to refer to something unknown, so we could use any other expedient just by indicating it with the question mark (?). This means that

$$y? [2y = y^2]$$

is exactly the same problem. We call this equivalence α -conversion, following Curry & Feys (1958), §3D.

¶5 · Natural languages do not provide inexhaustible collections of unknowns or variables, as lambda-calculus and LISP and mathematics do, but only a few wh-terms that are not completely free of meaning. As observed by Hamblin (1973), ‘who’ is not completely free of meaning because it refers to a person or, in other words, it refers to an indefinite individual belonging to the set of persons. Because of this, there is not α -conversion in natural languages, and it is very difficult to express and to understand problems involving several unknowns of the same wh-type in natural languages. A consequence is that a mathematical notation to complement the natural language is needed. For us, the difficulty is an indication that, concerning unknowns or variables, language was first and problem solving last.

§5.3 Sentence

¶1 · It is important to note that the unknown has to be part of the condition, in order to determine if a value is a solution to the problem, or not. In the condition, the unknown x is a free variable, and therefore the condition is an open expression, that is, a function.

¶2 · In summary, in order to refer to the freedom of a problem we have to use free variables, which are words free of meaning that do not refer to anything. These free words are useless by themselves, we can even substitute one for another using an α -conversion, so they have to be combined with other words to compose the condition of the problem. We will call this structure of words a *sentence*. All things related to the sentence, as, for example, the rules for word combination, are what is usually called *syntax*. So, as separate words are not enough to express problems, then some structure of words is needed, and therefore syntax is needed to represent problems.

¶3 · LISP uses binary trees for its syntax. In LISP, any expression, technically called S-expression, is either a word or a sentence, where a sentence is an ordered pair. In LISP the function to generate a sentence is **cons**, already seen in §4.8, that takes two expressions and returns an ordered pair. Because the arguments of **cons** can also be ordered pairs, the resulting structure is a binary tree. Then, the result of (**cons** X Y) is a tree where the expression X is the left branch and the expression Y is the right branch. LISP also provides the corresponding deconstructing functions: **car**, that takes a tree as argument and returns the left branch of the tree, and **cdr**, that returns the right branch. The analysis of sentences in LISP requires also a predicate, **atom?**, that returns **t**, for TRUE, when it is applied to a word, and **nil**, for FALSE, when it is applied to a sentence. This is the complete apparatus that LISP needs to generate any sentence from its constituent words, and to reach the words that are any sentence. Then, for LISP syntax, generation is simpler than proper parsing, but in any case all these operations are decidable, so the LISP syntax parser is decidable.

¶4 · The LISP syntax parser is arguably the simplest two-way, word to sentence and sentence to word, parser. The two simplest composing functions are **cons** and Merge, because they compose only two objects, being the only difference that the result of **cons** is ordered, while the result of Merge is unordered. If we use Merge instead of **cons**, then

some marker or label, as a preposition or simply stress, is needed to signal the elements of the sentence. Markers give more flexibility, because then it is a pragmatic decision how to order the information conveyed by a sentence. On the other hand, when order is prescribed by syntax, parsing is simpler and needs less memory, because then the parser always knows what to expect. In summary: the Merge parser is nearly as efficient as the optimal `cons` parser, and much more flexible.

¶5 · This shows that there is a trade-off between the parsing efficiency of `cons`, and the pragmatic convenience of Merge. Some natural languages use order, some markers, and most both, suggesting that evolution has selected the more flexible Merge as an answer to the trade-off between syntactic efficiency and pragmatic convenience. In that case, Merge was selected.

¶6 · We saw that syntax is needed to express problems, and that LISP uses a tree sentence structure, which is simpler than, but similar to, the natural languages sentence structure generated by Merge. What we left pending is the check against problem solving, that is, whether problem solving requires a tree sentence structure, or not.

§5.4 Problem

¶1 · We have seen why we need sentences, that is, syntax, to express problems, and why separate words are not enough. We have also seen two types of word: semantic or content words with meaning, and syntactic or function words without meaning. Well, although it seems impossible, there is a third type of word: defined words.

¶2 · A defined word is just an abbreviation, so we could go on without them, but they are handy. That way we substitute a word for a whole expression. It is the same expedient that we use when we introduce a technical term to encapsulate a phrase in order to avoid repeating the phrase. We can, for example, define a word to refer to a condition:

$$p_x := [2x = x^2].$$

¶3 · To state a problem is to state the condition that its solutions have to satisfy. This is the same as defining ‘something’ by listing the properties, uses, and, in general, the requirements that anything has to fulfill to be called ‘something’. Perhaps because defining is like stating a problem, it has not been studied in linguistics, though it happens necessarily in a language. But the first LISP of McCarthy (1960) already used definitions, then introduced by the operation `label`, “[i]n order to be able to write expressions for recursive functions,” page 186. And that is because the genuine recursion needs names to refer to functions.

¶4 · LISP itself would not lose any computing power without `label`, nor without its replacement `define`, see McCarthy et al. (1962), §2.3. This is because LISP also implements the functional semantics of lambda-calculus, and in lambda-calculus, which only uses anonymous functions, a function can call itself using the paradoxical combinator `Y`, see Curry & Feys (1958), §5G. As Friedman & Felleisen (1987) show, LISP does not need `define`, but without names the code goes easily beyond comprehension.

¶5 · So, in the end, definitions are more than handy, and we cannot go easily without them. Our processing limitation to about seven chunks of information, see Miller (1956), can be the reason why we real persons need defined words.

§5.5 Routine

¶1 · A *resolution* is a way of solving a problem, that is, a process that takes a problem and tries to return the solutions to the problem. A successful resolution of the problem $x? [2x = x^2]$ can proceed this way:

$$\begin{aligned}
 & [2x = x^2] \\
 & [2x - 2x = x^2 - 2x] \\
 & [0 = xx - 2x] \\
 & [0 = (x - 2)x] \\
 & [x - 2 = 0] \vee [x = 0] \\
 & [x - 2 + 2 = 0 + 2] \vee [x = 0] \\
 & [x = 2] \vee [x = 0] \\
 & \{2\} \cup \{0\} = \{2, 0\} .
 \end{aligned}$$

In this case the resolution was achieved by analogy transforming the problem, or rather its condition, until we found two subproblems with a known solution, $x? [x = 2]$ and $x? [x = 0]$, which we could then resolve by routine. So the problem has two solutions, two and zero.

¶2 · To represent each transformation the simplest expedient is the ordered pair, (s, t) , where s and t represent two expressions: s before being transformed, and t once it has been transformed. Note that we can use a single ordered pair to express each transformation, or the whole sequence. For example, the complete sequence can be summarized in just one ordered pair, which, in this case, describes how to resolve the problem by routine, because s is the problem and t is the set of its solutions:

$$(x? [2x = x^2], \{0, 2\}) .$$

¶3 · In summary: To resolve a problem by routine you don't need to reason, you only need to remember what are the solutions to the problem, and that information can be expressed using an ordered pair. Therefore, we can use LISP to express any resolution by routine. A resolution by routine can also be expressed using a two-element set, though then we should mark at least one element in order to distinguish which one is the problem and which one the set of solutions, so Merge would also work.

§5.6 Trial

¶1 · We can also resolve by trial and error. In the trial we have to test if a value satisfies the condition, or not, and the condition is an open expression. To note mathematically open expressions, also known as functions, we will use lambda abstractions $(\lambda x p_x)$, already seen in §4.3, where x is the free variable and p_x the open expression. In the case of our problem:

$$(\lambda x [2x = x^2]).$$

¶2 · Now, to test a particular value a , we have to bind that value a to the free variable inside the condition. We also say that we apply value a to the function $(\lambda x p_x)$. In any case we write function applications this way: $((\lambda x p_x) a)$. We can abbreviate the expression naming the function, for example $f := (\lambda x p_x)$, to get $(f a)$, which is just the

lambda way to write the typical $f(a)$. In our case, ‘to test if number 2 is equal when it is doubled to when it is squared’ is written $((\lambda x [2x = x^2]) 2)$. And to calculate if a value satisfies the condition, we replace the free variable with the binding value; this process is called β -reduction, also seen in §4.3. In our case we replace x with 2 ($x := 2$), this way:

$$((\lambda x [2x = x^2]) 2) \rightarrow [2.2 = 2^2] \rightarrow [4 = 4] \rightarrow \text{TRUE}.$$

¶3 · Lambda abstraction, function application, β -reduction, and an inexhaustible source of variables, are all the ingredients needed to concoct a lambda-calculus interpreter. LISP implements all of them: any non-reserved word can name a variable, the reserved word `lambda` as the left branch of a sentence is a lambda abstraction, and a lambda abstraction as the left branch of a sentence is a function application on which a β -reduction is automatically executed. This way LISP implements the whole lambda functional semantics.

¶4 · LISP implements the simplest β -reduction: it uses a call-by-value reduction strategy, and a positional assignment, that is, the assignment of actual to formal parameters is based on position, the first actual parameter to the first formal parameter, the second to the second, and so on. As with `cons` and `Merge`, see §5.3, natural language could be using a more flexible kind of β -reduction, as Unification proposed by Jackendoff (2011). Unification, see Shieber (1986), works on feature structures, which are labeled tree data structures that play the rôle of lambda abstractions, and then any device implementing Unification needs also operations to parse tree data structures, as `Merge`. Therefore, Unification and `Merge` can be complementary but not mutually exclusive.

¶5 · A condition can only return two values, which we will call `TRUE` and `FALSE`. In case we want to make a difference depending on the condition, and what else could we want?, then we have to follow one path when the condition is satisfied (`TRUE`), and a distinct path when the condition is not satisfied (`FALSE`). LISP provides a conditional command `cond` that can do it, although it is more general:

$$(\text{cond } (\langle \text{condition} \rangle \langle \text{TRUE case} \rangle) (\text{t } \langle \text{FALSE case} \rangle)).$$

¶6 · We can write down completely any trial just by using these two expedients: binding values to free variables in open expressions, and a conditional command, as `cond`. Suppose, for example, that we guess that the solution to our problem is one of the first four numbers, in other words, that the solution is in set $\{1, 2, 3, 4\}$, and that we want to try them in increasing order. Then, defining

$$f := (\lambda x [2x = x^2])$$

and mixing notations freely, the trial would be:

```
(cond ((f 1) 1)
      ((f 2) 2)
      ((f 3) 3)
      ((f 4) 4)
      (t nil)) .
```

In our natural language this code reads: if the condition f is satisfied by 1, then 1 is the solution; or else if the condition f is satisfied by 2, then 2 is the solution; and so on. A conditional is the natural way to express any trial and error resolution.

¶7 · In fact, to code any computable function in LISP, that is, to code any Turing machine in LISP, we need the commands `cond` and `set!`, the predicate `eq?`, and the procedures to control the tape. From the inexhaustible source of words, a finite set is needed to name the internal states, and another finite set is needed to name the symbols on the tape. Command `set!` assigns a value to a variable, and it is needed to set the internal state of the Turing machine. Predicate `eq?` compares two words, and it is needed to determine which clause of the `cond`-itional applies, that is, which is the right clause for the current internal state and read symbol. The procedures to control the tape of a Turing machine are just: `read`, which returns the word that is in the current position of the tape; `write`, which writes its argument in the current position of the tape; and `move`, which accepts as argument one of *left*, *right*, or *halt*, and acts accordingly. Finally, the conditional has to be executed in a loop. This way LISP also implements the whole functional semantics of computing. You can find an example in Casares (T) §7.1.

§5.7 Analogy

¶1 · If we know the solutions of a problem, then we can resolve it by routine, which can be written as an ordered pair, $(x? [2x = x^2], \{0, 2\})$, as we have already done in §5.5. If we do not know the solutions, but we suspect of some possible solutions, then we can resolve it by trial and error, which can be written as a conditional,

$$(\text{cond } ((f\ 1)\ 1)\ ((f\ 2)\ 2)\ ((f\ 3)\ 3)\ ((f\ 4)\ 4)\ (t\ \text{nil})),$$

as we have already done in §5.6. There is a third way of solving problems: by analogy.

¶2 · By analogy we transform a problem into another problem, or problems. Most times the outcome will be more than one problem, because ‘divide and conquer’ is usually a good strategy for complex problems. So, in general, the resolution of a problem will be a tree, being the original problem its trunk. If we use analogy to resolve it and we get, for example, four easier subproblems, then the tree has four main branches. But, again, from each branch we can use analogy, and then we get some sub-branches, or we can use routine or trial. We resolve by routine when we know the solutions, so the subproblem is solved; these are the leaves of the resolution tree. Trial ...

¶3 · You got lost? Don’t worry, even myself get lost in this tree of trees, and to what follows the only important thing to keep in mind is one easy and true conclusion: expressions that represent resolutions of problems have a tree structure, because they describe the resolution tree. And now we can answer the question that we left pending at the end of Subsection §5.3: problem solving does indeed require a tree sentence structure.

¶4 · Both LISP and natural language use tree sentence structures, conditionals and pairs, so both can be used to represent resolutions of problems. The question now is to ascertain the additional requirements that calculating resolutions of problems impose. In particular, we will investigate the requirements that a full problem solver has to fulfill in order to solve as many problems as possible.

§5.8 Resolution

¶1 · A condition, also known as predicate, is a function with two possible outcomes, the loved TRUE and the hated FALSE. This means that each time we apply it we get a TRUE or a FALSE. For example, defining again $f := (\lambda x [2x = x^2])$, we get:

$$\begin{aligned}(f\ 0) &\rightarrow \text{TRUE} \\(f\ 1) &\rightarrow \text{FALSE} \\(f\ 2) &\rightarrow \text{TRUE} \\(f\ 3) &\rightarrow \text{FALSE} \\(f\ 4) &\rightarrow \text{FALSE} \\&\vdots\end{aligned}$$

¶2 · What solves definitively a problem is the inverse function of the condition of the problem. Because the inverse function just reverses the condition, from a $0 \rightarrow \text{TRUE}$ to a $\text{TRUE} \rightarrow 0$, so when we apply TRUE to the inverse condition we get the solutions, and when we apply FALSE we get the set of the no-solutions. Thus we get:

$$\begin{aligned}(\check{f}\ \text{TRUE}) &\rightarrow \{0, 2\} \\(\check{f}\ \text{FALSE}) &\rightarrow \{1, 3, 4, \dots\}.\end{aligned}$$

¶3 · We can use the condition in both directions: the natural direction, when we apply a value to test if it satisfies the condition, and the opposite direction, when we want to know what values satisfy the condition. To express a problem is enough to write the condition and to indicate which are its free variables, and to solve a problem is enough to apply the condition in the opposite direction.

¶4 · It is too easy to say that the inverse function of its condition solves the problem; I have just done it. Unfortunately, it is nearly always very difficult to find the inverse of a condition, and sometimes the condition is inexpressible or unknown.

¶5 · We should distinguish solving, which is finding solutions, as the inverse condition does, from resolving, which is finding resolutions, that is, finding ways of finding solutions. Then resolving is calculating the inverse function of a condition, given the condition, and therefore a resolution is a function that, when it is successful, takes a problem, or its condition, and returns the problem solutions.

$$\text{Problem} \xrightarrow{\text{Resolution}} \text{Solution}$$

¶6 · In any case, finding a resolution to a given problem is a problem, its *metaproblem*. Finding a resolution to a problem is a problem because it has its two mandatory ingredients. There is freedom, because there are several ways to solve a problem, which are the routine and the different trials and analogies, and there is a condition, because not every function is a valid resolution for a problem, but only those that return the solutions to the problem. And then, being a problem, we need a resolution to find a resolution to the problem. And, of course, to find a resolution to find a resolution to the problem we need, can you guess it?, another resolution. What do you think? Better than ‘convoluted’ say ‘recursive’.

§5.9 Recursivity

¶1 · The solution to a problem can also be a problem or a resolution. For example, when a teacher is looking for a question to ask in an exam, her solution is a problem. And when an engineer is designing an algorithm to calculate some type of electrical circuits, her solution is a resolution. In this last case, the next problem could be to enhance the algorithm found, and then the resolution of this optimization problem would take a resolution to return a better resolution. Solutions can also be enhanced.

¶2 · The condition of a problem takes a possible solution as argument and returns TRUE when the argument is a solution, and FALSE when it is not a solution. Therefore, if the solutions of a problem are problems, then the condition of the problem has to take problems, or conditions, as arguments, and if the solutions of a problem are resolutions, then the condition of the problem has to take resolutions as arguments. This last case is the case of metaproblems, for example.

¶3 · A full problem solver has to solve as many problems as possible, and then it should take any problem, or its condition, and it should return the problem solutions. Any condition is a problem, if we are interested in the values that satisfy the condition, and then the full problem solver has to take any condition as input, including conditions on problems and on resolutions. Not every function is a condition, but only generalizing to functions we can cope with the resolution to resolution transformation that the engineer above needed. And the solutions returned by the full problem solver can be also problems and resolutions.

¶4 · The conclusion is then that a full problem solver resolution should input any expression and can output any expression. So a full resolution has to be able to transform any expression into any expression.

$$\left. \begin{array}{l} \text{Problem} \\ \text{Resolution} \\ \text{Solution} \end{array} \right\} \xrightarrow{\text{Resolution}} \left\{ \begin{array}{l} \text{Problem} \\ \text{Resolution} \\ \text{Solution} \end{array} \right.$$

The resolution can then be the transformation, but also what is to be transformed, and what has been transformed into. We call this property of transformations that can act on or result in transformations without any restriction *recursivity*.

¶5 · We can now answer the question asked at the end of Subsection §5.7. Full problem solving requires recursion, because full problem solving requires functions taking functions and returning functions as generally as possible, without restrictions, and then a whole functional semantics is needed. The function is the mathematical concept for transformation, for change. Therefore, provided with functional semantics, the recursive problem solver masters change.

¶6 · When the current state of things makes survival difficult or inconvenient or simply improvable, the problem is to find a way to go to a better state. In these circumstances, a living being that does not implement recursion can try some possible ways of solving from a set engineered by evolution, which can include imitating others, or trying with sticks or some other tools. But a recursive individual, with functional semantics, can calculate any change and its results before executing one. Being a tool any means to make change easier, the calculation can introduce a subproblem that asks for a tool, and then the Turing complete individual can devise the proper tool, and even devise a tool to make the tool, and so on. Mastering change, we recursive solvers can foresee the effects of our actions, and, at least to some extent, we can match the future to our own will.

§5.10 Quotation

¶1 · Recursivity needs a quoting expedient to indicate if an expression is referring to a transformation, or it is just an expression to be transformed.

¶2 · Joke:

- Tell me your name.
- Your name.

Who is answering has understood ‘tell me “your name” ’, which could be the case, although unlikely. In any case, quotation explains the confusion.

¶3 · Quotation is very powerful. Suppose that I write:

En cuanto llegué, el inglés me dijo: “I was waiting you to tell you something. ...
⋮
[a three hundred pages story in English]
⋮
... And even if you don’t believe me, it is true”. Y se fue, sin dejarme decirle ni una sola palabra.

Technically, I would have written a book in Spanish!

¶4 · Technically, using the quoting mechanism, all recursive languages are one. This is because, by using quotation, recursive languages can incorporate any foreign or artificial construction. For example, in a LISP manual written in English, as McCarthy et al. (1962), there would be some parts following the English syntax, and some other quoted parts following the LISP syntax that is being explained in the manual. LISP is a recursive language because a LISP interpreter can calculate any recursive function, and then English is a recursive language because we can learn LISP from a manual in English.

¶5 · That all recursive languages are one is in line with Chomsky’s universal grammar, and with his idea that seen globally “there is a single human language, with differences only at the margins”, quoted from Chomsky (2000), page 7.

¶6 · Mathematically, we have already dealt with this situation in §4.4: Any recursive language can produce any recursive language, and even any language, because the device that implements the recursive language is a universal computer, that is, a universal grammar.

§5.11 Lisp

¶1 · Lastly, we should review a complete LISP interpreter. Instead of the first LISP proposal by McCarthy (1960), we will analyze the complete implementation by Abelson & Sussman (1985). In fact, they implement two LISP interpreters: one coded in LISP in §4.1, and another one coded on a register machine in §5.2. Both interpreters code a read-eval-print loop, §4.1.4 and §5.2.4, where the core is the evaluator, §4.1.1 and §5.2.1. The evaluator is a dispatcher.

¶2 · The first task of the evaluator is to treat the self-evaluating words. Self-evaluating words have a meaning by themselves, and therefore they do not need any syntactic treatment. So this is the interface to semantics. In the case of this LISP interpreter by Abelson & Sussman, only numbers are self-evaluating, but on commercial ones there can be several kinds of semantic words, as booleans or strings or vectors. In the case of natural language, there are thousands of self-evaluating words, also known as content or autosemantic words, mainly nouns, adjectives, verbs, and adverbs.

¶3 · The second task is to deal with quoted text, which is also excepted of syntactic treatment. This LISP interpreter implements only `quote`, which is the literal quotation operation, but commercial ones have additional quotation operators, as `quasiquote` or `macro`. Using a general version of quotation, like the one used in written natural language, each kind of quotation would be dispatched to its proper parser, and then a book in English should quote differently its parts in Spanish and its parts in LISP, for example.

¶4 · After the exceptions, the dictionaries, where a dictionary is a list of variable-value pairs, and a variable can be any non-reserved word. This LISP has three kinds of variables:

- from genuine recursion, those introduced by `define`, which cannot be modified, and then they are like technical definitions in natural language;
- from computing, those introduced by `set!`, which can be modified and have some similarities to the pronouns of natural language; and
- from lambda-calculus, those introduced by lambda abstractions, which are the names of the function arguments, and then, comparing a function to a verb, these variables are the subject and objects of the sentence, or clause.

This LISP is strict and every variable has a well-defined scope, so its interpreters have to construct a hierarchy of dictionaries, and their look-up procedures have to start from the nearest one at the bottom, and they have to go up to the next one when searching fails.

¶5 · Then the evaluator deals with `lambda` and `cond`. The first implements the lambda abstraction of lambda-calculus, and the second the conditional of computing. These two plus the `define` of genuine recursion are each one the core of a functional semantics. Natural language can be using a more flexible version of lambda abstraction, as the feature structure of Unification, for example, that could work also as a dictionary.

¶6 · The only remaining task of the evaluator is to apply functions. This LISP has two kinds of functions: primitive functions, which are implemented by the interpreter, and compound functions, which are calculated by β -reduction. This LISP interpreter implements the following primitive functions:

- `cons`, `car`, `cdr`, and `atom?`, which are needed to implement its syntax;
- `eq?`, which is needed to implement the functional semantics of computing;
- some procedures needed to access media, as `read`, `write`, and `move`, for tapes;
- some procedures needed to deal with self-evaluating, or semantic, words; and
- some other functions implemented just for efficiency.

Because numbers are self-evaluating for this interpreter, then it has to implement some primitive functions for numbers, as the successor function, `1+`, or the predicate `number?`, for example. In the case of natural language, with thousands of semantic words, we should expect to find also thousands of primitive functions, and though most of them would not be language specific, as for example the one for face recognition, some would be, as Merge for syntax and also those needed to access media providing an interface to the sound system for speech generation and recognition. And, instead of β -reduction, natural language can be using a more flexible version, as Unification, for example.

¶7 · Comparing `cons` and its sibling syntax operations with the complete LISP interpreter, the conclusion should be again the same: the very reason of any recursive device is not syntax parsing, but full problem solving. Then, the interface to pragmatics is the very reason of our natural recursive machinery.

¶8 · *The devil is in the details.* To study the evolution of the eye, it is important to know the laws of optics. A video camera does not work as an eye because the materials and methods of an engineer are different from those of evolution, but the limitations of optics are the same for both. Here, the LISP interpreter plays the rôle of the video camera, providing a complete and detailed list with all the requirements that a recursive device has to implement, and a working model with the reasons why each and every element of the whole does what it does. Unfortunately, we lack other living implementations to compare human recursion with, but at least we have LISP. With this warning note about the use of models, we conclude our tour around LISP, a masterpiece of engineering.

§5.12 Summary

¶1 · The requirements found in this Section §5 are evolutionary, so they are not pre-conditions of design, but post-conditions. Each evolutionary requirement appears by chance, grows in the short term by a mixture of chance and fitness, is kept in the mid term because of its fitness, and is entrenched in the long term as part of a bigger structure, which is the recursive engine, in our case.

¶2 · From a mathematical point of view, the list of requirements found in this Section §5 is redundant. For example, lambda-calculus does not need definitions, because it does everything with anonymous functions. But, for a function to refer to itself, we need names, and therefore to define genuine recursive functions we need definitions! It is nearly the same with conditionals. Lambda-calculus does not need them because they can be defined as any other function, but the conditional is an instruction that a universal Turing machine needs. That the list of requirements is redundant only means that evolution does not value succinctness, but efficiency. And though the different versions of recursion are mathematically equivalent, their implementations are distinct from an engineering point of view. Then, by implementing more than one, the problem solver has the opportunity to use each time the version that is best suited to the problem. And redundancy also pays in robustness.

¶3 · From a very abstract point of view, all the requirements found in this Section §5 can be connected to the concept of function, which is the key concept of recursion. A function is the general form of conditional, because the result of the function is conditioned on the values it takes; a computable function calculated by a universal Turing machine is expressed just as a list of conditionals. But a function can also be expressed as a structure with holes that, depending on how the holes are filled, results differently. The holes are pronouns, and the rest of the structure are other words that compose the sentence. The pair is arguably the simplest structure, and then a tree built from pairs is surely the simplest open structure, and therefore the binary tree is the first candidate for the sentence. See that the Merge operation of the minimalist program builds binary trees. And, finally, we need recursion to get the most general functions that are still effectively calculable, as first stated by Church (1935).

¶4 · The point of this Section §5 was to show some requirements from full problem solving that our natural languages exhibit. This does not prove that they were originated by problem solving, and below in Subsection §6.3 we will argue that some were possibly not, but, in any case, they enhanced our problem solving abilities, and then these ‘problematic’ features of language enhanced our survival fitness because of problem solving.

§6 Discussion

§6.1 Problems and Evolution

¶1 · Life can be assimilated to the survival problem. From this point of view, which we will call the problematic view of life, evolution is a problem solver. Evolution solved some subproblems of the survival problem by designing systems, as the cardiovascular system, and it solved their sub-subproblems by designing organs, as the heart.

¶2 · But evolution cannot solve the problems faced by individuals in their day-to-day living, because those problems depend on casual circumstances. For solving those problems faced by individuals, evolution designed the nervous system, the brain organ, and even the specialized nervous tissue and neuron cell. Then, broadly speaking, the function of the nervous system is to deal with information, and the function of the brain is to take information from the body, compute what to do in order to resolve according to the circumstances, and send the resulting command information back to the body. In other words, the brain is the solver of the problems of the individual.

¶3 · There is a delicate interaction between the brain and the rest of the body, which is calibrated by the proper distribution of responsibilities between the two problem solvers involved, evolution and the brain. For example, heart beat rate can be autonomous, as shown by a separated heart beating, but the brain can command to increase the beat rate when running, for example, to flee from a predator.

¶4 · We are Turing complete because our brain is Turing complete, but not all living individuals are Turing complete, so we might wonder what difference does this make. To start, we have to distinguish solving, or finding solutions, from resolving, or finding resolutions, where a resolution is a way of solving, that is, a process for solving problems, and then mathematically, as seen in §5.8, a resolution is a function that takes a problem and returns solutions, right and wrong solutions. So, being a function, a Turing complete individual can express and understand, that is, imagine, any resolution in her recursive language, while a more limited individual will apply its limited set of resolutions to any problem. The key point is that a single Turing complete individual can imagine any possible resolution, that is, any possible way of solving a problem, and then she can execute any of the imagined resolutions that returns right solutions, while an individual that is not Turing complete can only apply those resolutions that are implemented in the hardware of its body, mainly in the hardware of its brain.

¶5 · Species that are not Turing complete need a genetic change to modify their set of resolutions, while Turing complete individuals can apply new resolutions without any hardware change, but just by a software change. So the timing of creativity depends on evolution until Turing completeness is achieved, and it does not depend on evolution after that point for the species that achieve Turing completeness. We will call every point where an evolutionary path achieves Turing completeness an *evolutionary singularity*. In other words, an evolutionary singularity is any evolutionary moment when the brain surpasses evolution in problem solving.

¶6 · New behaviors related to the solution of problems, such as feeding, mating, and, in general, surviving, should proliferate after an evolutionary singularity. And noting that a tool is the physical realization of a resolution, then an explosion of new tools should start whenever an evolutionary singularity happens. So the archaeological record of human tools should point to our evolutionary singularity.

¶7 · In summary, creativity is slow until an evolutionary singularity and creativity ex-

plodes after every evolutionary singularity because, after achieving Turing completeness, performing new ways of solving the survival problems becomes cheap and available to single individuals while, before achieving it, any single new way of solving them required genetic changes on species over evolutionary time spans.

¶8 · Nevertheless, Turing completeness is defined by a single condition: pan-computability. A universal Turing machine, which is the prototype of Turing completeness, is every Turing machine that can compute whatever any Turing machine can compute, but no more. This means that to perform any specific computation you can, either build the Turing machine specific for that computation, or write the program for that specific computation on a universal Turing machine. Either way the outcome of the computation will be the same, and the only differences would be that the first can run faster, and that the second can be implemented faster, once you have a universal computer. The second is better for modeling, that is, for imagining, because writing models is much faster than building models, but, again, the results of the computations are the same.

¶9 · In other words, creativity is the only exclusive feature of Turing complete problem solvers. And this explains an elusive fact: every time a specific behavior is presented as uniquely human, it is later rejected when it is found in another species. The point is not that we behave in some specific way to solve a problem, but that we are free to imagine any way to solve our problems. Creativity is the mark of Turing complete solvers.

§6.2 Syntax and Problems

¶1 · We could equate problem solving with syntax, because syntax is needed just to express problems, but mainly because both syntax and problem solving are computing. And full problem solving requires universal computability, which is the maximum syntactic requirement. So full problem solving needs syntax, and needs all of it: full problem solving needs a complete syntax.

$$\left. \begin{array}{c} \text{Complete} \\ \text{Syntax} \end{array} \right\} = \left\{ \begin{array}{c} \text{Universal} \\ \text{Computing} \end{array} \right\} = \left\{ \begin{array}{c} \text{Full} \\ \text{Problem Solving} \end{array} \right\}$$

¶2 · By identifying syntax with problem solving we have solved the syntax purpose paradox presented in §2.2: when a brain becomes a complete syntax engine, then that brain becomes a full resolution machine. So syntax can be meaningless, but it is very useful and its purpose is to solve problems in any possible way. The key is that while a complete syntax is still mechanical, it is not really meaningless, because it has a functional semantics.

¶3 · That both syntax and problem solving are computing is a coincidence that can also be derived from the problematic view of life. In this view, whatever the brain does is to compute for solving the problems of the individual. Communication, or sharing information with other individuals, is just one problem out of those that the brain has to solve, and language is just a particular solution for communication. This explains the coincidence, but here we distinguish communication from thinking, defining very artificially thinking as what the brain does that is not communication, to separate the linguistic requirements from the much more general cognitive requirements. The reason is that syntax, which is in principle a solution to the communication problem, was instrumental in achieving full problem solving, which concerns cognition generally.

¶4 · Syntax and problem solving use the same computing machinery, but their particular requirements are different: better syntax is for better communication, while better problem solving is for better thinking. Then our computing machinery should have evolved by satisfying requirements from syntax (communication), or from problem solving (thinking), or from both. And then we should find problem solving requirements in syntax, and syntax requirements in problem solving, or, in fewer words, language and thinking co-evolved in humans.

¶5 · The description by [Vygotsky \(1934\)](#) of the co-development of language and thinking in children can be seen as a possible model for the co-evolution of language and thinking under our hypothesis. In fact, at this very abstract level, the development of syntax and problem solving in children can use every argument used here to base our hypothesis, and then, the very same arguments can be used to base a parallel hypothesis: ‘syntax and problem solving co-develop in children towards recursion’. This parallel hypothesis is easier to examine, and if it is found true, then the original hypothesis, which is based on the same arguments though applied to a different process, would gain some support.

¶6 · Syntax recursivity is credited with making possible the infinite use of finite means, which is obviously true. But our wanderings in the problem resolution realm have shown us that there are other features provided by recursion that are, at least, as important as infinity; infinity that, in any case, we cannot reach. The main three ones are: sentences, or hierarchical tree structures; functions, or open expressions with free variables; and conditionals, or expressing possibilities: *I cannot imagine how I would see the world if there were no conditionals, can you?* Conditionals, for example, are necessary to think plans in our head that can solve our problems, and then problem solving gives conditionals a purpose and an evolutionary reason that they cannot get from language.

§6.3 Syntax and Evolution

¶1 · We have seen that syntax and problem solving co-evolved towards Turing completeness, because both are computing, and Turing completeness is the maximum computing power. But, how was that co-evolution?, that is, which one, syntax or problem solving, was driving each step of the co-evolution?

¶2 · Before answering these questions, we must set the starting point. As syntax requires words to work on, we will start when there were already words. [Pavlov \(1927\)](#) showed that the assignment of arbitrary sounds to meanings is a capability that dogs already have, and then words should have preceded syntax. But dog words are indices for [Pierce \(1867\)](#), and neither icons nor indices can be produced in high quantities, so we should expect to have symbols before syntax developed. For these and other reasons, our starting point will be a protolanguage, as presented by [Bickerton \(1990\)](#), that has already developed phonology and a lexicon with its semantics.

¶3 · We argue that the very first steps, leading to the sentence, were driven by syntax. The main reason is that syntactic languages are exponentially more expressive than asyntactic languages, see [Casares \(H\)](#). By giving meaning to sentences, a syntactic language can assign different meanings to word sequences as ‘dog bit child’ and ‘child bit dog’ that would mean exactly the same in asyntactic languages. The implication is that ambiguity is reduced efficiently just by imposing some structure to a set of words, and this is the case even without free variables, which was possibly the first reason why problem solving needed sentences. For example, the first step of syntax could be very simple: just put the agent before every other word. This simple expedient by itself would prevent efficiently

some ambiguities, explaining who did and who was done something. Marking who is the agent would also achieve it. And that natural languages do not provide inexhaustible sets of variables points in this direction, too.

¶4 · We argue that the last steps, fulfilling universal computability, were driven by problem solving. The reason, now, is that Turing completeness is useful for problem solving to detach itself from hardware causing an explosion of creativity, but it is detrimental to natural language syntax, as stated by the anomaly of syntax.

¶5 · I don't know which one, syntax or problem solving, was driving each of the other steps that provided each of the capabilities that are needed to implement a Turing complete brain. But what I take for sure is that we humans evolved to Turing completeness, not to achieve a full resolution machinery, which was the final prize, but because each recursive feature, such as the sentence, the open expression, and the conditional, made thinking (problem solving) or communication (syntax), or both, more efficient, resulting in more fitted individuals. A more detailed study of the requirements found in Section §5, in addition to those found in a grammar text book, will be needed to disentangle the co-evolution of syntax and problem solving.

¶6 · Our general view of the evolution from an animal communication system to our recursive language has then a middle point, protolanguage.

Animal Communication → Protolanguage → Recursive Language

To reach the protolanguage, evolution had to provide facilities that were both cognitive, to store and retrieve many words, and anatomical, to utter and decode many words. Tree data structures are simple and useful, so they were probably already used for vision, see [Marr \(1982\)](#), and then the next step was to repurpose, or just to reinvent, the tree data structure computing capability to use it with words, yielding syntax. After syntax, some more evolutionary steps were required to implement a functional semantics, reaching Turing completeness. Therefore, the evolution from protolanguage to recursion was only cognitive, or more precisely, it was completely computational, and then it should have left no anatomical clues, and it could have happened entirely within the time frame of the anatomically modern *Homo sapiens*.

¶7 · In any case, it would be difficult to deny that syntax was, at least, instrumental in achieving Turing completeness, and therefore that syntax was influential in reaching our evolutionary singularity.

§6.4 Beyond the Singularity

¶1 · These results should help us to reevaluate syntax. In language evolution there are two main positions regarding syntax; see [Kenneally \(2007\)](#), Chapter 15. The gradualist side defends a more gradual evolution, where syntax is just a little step forward that prevents some ambiguities and makes a more fluid communication; see [Bickerton \(2009\)](#), page 234. For the other side, led by Chomsky, syntax is a hiatus that separates our own species from the rest; see [Hauser, Chomsky, and Fitch \(2002\)](#). What position should we take?

¶2 · The co-evolution of syntax and problem solving explains that, once our ancestors reached Turing completeness, they acquired complete syntax, also known as recursion, and thus they mastered change and they became full problem solvers. Then they could see a different the world. How much different? Very much.

¶3 · Seeing the world as a problem to solve implies that we look for the causes that are hidden behind the apparent phenomena. But it is more than that. Being able to calculate resolutions inside the brain is thinking about plans, goals, purposes, intentions, doubts, possibilities. This is more than foreseeing the future, it is building the future to our own will. Talking about building, what about tools? A tool is the physical realization of a resolution, and then with complete syntax we can design tools, and even design tools to design tools, and so on recursively.

¶4 · We are creative because we master change. With our complete syntax, we can imagine any possible transformation and its effects, so we can create our own imagined worlds, and we can drive the real world in the direction we want, creating artificial domains in it. Mainly, we adapt the world to us, instead of the converse. Rather than just surviving, we are always looking for new ways to improve our lives.

¶5 · We need syntax to express the freedom of problems, and with complete syntax we can calculate any way of solving problems. Note that calculating different ways of solving our problems is calculating how to use our freedom in order to achieve our ends. So, yes, we are free because of syntax.

¶6 · Summarizing, the identification of syntax with problem solving explains why syntax, being so little thing, has made us so different. Then we agree with Chomsky on that syntax was instrumental in creating the hiatus that separates our species from the rest. But recursion is more than just one operation, so we have probably acquired the complete syntax in more than one evolutionary step, and on this we agree with the gradualist party.

§7 Conclusion

¶1 · Only our species is Turing complete. Therefore, we must explain the evolution of Turing completeness to understand our uniqueness.

¶2 · Turing complete individuals can transform strings of symbols, irrespective of the symbols meanings, but according to any possible finite set of well-defined rules. It seems a nice capability but, being meaningless, not very useful. It could be, but syntax is also about meaningless transformations of strings of symbols according to rules. Turing completeness is then a pan-syntactical capability, because the syntax of a natural language does not need to follow any possible set of rules, but only one specific set of rules.

¶3 · Syntax is very peculiar indeed, because syntax is a peculiarity of human language, which is our most peculiar faculty. For this reason, we can argue that syntax is what defines our species, and yet it seems improbable to explain how some mechanical transformations of strings of symbols have made us like we are. In addition, syntax is not our only peculiarity: is it a coincidence that we are the only species with a syntactic language and also the most creative?

¶4 · But then, after realizing that Turing completeness is closely related to syntax, which is a human peculiarity, we have to halt, because we cannot progress anymore without new inputs. In order to overcome that impasse, this paper introduces a new piece of the jigsaw puzzle that links areas that were previously unconnected: the new piece is problem solving. Problem solving is a piece of problem that fits with syntax on one side, and with Turing completeness on the other side.

¶5 · We are now ready to understand Turing completeness from the perspective of problem solving. After finding that creativity is the peculiarity of Turing complete problem solvers, we see how our Turing complete problem solving machinery explains both our

creativity and our pan-syntactical capability. But before that, we have also seen that Turing complete problem solving needs sentences, functions, and conditionals; all of them employed by syntax.

¶6 · The conclusion is that syntax and problem solving should have co-evolved in humans towards Turing completeness.

Acknowledgements

¶1 · Thanks to Kleantes Grohmann for his encouragement, and to Mark Hauser, Ray Jackendoff, Diego Krivochen, and the anonymous reviewers of Biolinguistics, for their comments on previous versions of this paper.

References

- Abelson & Sussman (1985): Harold Abelson, and Gerald Sussman, with Julie Sussman, *Structure and Interpretation of Computer Programs*; The MIT Press, Cambridge MA, 1985, ISBN: 978-0-262-01077-1.
- Bickerton (1990): Derek Bickerton, *Language & Species*; The University of Chicago Press, Chicago, 1990, ISBN: 978-0-226-04611-2.
- Bickerton (2009): Derek Bickerton, *Adam's Tongue: How Humans Made Language, How Language Made Humans*; Hill and Wang, New York, 2009, ISBN: 978-0-8090-1647-1.
- Bickerton (2014): Derek Bickerton, *More than Nature Needs: Language, Mind, and Evolution*; Harvard University Press, Cambridge MA, 2014, ISBN: 978-0-674-72490-7.
- Casares (H): Ramón Casares, "A Complete Hierarchy of Languages";
DOI: [10.6084/m9.figshare.6126917](https://doi.org/10.6084/m9.figshare.6126917).
- Casares (P): Ramón Casares, "Problem Theory";
DOI: [10.6084/m9.figshare.4956353](https://doi.org/10.6084/m9.figshare.4956353), [arXiv:1412.1044](https://arxiv.org/abs/1412.1044).
- Casares (R): Ramón Casares, "On 'On Recursion'";
DOI: [10.6084/m9.figshare.5097691](https://doi.org/10.6084/m9.figshare.5097691).
- Casares (T): Ramón Casares, "On Turing Completeness, or Why We Are So Many";
DOI: [10.6084/m9.figshare.5631922](https://doi.org/10.6084/m9.figshare.5631922).
- Casares (U): Ramón Casares, "Universal Grammar is a universal grammar";
DOI: [10.6084/m9.figshare.4956764](https://doi.org/10.6084/m9.figshare.4956764).
- Chomsky (1959): Noam Chomsky, "On Certain Formal Properties of Grammars"; in *Information and Control*, vol. 2, no. 2, pp. 137–167, June 1959, DOI: [10.1016/S0019-9958\(59\)90362-6](https://doi.org/10.1016/S0019-9958(59)90362-6).
- Chomsky (1965): Noam Chomsky, *Aspects of the Theory of Syntax*; The MIT Press, Cambridge MA, 1965, ISBN: 978-0-262-53007-1.
- Chomsky (2000): Noam Chomsky, *New Horizons in the Study of Language and Mind*; Cambridge University Press, Cambridge, 2000, ISBN: 978-0-521-65822-5.
- Chomsky (2005a): Noam Chomsky, "Some Simple Evo-Devo Theses: How True Might They Be for Language?"; Paper of the talk given to the Morris Symposium on the Evolution of Language, held at Stony Brook University, New York, in October 15, 2005; URL: <https://linguistics.stonybrook.edu/events/morris/05/program>.
- Chomsky (2005b): Noam Chomsky, "Three Factors in Language Design"; in *Linguistic Inquiry*, vol. 36, no. 1, pp. 1–22, Winter 2005, DOI: [10.1162/0024389052993655](https://doi.org/10.1162/0024389052993655).

- Chomsky (2006): Noam Chomsky, *Language and Mind*, Third Edition; Cambridge University Press, Cambridge, 2006, ISBN: 978-0-521-67493-5.
- Chomsky (2007): Noam Chomsky, “Of Minds and Language”; in *Biolinguistics*, vol. 1, pp. 9–27, 2007,
URL: <http://www.biolinguistics.eu/index.php/biolinguistics/article/view/19>.
- Church (1935): Alonzo Church, “An Unsolvable Problem of Elementary Number Theory”; in *American Journal of Mathematics*, vol. 58, no. 2, pp. 345–363, April 1936, DOI: [10.2307/2371045](https://doi.org/10.2307/2371045). Presented to the American Mathematical Society, April 19, 1935.
- Curry & Feys (1958): Haskell B. Curry, and Robert Feys, with William Craig, *Combinatory Logic*, Vol. I; North-Holland, Amsterdam, 1958, ISBN: 978-0-7204-2207-8.
- Davis (1965): Martin Davis (editor), *The Undecidable: Basic Papers on Undecidable Propositions, Unsolvable Problems and Computable Functions*; Dover, Mineola, New York, 2004, ISBN: 978-0-486-43228-1. Corrected republication of the same title by Raven, Hewlett, New York, 1965.
- Davis (1982): Martin Davis, “Why Gödel Didn’t Have Church’s Thesis”; in *Information and Control*, vol. 54, pp. 3–24, 1982, DOI: [10.1016/s0019-9958\(82\)91226-8](https://doi.org/10.1016/s0019-9958(82)91226-8).
- Everett (2008): Daniel L. Everett, *Don’t Sleep, There Are Snakes: Life and Language in the Amazonian Jungle*; Vintage, New York, 2008, ISBN: 978-0-307-38612-0.
- Fitch, Hauser, and Chomsky (2005): Tecumseh Fitch, Marc Hauser, and Noam Chomsky, “The Evolution of the Language Faculty: Clarifications and Implications”; in *Cognition*, vol. 97, pp. 179–210, 2005, DOI: [10.1016/j.cognition.2005.02.005](https://doi.org/10.1016/j.cognition.2005.02.005).
- Friedman & Felleisen (1987): Daniel Friedman, and Matthias Felleisen, *The Little LISPer*, Trade Edition; The MIT Press, Cambridge, Massachusetts, 1987, ISBN: 978-0-262-56038-2.
- Gödel (1930): Kurt Gödel, “Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I”; in *Monatshefte für Mathematik und Physik*, vol. 38, pp. 173–198, 1931, DOI: [10.1007/BF01700692](https://doi.org/10.1007/BF01700692). Received November 17, 1930. Extended by Rosser (1936). English translation in Davis (1965).
- Hamblin (1973): Charles Hamblin, “Questions in Montague English”; in *Foundations of Language*, vol. 10, pp. 41–53, 1973.
- Hauser, Chomsky, and Fitch (2002): Marc Hauser, Noam Chomsky, and Tecumseh Fitch, “The Language Faculty: Who Has It, What Is It, and How Did It Evolved?”; in *Science* 298, pp. 1569–1579, 2002, DOI: [10.1126/science.298.5598.1569](https://doi.org/10.1126/science.298.5598.1569).
- Jackendoff (2011): Ray Jackendoff, “What Is the Human Language Faculty?: Two Views”; in *Language*, vol. 87, no. 3, pp. 586–624, September 2011, DOI: [10.1353/lan.2011.0063](https://doi.org/10.1353/lan.2011.0063).
- Kenneally (2007): Christine Kenneally, *The First Word: The Search for the Origins of Language*; Penguin Books, New York, 2008, ISBN: 978-0-14-311374-4.
- Kleene (1936a): Stephen Kleene, “General Recursive Functions of Natural Numbers”; in *Mathematische Annalen*, vol. 112, pp. 727–742, 1936, DOI: [10.1007/BF01565439](https://doi.org/10.1007/BF01565439).
- Kleene (1936b): Stephen Kleene, “ λ -Definability and Recursiveness”; in *Duke Mathematical Journal*, vol. 2, pp. 340–353, 1936, DOI: [10.1215/s0012-7094-36-00227-2](https://doi.org/10.1215/s0012-7094-36-00227-2).
- Kleene (1952): Stephen Kleene, *Introduction to Meta-Mathematics*; Ishi Press, New York, 2009, ISBN: 978-0-923891-57-2. Reprint of the same title by North-Holland, Amsterdam, 1952.

- Krifka (2011): Manfred Krifka, “Questions”, DOI: [10.1515/9783110255072.1742](https://doi.org/10.1515/9783110255072.1742); in *Semantics: An International Handbook of Natural Language Meaning*, Volume 2, HSK 33.2, K. von Heusinger, C. Maienborn, P. Portner (editors), pp. 1742–1785, De Gruyter Mouton, Berlin/Boston, 2011, ISBN: 978-3-11-018523-2.
- Marr (1982): David Marr, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*; W.H. Freeman and Company, San Francisco, 1982. ISBN: 978-0-7167-1284-8.
- McCarthy (1960): John McCarthy, “Recursive Functions of Symbolic Expressions and Their Computation by Machine, Part I”; in *Communications of the ACM*, vol. 3, no. 4, pp. 184–195, April 1960, DOI: [10.1145/367177.367199](https://doi.org/10.1145/367177.367199).
- McCarthy et al. (1962): John McCarthy, Paul Abrahams, Daniel Edwards, Timothy Hart, and Michael Levin, *LISP 1.5 Programmer’s Manual*; The MIT Press, Cambridge, Massachusetts, 1962, ISBN: 978-0-262-13011-0.
- Miller (1956): George Miller, “The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information”; in *Psychological Review*, vol. 63, No. 2, pp. 81–97, 1956, DOI: [10.1037/h0043158](https://doi.org/10.1037/h0043158).
- Pavlov (1927): Ivan Pavlov, *Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex*, translated by G.V. Anrep; Oxford University Press, London, 1927. Reprinted by Dover, Minneola, NY, 1960, 2003; ISBN: 0-486-43093-6.
- Pierce (1867): Charles Sanders Peirce, “On a New List of Categories”; in *Proceedings of the American Academy of Arts and Sciences*, vol. 7, pp. 287–298, 1868, DOI: [10.2307/20179567](https://doi.org/10.2307/20179567). Presented May 14, 1867.
- Pinker & Jackendoff (2004): Steven Pinker, and Ray Jackendoff, “The Faculty of Language: What’s Special About It?”; in *Cognition*, vol. 95, pp. 201–236, 2005, DOI: [10.1016/j.cognition.2004.08.004](https://doi.org/10.1016/j.cognition.2004.08.004). Received 16 January 2004; accepted 31 August 2004.
- Post (1936): Emil Post, “Finite Combinatory Processes — Formulation 1”; in *The Journal of Symbolic Logic*, vol. 1, no. 3, pp. 103–105, September 1936, DOI: [10.2307/2269031](https://doi.org/10.2307/2269031). Received October 7, 1936.
- Post (1944): Emil Post, “Recursively Enumerable Sets of Positive Integers and their Decision Problems”; in *Bulletin of the American Mathematical Society*, vol. 50, no. 5, pp. 284–316, 1944, DOI: [10.1090/s0002-9904-1944-08111-1](https://doi.org/10.1090/s0002-9904-1944-08111-1).
- Rosser (1936): J. Barkley Rosser, “Extensions of Some Theorems of Gödel and Church”; in *Journal of Symbolic Logic*, vol. 1, pp. 87–91, 1936, DOI: [10.2307/2269028](https://doi.org/10.2307/2269028).
- Shieber (1986): Stuart Shieber, “An Introduction to Unification-Based Approaches to Grammar”; Microtome Publishing, Brookline, Massachusetts, 2003. Reissue of the same title by CSLI Publications, Stanford, California, 1986.
URL: <http://nrs.harvard.edu/urn-3:HUL.InstRepos:11576719>.
- Simon & Newell (1971): Herbert Simon and Allen Newell, “Human Problem Solving: The State of the Theory in 1970”; in *American Psychologist*, vol. 26, no. 2, pp. 145–159, February 1971, DOI: [10.1037/h0030806](https://doi.org/10.1037/h0030806).
- Stabler (2014): Edward Stabler, “Recursion in Grammar and Performance”, DOI: [10.1007/978-3-319-05086-7_8](https://doi.org/10.1007/978-3-319-05086-7_8); in *Recursion: Complexity in Cognition*, Volume 43 of the series ‘Studies in Theoretical Psycholinguistics’, Tom Roeper, Margaret Speas (editors), pp. 159–177, Springer, Cham, Switzerland, 2014, ISBN: 978-3-319-05085-0.
- Tomasello (2008): Michael Tomasello, *Origins of Human Communication*; The MIT Press, Cambridge, Massachusetts, 2008, ISBN: 978-0-262-51520-7.

- Turing (1936): Alan Turing, “On Computable Numbers, with an Application to the Entscheidungsproblem”; in *Proceedings of the London Mathematical Society*, vol. s2-42, no. 1, pp. 230–265, 1937, DOI: [10.1112/plms/s2-42.1.230](https://doi.org/10.1112/plms/s2-42.1.230). Received 28 May, 1936. Read 12 November, 1936.
- Turing (1937): Alan Turing, “Computability and λ -Definability”; in *The Journal of Symbolic Logic*, vol. 2, no. 4, pp. 153–163, December 1937, DOI: [10.2307/2268280](https://doi.org/10.2307/2268280).
- Vygotsky (1934): Lev Vygotsky, *Thought and Language*, newly revised and edited by Alex Kozulin; The MIT Press, Cambridge, Massachusetts, 1986, ISBN: 978-0-262-72010-6.
- Watumull et al. (2014): Jeffrey Watumull, Marc Hauser, Ian Roberts & Norbert Hornstein, “On Recursion”; in *Frontiers in Psychology*, vol. 4, article 1017, pp. 1–7, 2014, DOI: [10.3389/fpsyg.2013.01017](https://doi.org/10.3389/fpsyg.2013.01017).
- Zermelo (1908): Ernst Zermelo, “Untersuchungen über die Grundlagen der Mengenlehre I”, DOI: [10.1007/978-3-540-79384-7_6](https://doi.org/10.1007/978-3-540-79384-7_6); in *Ernst Zermelo Collected Works* Volume I, H.-D. Ebbinghaus, A. Kanamori (editors), pp. 160–229, Springer-Verlag, Berlin Heidelberg, 2010, ISBN: 978-3-540-79383-0.