The CC-Theory of the Origin of Language

Jan Odijk, UiL-OTS, Utrecht University

j.odijk@uu.nl

Version 3, Sep 8, 2011

Abstract

In this paper I propose a research programme based on a theory called the CC-theory, consisting of three very tentative and speculative hypotheses that together account for the origin of the major aspects of natural language. The core hypothesis (which I will call the Conceptual Copy Hypothesis or CC-Hypothesis) states that a very small change in the genes of our ancestors had the effect that a second copy of the conceptual-intentional (C-I) component develops: this small change at the genotype level is argued to have dramatic consequences at the phenotype level: this new copy of the C-I component starts to function as the grammatical component and many properties of the grammatical component are derived from assuming this origin. The second hypothesis states that this new component makes a link with the already existing system to generate and interpret mouth-produced sounds, gestures and facial expressions in use in primates and our direct ancestor for emotive calls and social interaction. It thus accounts for the fact that the primary media for language are speech and gesturing, and, together with independent facts and assumptions, for the fact that natural language uses symbolic reference rather than indexical or iconic reference. The third hypothesis states that integrating two elements into a structure (structure building) is an operation in an independent module, unique to humans but not to language. Recursion can arise from this structure building under specific conditions. It thus accounts for the fact that structure building (and in many cases also recursion) can, as I will argue, be used for objects from many faculties and modules, both linguistic ones (orthography, phonology, morphology, syntax, and discourse grammar) and non-linguistic ones (counting (mathematics), music, dance, thinking, vision, artificial languages in logic, mathematics and programming). The CC-theory is tentative and highly speculative. However, even if the theory I propose here would turn out to be completely wrong, I believe it is interesting to study since the theory, in my view, has all the right properties of a theory of the origin of language, so the correct theory of the origin of language should have the same or similar properties.

1 Introduction

In this paper I propose a research programme based on a theory, called the CC-theory, consisting of three very tentative and speculative hypotheses that together account for the origin of the major aspects of natural language. The three hypotheses are independent o f each other, i.e. each could individually turn out to be false without affecting the status of the other hypotheses, but together they form a coherent whole to account for the major aspects of the origin of language.

The core hypothesis (which I will call the Conceptual Copy Hypothesis or CC-Hypothesis) states that a very small change in the genes of our ancestors (most probably in the genes regulating the development of the body, more specifically the brain) had the effect that a second copy of the conceptual-intentional (C-I) component develops: this small change at the genotype level is argued to have dramatic consequences at the phenotype level: this new copy of the C-I component starts to function as the grammatical component and many properties of the grammatical component are derived from assuming this origin, including the existence and nature of the basic grammatical units, the fact that these are not atomic but have properties, the combinatorial nature of the grammatical component, the fact that open class and closed class word types exist, the precise properties for each word type, and also the word types themselves. The hypothesis also predicts that a large lexicon (at least several tens of thousands of entries) can be processed efficiently in the grammatical component, and it even makes (intensional) predictions on what kind of mechanism is used for that. Finally it also accounts for the fact that the grammatical system is such a learning-eager (or acquisition-eager) component to such a degree that language learning (as suggested by Chomsky), is better described as *language growth*. The hypothesis also shines a new light on the relation between grammar and semantics, in particular it offers a completely new perspective on the intuition of many linguists that grammatical categories are somehow grounded in semantic categories, but it does so while maintaining the independence and the autonomous character of the grammatical component. It also predicts that there is a universal set of grammatical properties, from which each language uses a subset, (contra [Evans & Levinson 2009]) and provides a new source of evidence for determining the exact nature of this set. The basic idea, if correct provides a solution for the paradox that language evolution must have involved just one or a few very small changes while at the same time the grammatical component is a very rich and intricate system that is to a large extent specific to humans and to language.

The second hypothesis states that this new component makes a link with the already existing system to generate and interpret mouth-produced sounds, gestures and facial expressions (I will call these the *SGF-modules*) in use in primates and our direct ancestor for emotive calls and social interaction. It thus accounts for the fact that the primary media for language are speech and gesturing, and, together with independent facts and assumptions, for the familiar sound-meaning relation, including its arbitrariness and the fact that natural language uses *symbolic reference* rather than indexical or iconic reference.

The third hypothesis states that integrating two elements into a structure (structure building) is an operation **in an independent module**, unique to humans but not to language. Recursion can arise from this structure building under special conditions (i.e. if a structure formed by (inter alia) a particular structure building operation can enter the same structure building operation again). It thus accounts for the fact that structure building can, as I will argue, be used for objects from many faculties and modules, both linguistic ones (in orthography, phonology, morphology, syntax, and discourse grammar) and non-linguistic ones (counting (mathematics), music, dance, thinking, vision, and in artificial languages in logic, mathematics and programming), and in many of these also recursive structure building occurs. I will argue that structure building plays a crucial role in alleviating the restrictions that Short Term Memory (STM) imposes on processing input and output.

Though the theory proposed here is a theory on the origin of language, it is a theory that is falsifiable, since it basically describes the organization of the human brain with regard to language and how it differs from the brain of our direct ancestor and our fellow primates. The theory also makes very clear linguistic predictions. Evidence in favor or against the theory can therefore be obtained in principle and in practice from linguistics (orthography, phonology, morphology and syntax), psychology, human biology, primate biology and comparative biology, and for certain aspects also from other human faculties such as counting, music, dance, vision and from cognitive (thinking) processes. Even if the theory I propose here would turn out to be completely wrong, I believe it is interesting to study since the theory, in my view, has all the properties that a theory of language origin should have, so the correct theory of language origin should have the same or similar properties.

2 Context

Hauser, Chomsky & Fitch 2002 [HCF 2002] make a specific proposal about the evolution of natural language. They observe that there is general agreement that “although bees dance, birds sing, and chimpanzees grunge, these systems of communication lack the rich expressive and open-ended power of human language”, which, they claim is “based on humans’ capacity for recursion”. The evolutionary puzzle, therefore, lies, as they state, “in working out how we got from there to here, given this apparent discontinuity”. A second issue they mention revolves around whether the evolution of language was gradual versus saltational. And as a third issue, they mention “continuity” versus “exaptation”, i.e. “whether human language evolved by gradual extension of preexisting communication systems, or whether important aspects of language have been exapted away from their previous adaptive function”

In order to make progress on these questions, [HCF 2002] distinguish Faculty of the Language – narrow sense (FLN) from Faculty of Language – broad sense (FLB). FLN contains components that are both unique to humans and unique to language. FLB contains all components that play a role in the knowledge and processing of language, also the ones that are not both unique to humans and unique to language. By definition, FLB contains FLN. [HCF 2002] argue that FLB includes, apart from FLN, at least two organism-internal systems, the “sensory-motor” (S-M) system and the “conceptual-intentional”(C-I) system. They characterize the content of FLN as the abstract linguistic computational system alone, independent of the other systems with which it interacts and interfaces:

“A key component of FLN is a computational system (narrow syntax) that generates internal representations and maps them into the sensory-motor interface by the phonological system, and into the conceptual-intentional interface by the (formal) semantic system. All approaches agree that a core property of FLN is recursion, attributed to the narrow syntax in the conception just outlined.” [HCF 2002: 1570-1571]

And they add that “This capacity of FLN yields discrete infinity”.

Concerning evolution of language, [HCF 2002: 1573] suggest that FLB as a whole is not uniquely human, that it has homologues (perhaps in less developed or otherwise modified form) in nonhuman animals, and that only FLN is uniquely human. Next, they appear to equate FLN with the computational mechanism of recursion (“we suggest that FLN- the computational mechanism of recursion –is recently evolved and unique to our species”, p. 1573). In fact, just a few sentences later, they make this explicit, by stating that “in fact, we propose in this hypothesis that FLN comprises **only** the core computational mechanisms of recursion as they appear in narrow syntax” (emphasis mine) though they include its “mappings to the interfaces”, and they conclude that FLB , as a whole, thus has a long evolutionary history, long predating the emergence of language”, while “the status of FLN as an adaptation is open to question”, since “the argument from design” is “nullified”.

Chomsky [2004] elaborates on this:

“If this general picture has some validity, than the evolution of language may be a very brief affair, even though it is a very recent product of evolution. Of course, there are innumerable precursors, and they doubtless had a long evolutionary history. For example, the bones of the middle ear are a marvelous sound-amplifying system, wonderfully designed for interpreting speech, but they appear to have migrated from the reptilian jaw as a mechanical effect of growth of the neocortex in mammals that began 160 million years ago, so it is reported. We know far too little about conceptual systems to say much, but it’s reasonable to suppose that they too had a long history after the separation of hominids, yielding results with no close similarity elsewhere. But the question of evolution of language itself has to do with how these various precursors were organized into the faculty of language, perhaps through some slight genetic event that brought a crucial innovation. If that is so, then the evolution of language itself is brief, speculations that have some bearing on the kind of inquiry into language that is likely to be productive.”

[Pinker & Jackendoff 2005] (P&J2005), however, take a different view on the evolution of language, "namely that the language faculty, like other biological systems showing signs of complex adaptive design, is a system of co-adapted traits that evolved by natural selection" (p. 204),[1] and they support their view by pointing out that there is much more to the language faculty that is highly specific to human language (FLN) than just recursion and interfaces to other components. Restricting attention to syntax, they mention examples such as words, word order, agreement, and case marking, all kinds of grammatical words of categories such as auxiliary, complementizer, determiner, etc, and grammatical categories such as case, gender, voice, mood, etc. P&J2005 suggest that it is likely that such aspects of language are specific to human language: "it is specialized machinery for regulating the relation of sound and meaning" (P&J2005: 216)

In this paper I adopt the position suggested by [HCF 2002] that the central part of the language faculty (FLN) is a relatively recent development and is not an adaptation evolved by natural selection. On the other hand, I also agree with P&J2005, that the central part of the language faculty (FLN) is more than just recursion and interfaces to other systems, but is a richly structured system with many properties that are highly specific to humans and to human language (and where recursion, in my view, is not specific to language).

This creates a paradox, and I tentatively, and very speculatively, propose a theory consisting of three specific hypotheses to resolve this paradox for at least the majority of these properties.

3 Resolving the Paradox

3.1 General

One consequence of the hypothesis that the central part of the language faculty (FLN) is a relatively recent development and is not an adaptation evolved by natural selection is that the change that caused it must have been very small. After all, a big change is very unlikely to lead to a well-functioning organism and is unseen elsewhere in evolution. Indeed, the change that [HCF 2002] propose, namely only the addition of a recursive combinatorial system, may be considered small, and thus consistent with their suggestion that FLN is not an adaptation evolved over time. But this is in conflict with P&J2005's

[1] Consistent with earlier views on language evolution that they have published, e.g. [Pinker & Bloom 1990], [Pinker 1994:333], [Jackendoff 2002: chapter 8].

observation (correct in my view) that FLN is a richly structured system with many properties that are highly specific to human language. And if the hypothesis by [HCF 2002] is to be interpreted in such a way that recursion is *just one* of the components of FLN (as Fitch, Hauser and Chomsky 2005 [FHC 2005:183] emphasize: “at a minimum, then, FLN includes the capacity of Recursion”, then [HCF 2002] face the same problem.

We start our attempt to resolve this paradox by making a very general observation that is not specific to language at all. The observation is that the change or changes involved must be small at the **genotype** level, but this does not necessarily imply that the result of the small change at the genotype level is also small at the **phenotype** level.[2]

One example of this, outside of the domain of language, is the neoteny hypothesis, which illustrates that a small change at the genotype level can lead to a big change at the phenotype level.

Neoteny, also called juvenilization, is the retention, by adults in a species, of traits previously seen only in juveniles In neoteny, the physiological (or somatic) development of an animal or organism is slowed or delayed Ultimately this process results in the retention, in the adults of a species, of juvenile physical characteristics well into maturity. Neoteny has been invoked as an explanation for the evolutionary development of flightless birds, pets such as dogs (claimed to retain juvenile properties of wolfs), domesticated silver foxes, and humans. (from Wikipedia)

The idea that adult humans exhibit certain neotenous (juvenile) features, not evinced in the great apes, is about a century old, and originates with [Louis Bolk, 1926]: "Man, in his bodily development, is a primate foetus that has become sexually mature". Bolk made a long list of such traits. These include sparse body hair, enlarged heads with high relative brain weight, lactose tolerance, the flatness of the human face compared with other primates, the late arrival of the teeth, loss of pigmentation in skin, eyes, and hair, the form of the external ear, the epicanthic (or Mongolian) eyeful, the central position of the foramen magnum (it migrates backward during the ontogeny of primates), persistence of the cranial sutures to an advanced age, the labia majora of women, the structure of the hand and foot, the form of the pelvis, the ventrally directed position of the sexual canal in women, certain variations of the tooth row and cranial sutures, and a dozen of additional features. [Gould 1977]

Neoteny is thus a clear example of a small change at the genotype level (only a few genes for regulating the development of the human body changed) causing a large number of differences at the phenotype level.

---

[2] Cf. [Fitch 2010:55]

Another example where small changes at the genotype level may cause large differences at the phenotype level is formed by the so-called homeotic genes, in particular the so-called *hox genes*, a special cluster of homeotic genes that determine how a cell has to develop depending on its location in the developing body. [Mayr 2002: 110-111; Deacon 1997:165-192; Ridley 1999; Fitch 2010:53-54, p. 216 on 'serial homology' in vertebrates, and p. 219 on 'gene duplication'; Ronshaugen et al 2020]. It was discovered that these genes have this role by mutating them with X-rays: flies with such mutated genes started to develop in unusual ways, e.g. "they had legs where they should have antennae, or wings where they should have small stabilizers called haltered" [Ridley 1999:176]. Transgenic mice in which both copies of the homeotic gene LIM1 (which appears also to be involved in Williams syndrome) have been knocked out fail to develop heads altogether, even though the remainder of the body develops normally [Deacon 1997:273]. So this constitutes another clear example where small mutations at the genotype level result in big changes at the phenotype level, especially caused by mutations in the so-called 'gap' genes, which "had big effects, defining whole areas of the body" (Ridley 1999:176).

We thus have plausible mechanisms, hypothesized and partly observed in completely different areas of evolutionary development where small changes at the genotype level cause large differences at the phenotype level. We suggest that such a mechanism is also responsible for the development of FLN. In fact, we might go further and suggest that this mechanism might be a special instance of the juvenalization process that turned great apes into humans (as argued for by [Koster 1982]), though this is an independent and stronger hypothesis, which requires independent evidence, and which might turn out to be false even if FLN arose through other small changes in genes with large consequences at the phenotype level.

Even if we have plausible mechanisms such as the ones described above to resolve our paradox, they say nothing specific about how FLN arose. For that, we need a theory of such a plausible mechanism that is much more specific to FLN and how it arose. We present, very tentatively and largely speculatively, such a theory in the next section.

3.2 Resolving the Paradox (specifically for language)

Before turning to the theory I propose to resolve the paradox, I will first sketch some basic assumptions. I assume that the mind/brain consists of a number of *modules*. The notion of *module* as used here is inspired by its use in computer science (Chomsky, Pinker and Jackendoff hold similar views on the modular organization of the brain).[3] Each module operates autonomously (i.e., it functions by its own rules / principles / operations only) and is influenced by outputs of other modules only if an explicit interface between the modules enables this. A module is *informationally encapsulated*, i.e. the data structures and operations it uses are invisible to other modules unless they are explicitly made visible via an interface. We may group certain modules in a *faculty* (e.g. several modules that are assumed to account for language in the *language faculty*) to make it correspond to categories that we distinguish in normal life, but any such grouping is just for convenience and has no basis in reality.

I assume the following modules in our direct predecessor:

1. A cognitive-intentional (C-I) module
2. An emotional/affective module
3. A social interaction module
4. SGF modules for recognizing/generating mouth-created sounds, facial expressions and gestures
5. A Short Term Memory Module
6. Sensory-Motor (S-M) modules for the senses of vision, hearing, smell, touch and taste, and for motor control (actions)

It is well known that primates (and other animals) can express emotional states and acts of social interaction[4] by means of (mouth-originating) sounds, gestures and facial movements (e.g. for alarm calls), and that other animals of the same species interpret such sounds and gestures as symbols for emotional states and for social interaction.[5] I assume that primates have dedicated sensory-motor modules (different from the other S-M modules) to deal with such phenomena. I will call these the SGF modules. These modules have no interface to the C-I system, but only to the emotional module and to the social interaction module. Thus primates can express their emotional states by making sounds and gestures, and they can interpret sounds, gestures and facial expressions of other primates (of the same species) as emotional states. In humans, this system also exists and is used for acts such as laughing, sobbing, screaming with fright, crying with pain, groaning and sighing [Deacon 1997:418][6], I believe that the same system is also reused in humans for certain types of learned utterances, in particular for uttering and

---

[3] Other terms used for *module* in the language context include *component*, *subsystem*, *mental organ* and others. See also [Fitch 2010:81] . [Fitch 2010:17] speaks in this context of a 'multi-component approach to language'.

[4] [Fitch 2010:237] states that 'both chimpanzee species […] have a repertoire of grunts, hoots, and screams which play an important role in social behavior'.

[5] See [Hauser 1996] for an overview of animal communication systems in general, See also [Deacon 1997:234-235].

[6] [Fitch 2010:177] agrees and calls this system an 'innate human call system'

interpreting (certain classes of) interjections, and for swearing and expletives (although these clearly do have a linguistic basis).[7] I will come back on this below.

The C-I module has input-output relations with several S-M modules: vision, hearing, feeling, etc. The C-I system can use visual input to interpret images and sequences of images, and it can use hearing to interpret sounds as sounds. However, it does not have an interface specifically to the SGF modules, which deal with the analysis of sounds and gestures made by animates and which, I assume, are separate modules with the important characteristics that all such sounds and gestures are interpreted as symbols, i.e. as forms standing for something else.[8]

The SGF modules were already present before the modern human appeared.[9] It is present in primates, but it is crucially not linked to their C-I component.[10]

Modern humans have all the modules our predecessors have, but also several modules for language, in particular, (1) a grammatical module (encompassing parts of syntax and of morphology), including a lexicon of grammatical units; (2) (what Chomsky calls PF) modules of morphophonology, phonology and phonetics that make the link between the grammatical module and (as I assume) the SGF modules, and (3) a semantic module establishing the link between the grammatical module (including the lexicon) on the one hand and the C-I module on the other.

---

[7] See [Hurford 2003:47] and [Smits 2009:71-74], [Jackendoff 2002: 240] for similar views. [Jackendoff 1994:146] claims that some people having Broca's aphasia (agrammatism) can swear fluently, which might be evidence for the involvement of different modules in 'normal language use' v. swearing. [Pinker 2007:334] discusses this in some detail, and he argues that words of swearing are stored in a different place in the brain than other words.

[8] [Bickerton 2009:80-81] claims that apes "do get" the idea that an arbitrary symbol can stand in for something in the real world. However, apes clearly do not acquire this ability by themselves but only in unnatural circumstances in which they are trained to do so, they do it with great difficulty and it takes them a lot of time; also the size of their resulting mental lexicon remains very limited in comparison to humans' mental lexicons. I believe that these abilities differ qualitatively from humans' abilities (see also [Jackendoff 2002:242]): humans have dedicated systems for this in their brains, while apes, to the extent that they can do it at all, use more general brain systems. I believe that the ape's ability can in certain respects be compared to how humans calculate (e.g. addition, multiplication), an ability for which humans, in my view, have no dedicated modules, which they can generally only do after explicit teaching and by consciously carrying out an algorithm, making many mistakes along the way, requiring additional memory (e.g. in the form of pencil and paper), but aided by remembering all kinds of shortcuts (e.g. multiplication tables) by heart. Humans are typically beaten in this ability by the simplest of electronic devises (calculators, the simplest computers, etc.) in speed and precision. Though I believe this comparison is appropriate in the sense that they both illustrate mental activities for which no dedicated modules exist, surely human calculation –despite its limitations- is of course orders of magnitude better than the apes' abilities to deal with symbols.

[9] [Hauser & Fitch 2003:179] conclude (though tentatively) "we inherited from animals a suite of perceptual mechanisms for listening to speech—ones that are quite general and did not evolve for speech". Our assumptions in the main text are even more specific than this, so they might prove to be false independently of [Hauser & Fitch 2003]'s conclusion.

[10] [Fitch 2010:148] 'animals have surprisingly rich mental lives, and surprisingly limited abilities to express them as signals. Animals possess concepts […] but do not express them as signals'.

The theory that I propose postulates two changes in humans versus our immediate predecessors. I formulate these two postulated changes as two separate hypotheses in the theory. A third hypothesis relates these changes to an already existing module in primates The first and core hypothesis of this theory is called the Conceptual Copy Hypothesis (CC-Hypothesis, or CCH) and can be formulated as follows: a small genetic change causes the development of a second module in the brain, modeled after the conceptual-intentional module (basically a new copy of the conceptual-intentional module), and this new module starts to function as the grammatical module. From now on I will call this second copy the "grammatical module". This grammatical module is the core of FLN, i.e. it is unique to language and to humans.[11] Do note that when I speak of a copy, this holds at the genotype level: once the genes have changed, the result is that two such modules will develop where in an earlier evolutionary stage only one such module would develop.

Since this new module is a separate module, it develops and acts independently and autonomously from the conceptual-intentional module (and from other modules). It may also engage in other input and output relations with other modules than the conceptual-intentional module, and if it does, this will have dramatic consequences for the nature of this module: it will still be structured in the same way as the original C-I module but it contents will be completely different because it has a completely different input.

[Jackendoff 1992:39] already observed the high correspondence between the C-I module and the grammatical module: "syntax presumably evolved as a means to express conceptual structure, so it is natural to expect that some of the structural properties of concepts would be mirrored in the organization of syntax". However, here I go further by assuming that this arose because the grammatical module -- seen from the perspective of evolution -- basically **is** a copy of the C-I module, reinterpreted for different input.

[Fitch 2010:430-431] observes that 'we might profitably seek the origins of some aspects of grammar […] in pre-existing, innately determined concepts'; such an approach 'may have considerable explanatory power in this regard', and 'if much of the observed complexity in syntax derives from the pre-existing conceptual mechanisms involved in cognition […], we would face less difficulty in understanding how they evolved, given the fact that 'two million years is not a lot of time, in evolutionary terms, for a complex suite of syntax-specific mechanisms to evolve and reach fixation'. And such an approach is exactly what I am pursuing here. Fitch also states that 'it less obvious how such biological constraints [on conceptual systems, JO] could lead to the origin of function words, or explain the specific forms that grammatical changes tend to take', and he concludes that 'however powerful conceptual constraints might be helping to explain the evolution of syntax, few scholars think that such constraints can shoulder the entire

[11] I call it the grammatical component (and not the syntactic component) because it includes syntax and aspects of what is traditionally called morphology. Cf. [Selkirk 1982]. This module also contains the lexicon of basic grammatical units, so perhaps the lexico-grammatical module would be a more appropriate name.

burden'. I partially agree with Fitch on these objections,[12] and I will show below that they do not apply to the specific version of grounding grammar in cognition that I propose.

The grammatical module, as soon as it arose, did not make any connection with the S-M modules, but only to the SGF modules dedicated to dealing with sounds and gestures and facial movements made by animals, thus accounting for the fact that human language uses speech, gestures and facial expressions as its modalities.[13] I cannot explain at this point why this happened, and I simply have to stipulate it, to be compatible with (what I believe to be) reality. However, the hypothesis proposed here makes the clear prediction that this fact should follow from the design of the grammar module (and hence also of the C-I module), where the differences in connections between the C-I and grammar module are predicted to follow from their different positions (physiologically and/or functionally) inside the brain.

This kind of development can be caused by a small genetic change, since basically the genetic instructions just have to say that an additional copy of the same module is created. [14] The different input-output relations are determined, as stated above, by the different location (functionally or physiologically) of this new module in the overall brain: it is determined genetically that it will create connections to other modules, but where these connections actually end up is determined by the location of the module, and therefore differs from the connections the conceptual-intentional system makes even though the genetic blueprint for the two modules is identical.

Obviously, this hypothesis on the evolution of language is highly tentative and speculative, but it makes concrete predictions on the structure of the brain of the modern human, the modules it contains, the relations they have with other modules, and the internal organization of these modules. As such, it is a hypothesis that is (in principle) empirically testable by studying the modern human brain and its activity patterns and comparing it with brains of primates. Of course, studying the organization and working of the modern human brain is quite formidable a task, and though enormous progress has been made in this area with technologies such as fMRI, PET and CAT scans, it appears that so far only rather global properties of brain structure and brain activities can be

---

[12] For example, I do not understand why it is less obvious to him how it could lead to the origin of function words: after all, in a certain sense one could see the whole enterprise of formal semantics ([Montague 1974] and related work), that Fitch is familiar with ([Fitch 2010: 120] as the semantics of function words.

[13] Cf. [Fitch 2010:495] 'the study of living apes shows that sophisticated gestural capabilities were *already present* in the LCA [Last Common Ancestor, JO]'. I cannot explain why the speech modality is much more dominant than gesturing (as in sign language).

[14] The human (and animal) body has many components in two copies, e.g. two eyes, two ears, two arms, two lungs, two legs, etc. This duplication surely arose in a different time frame than language, but it remains to be investigated whether the mechanism I propose here is the same as or similar to the one that accounts for these duplicates. [Fitch 2010:54] mentions *gene duplication* as 'a core trick' (p. 55); small changes in regulatory genes 'play a critical role in evolution', and even though 'we find little evidence that such [homeotic] mutations have gone on to form successful new species', 'rather drastic phenotypic changes might potentially be beneficial', 'particularly after gene duplication', and 'regulatory genes provide an unquestionable path by which minor genotypic changes can have a major phenotypic effect'. [Fitch 2010:55.]

reliably monitored.[15] In any case, I cannot provide empirical evidence of this nature in this paper. My only hope is that this paper may lead to a research programme that will explore the theory suggested here in a systematic way, making use of all the evidence that can be found. Before starting on this, however, I would like to provide several arguments that make the hypothesis postulated at least plausible as a good candidate for such an investigation.

Before discussing the grammatical module itself, I first have to make some assumptions on the organization and working of the conceptual module. I basically follow [Jackendoff 1992] who describes the human conceptual module, but, of course, my assumptions must be about the C-I system of the predecessor of modern humans, i.e. of hominids before they had language. It is, of course, almost impossible to obtain empirical evidence for this, and the only way so far to assess the plausibility of these assumptions is to make comparisons with the C-I system of our closest relatives, and with the C-I system of young children in a pre-language phase. Despite these problems, I assume that the conceptual-intentional module operates on basic units that I will call concepts (*I-concepts* in [Jackendoff 1992:22]).[16,17] The C-I module in my view does not contain innate concepts (see also [Jackendoff 1994:190]), but it innately determines to a high degree how concepts can be made, how they are structured, what properties they can have, how they are distinguished from other concepts, what relations they can have with other concepts, etc.; most if not all of these aspects are genetically determined: [Jackendoff 1992:25-26; 43] speaks of a *grammar of lexical concepts*.[18] It also has mechanisms to determine, based on input from the external world (e.g. via the visual, tactile, olfactory and auditory senses) which concepts to form and how to distinguish them from each other. In other words, the development of the C-I-module, its internal organization and working is to a high extent determined genetically, but at least for primates including humans there is plasticity in the sense that its input from other modules determines how exactly the C-I module is filled with concepts. The great (phenotypic[19]) plasticity of the C-I module in hominids accounts for the fact that humans can continue to form new concepts during all of their life.

---

[15] See [Musso 2003] , [Mitchell et al.2008] and reference there, and [Pallier et al 2011] for some relevant work.

[16] [Fitch 2010:122-123] states that ‘concepts exist prelinguistically: we can have concepts before we know words for them’

[17] Cf. [Hurford 2003:44] ‘Many species […] may not possess very many concepts, but they do nevertheless posses them’; [Fitch 2010:18] states: ‘many aspects of conceptual structure, and thus components of semantics, are shared with primates and other animals’, and ‘animals can also combine learned cognitive representations and incorporate them into novel behavior in useful ways’ (p. 151), and ‘there can be little doubt that animals “think”, though perhaps unconsciously’ (p.151). See also p. 201.

[18] [Fitch 2010:125], discussing studies on language acquisition, concludes that ‘studies on children’s acquisition of word meanings provide strong arguments for innate constraints on human conceptual abilities’, and that the hypothesis space for possible meanings ‘appears constrained in certain ways’ (p. 126).

[19] [Fitch 2010:28]

Concepts are not atomic but have properties ([Jackendoff 1992:25] speaks of *composite* lexical concepts). Some of these properties they share with other concepts. I will notate properties of the C-I module in capitals.

I also assume that the C-I-module has options (though only limited options) for combining concepts, e.g. for combining a concept for an action with (some of) its participants.[20] However, I tentatively hypothesize that this limited combinatory potential does not include recursion or even structure building. Nothing in my proposal crucially depends on this assumption, but I believe that there is a better alternative explanation for structure building and recursion. We will come back on the issue of recursion, and its role in various modules (including the C-I module), if any, in section 7. The C-I module as conceived here is independent of language and receives and interprets input from the visual, auditory, tactile, olfactory and taste senses [Jackendoff 1992:33]. Humans have voluntary control over their conceptual module (they can determine what to think about, and this is not triggered by the direct environment). According to [Hurford 2003:45] some animals may have this to a limited degree, so this property is not fully dependent on language.[21]

In short, the C-I component contains instructions on how to build and combine concepts, and, after some development with the right kind of input in the proper environment, it also contains a set of concepts.

Given these assumptions about the C-I module, let's see how the basic grammatical module develops. By hypothesis, it should contain similar phenomena as the C-I module, and indeed we find correlates for (1) basic units and their properties, (2) input-output relations and (3) combinatorics that highly resemble the C-I phenomena, though they are different due to different input-output relations and the autonomous development of the grammatical module. We will discuss each of these in more detail

[20] [Bickerton 209:80-81] claims that chimpanzees and bonobos, when artificially taught language, spontaneously combine signs (though seldom, and never more than 2 signs), and these include combinations with the interpretation of predication. If he is correct, it could be interpreted as evidence for the non- and pre-linguistic nature of this ability. [Hurford 2003:45] states: about apes 'it seems certain that they have mental representations in predicate argument form'. See also [Jackendoff 2002:238]. [Fitch 2010:185] concludes that 'in many species there are rules (or constraints) upon vocal sequences that can reasonably be termed "animal syntax"', but 'the types of rules that govern these arrangements are very simple compared to human linguistic syntax' and they can be 'captured by trivial finite state grammars'.

[21] [Fitch 2010:180] even claims that there is some limited control over the expressions of emotions in chimpanzees: 'withholding [calls, JO] is difficult, but possible, for chimpanzees in the wild'. He goes on to illustrate this with the example that 'chimpanzees 'may cover their mouths to avoid vocalization'. But this example just illustrates, in my view, that the expression of emotions is automatic, but that a separate module (I assume the chimpanzees' C-I module) 'realizes' that expressing the emotion might be harmful, and it tries to avoid its expression by using a different mechanism that it does have voluntary control over (motor control of the hands) to minimize the effects of the fully automatic expression of emotions. If this analysis is correct (and similar phenomena also occur in humans, of course), it does not argue against but rather provides additional support for my claims in the main text.

Basic Units. There are basic units in the grammatical module, as in the C-I module. I will call these *basic grammatical units*. And the basic grammatical units are not atomic, but have properties. I will notate grammatical properties in lower case.
I use the term *basic grammatical unit* and not the term *word*. That is intentional. The term *word* is, unfortunately (but characteristically for human language), highly ambiguous, so I think it should be avoided anyway.[22] But if we restrict attention to the interpretation as a specific (somewhat autonomous) type of grammatical unit, as a type, it can be said that many basic grammatical units correspond to words. However, basic grammatical units can also be realized as roots, stems, affixes or clitics. The grammatical module thus contains instructions on how to form basic grammatical units, and, after some development with the right input, will contain a set of basic grammatical units (a lexicon).

The basic grammatical units and sounds both involve form, and I will distinguish them sometimes as "abstract form" (for grammatical units) v. "concrete form" (for sounds) for reasons that will become clear below. For clarity, some simple examples from English can illustrate the difference between abstract form and concrete form. Consider the following examples:

(1) Concrete form
  a. Boy-s
  b. men
  c. child-ren
(2) Abstract form:
  a. [lemma=boy, number=plural]
  b. [lemma=man, number= plural]
  c. [lemma=child, number= plural]

The examples in (1) involve, at the concrete form level, 3 different word forms with 3 different morphological phenomena (2 different suffixes and ablaut). However, the same 3 different word forms in (2), at the abstract form level, share one morphosyntactic property (number=plural). All grammatical theories of English agree that the concrete form words in (1) should be analyzed at an abstract form level as in (2). Notice also that under the hypothesis proposed here, the abstract grammatical units do not contain a phonological representation: the relation between an abstract grammatical relation and its phonological form is created in the PF-component. Of course, in order to maintain the correct relation between the abstract grammatical unit, its meaning, and it phonological representation, the basic grammatical unit must have a unique identifier mediating this relation. Thus *boy*, *man* and *child* in the examples of (2) are not phonological or orthographic representations but unique identifiers for these basic grammatical units. But this is necessary in any case, since an orthographic and/or phonological representation

[22] There is first of all the type v. the token reading; there is the grammatical word v. the phonological word reading; there is the orthographic unit v. the grammatical unit reading; there is the word form v. base form (lexeme) reading; a normal amount of ambiguity one expects in natural language but not really beneficial for carrying out high quality science.

cannot serve this purpose because they are very often ambiguous (e.g. *book* can be either a verb or a noun, and these have totally unrelated meanings).

Input-Output Relations Second, as stated before, the grammatical module will make input-output relations, in particular with the SGF modules (speech perception and production, gesture interpretation and production). The relation made between the grammatical component and the speech production and perception centers is mediated by what is usually called the PF-component[23], which starts its existence due to this link being made. I have nothing to say about the PF-component in this paper, though we will briefly come back to it when discussing recursion and structure building in section 7.
A link is created between the grammatical module and the C-I module, and we may call this relation the semantic module. I also have nothing to say about the semantic module, except that I expect it to be relatively simple because of the close correspondence between the C-I module and the grammatical module, at least much simpler than the PF module. See [Chomsky 2005:4,13] for a similar view and below for some examples of the complexity of the PF-modules.
The link between the C-I module and the SGF modules (indirectly, mediated via the grammatical component) creates the familiar pairing of sound and meaning.[24] Via this link the brain connects concepts from C-I to sounds via the grammatical units.
Thus the SGF subcomponents, which developed independently of language and existed long before language arose, are re-used here for a different purpose. A clear example of a component belonging to FLB but not to FLN.

It is important to be very explicit about what I do claim here and what I do not claim. It is generally assumed that human language cannot be seen as a continuation or extension of animal communication systems used for alarm calls or for social interaction. I fully agree. However, what I claim here is that the animal communication systems use two (groups of) modules and a link between them. These (groups of) modules are the emotive/affective and social interaction modules on the one hand and the SGF modules on the other. What I claim is that only the SGF modules are reused for human language.[25] When a particular kind of emotion is evoked (e.g. because the animal sees a predator), it causes the corresponding sound/gestures to be made (and this can hardly be stopped since animals have no control over this). Reversely, when the relevant sound/gesture is made, the corresponding emotion is automatically evoked (and that is uncontrollable as well).

---

[23] The PF component includes aspects of morphology that one might call "interpretive morphology", i.e. the realization of grammatical properties and morphemes as allomorphs, morphophonology and phonology. As for gesturing in sign language, [Fitch 2010:437] states that sign language has sign 'phonetics and phonology', so a similar module as the PF-module links the grammatical module and the actual gestures.
[24] However, given the crucial intermediate role of the grammatical component, this is in my view better represented as a triple (sound, grammar, meaning).
[25] [Fitch 2010:174] states that 'certain aspects of these [primate communication, JO] systems do seem to provide potential precursor mechanisms to mechanisms involved in human language (e.g. the ability to interpret signals made by others as meaningful), while other do not (e.g. the innately determined structure of primate vocalizations […]).

It has been argued ([Deacon 1997], [Bickerton 2009:47]), that the kind of reference used in animal communication systems is qualitatively different from reference as used in human language. I agree. The two kinds of reference are usually called indexical v. symbolic, following [Peirce 1978:143] (who made a trichotomous distinction[26] between iconic, indexical and symbolic reference). I tentatively propose that the difference in reference follows from the theory proposed here. Let's first be clear about what is involved in reference. According to [Peirce 1978:156], 3 entities must be distinguished: the Representamen, the Object, and the Interpretant (the idea evoked in the interpreter's mind). According to Peirce, a Sign is a Representamen with a (mental) Interpretant. Signs, in my view, only refer to internal mental objects and states, never to real world object outside of a mind/brain. There is a relation between internal objects and states and outside reality, but that relation is very complex and, in my view, irrelevant for Signs. If this is correct, we only have to consider the Representamen, the Interpretant, and their relation.

When comparing animal calls with human language, the kind of Representamen is constant (mouth-originating sounds, gestures), but the Interpretants are different and the link between the Representamen and the Interpretants is different, so a different kind of reference can arise.
Indexical and symbolic reference share the arbitrariness of the relation between sign and signified (in contrast to iconic reference). Thus the fact that the SGF subcomponents allow for such arbitrariness accounts for the fact that this also happens in human language, with no or only a minor role for iconic reference. With indexical reference there is always a physical or temporal connection between Representamen and the Object (the Object is 'really affected' by it), while no such connection need be present in symbolic reference (the *displacement property*). In my view, this is so because for Signs the relation with the Object is irrelevant for reference. Though it is in no way crucial to the main thesis of this paper, I tentatively propose that the difference in reference kind is a consequence of the following: first, as stated above, the Representamens that we are talking about refer to brain-internal states (emotions, or concepts or states-of-affairs in the C-I component); second, the emotive system is not under voluntary control (is part of the autonomous neural system); the C-I component, however, **is**, to a much greater extent, at least in humans (one can voluntarily decide to think about one thing instead of another). I therefore submit that the crucial difference is that the kinds of entities referred to differ significantly (under control or nor). In fact, there is an asymmetry: one can think about something even if it is not present in the direct environment; and when one thinks about something, one does not necessarily speak about it (the C-I component does not have to send it to the grammatical module and then further to the SFG modules); however, it is as far as I can see impossible for people not to interpret a sound from their native language as referring to some internal entity or state (see also [Pinker 2007:332] 'understanding the meaning of a word is automatic').[27] So the SGF modules interpret

[26] Actually, this was just one of multiple distinctions Peirce made

[27] If one's attention is focused on something else, the sound may not enter the SGF modules, and the speech not interpreted. Noninterpration may also happen in very special circumstances, e.g. I sometimes when

sounds and send them to the grammar component in a fully automatic and uncontrollable manner. In short, the account I propose here explains why human language uses symbolic reference (in contrast to animals).

[Deacon 1999:83], if I understand him correctly, also requires combinatorial potential for signs showing symbolic reference. I think that I disagree,[28] but anyway, limited combinatory potential for signs comes for free under my hypothesis, and unbounded combinatory potential arises thanks to structure building (see section 7), so in that respect my theory is compatible with [Deacon 1999].

Note that the link between the C-I module and the SGF modules is a new development made possible by the development of the grammatical module. In non-human primates, there is no grammatical module, and also no link between the C-I module and the SGF modules. That explains why humans can and primates cannot express their thoughts. As stated above, the emotional system does have a link with the S-M system, both in primates and in humans (and perhaps in many more animals). Primates use sounds, gestures and facial expressions to express emotional states and for social interaction. Gestures and facial expressions are of course also used by humans to express emotional states (sobbing, laughing, etc.). As claimed above a relic of using speech sound for emotional states is present in humans in the form of interjections. Indeed, interjections are used for social interaction ('pragmatic functions', greeting, etc,), and expressing emotions and sensations.[29] Since interjections are the result of a direct link between the emotional module and the social interaction module on the one hand and the SGF modules on the other hand, they have different properties than speech sounds that express concepts via the grammatical module: they do not have combinatorial potential, they do not allow any morphological variation (except repetition), and they need not satisfy phonological constraints that hold for sounds expressing concepts via the grammatical module, e.g. interjections can contain different sounds than the language allows (e.g., in Dutch, an alveolar click (expressing disapproval); an alveolar lateral click to incite a horse; trilled bilabials (*brr*) to express feeling cold, etc.) and they can violate phonotactic rules (e.g., in Dutch, interjections not containing any vowels (e.g. *sst, sht, pst, pff*), or ending in a lax vowel (*bah, hè*)) [e-ANS http://www.let.ru.nl/ans/e-ans/index.html].

There are other differences between animal language and human language. For example, in many species, especially our closest relatives, the relation between an emotion and a sound/gesture is not learned but innate (cf. e.g. [Hurford 2003: 48]), while the relation between sound and meaning in human language is of course learned. I cannot explain this

---

intensely studying a particular word from my native language (Dutch) fail, for a short while, to recognize it as a word of my native language—I do not get its meaning but only its form (letters, sound), and it gives the sensation as if it is a word from a foreign language (though consciously I know that it is a word of Dutch). It is unpredictable when this happens, it is not controllable, and the sensation, though not unpleasant, usually lasts very brief. It is perhaps caused by overloading the brain (and thus a performance error).

[28] [Jackendoff 2002:239] also explicitly rejects this assumption by Deacon.

[29] A third function is to express sounds made by humans, animals and entities (clocks, bells etc) by means of onomatopoeic expressions. I am not sure that a correlate of this exists in animals.

difference.[30] Clearly, the animal systems are superior here from a communicative point of view (in human language the relation must not only be learned, but it also differs from community to community, creating a formidable obstacle for successful communication).[31]

Combinatorics Third, since the C-I module has limited combinatorial potential, the grammar module has so too, and thus one can make (meaning, grammar, sound) triples not only for concepts and basic grammatical units but also for (limited) combinations of concepts and combinations of basic grammatical units   In particular, there are grammatical mechanisms for combining words (e.g. a word with its complements). However, since (by assumption), the C-I module has no structure building and no recursivity, the grammatical module has no structure building or recursivity either.

4 Correspondence between the C-I module and the Grammatical module

Let's now look in more detail to the correspondences between the C-I module and the grammatical module, which are claimed to exist under the hypothesis proposed here.
As we stated above, the C-I module is highly structured, and though probably no concepts are inborn, the way concepts can be formed and structured, and in particular the features that they are composed of are genetically determined. Obviously, proper development of concepts is possible because of the plasticity of the module but will take place only when there is appropriate input.
Exactly the same is true for the grammatical module. Though this module has the same structure and the same development plan as the C-I module, its input is of a completely different nature, so that it will develop differently.

I assume that a part of the conceptual system is more or less predetermined, given appropriate input. The correlates of these largely predetermined parts in the grammatical module develop into grammatical properties (attribute-value pairs, or features) and the closed class lexical items. Since this part of the conceptual module and of the grammatical module is more or less predetermined, its development is finished after a specific period, and cannot be changed or extended anymore, accounting for the degrading capacity of humans to learn language for these aspects over time. Of course, due to different input from the concrete forms this may differ somewhat from language to

[30] [Fitch 2010:175] claims that 'human linguistic signals [..] *must* be learned if they are to provide an open-ended shared lexicon of signals appropriate for linguistic communication'. I disagree. If humans would have a finite number of basic concepts with innate fixed associated forms, and recursive mechanisms to combine these concepts and their associated forms into an indefinite number of meaningful expressions, then I do not see why the relation between the concepts and the sounds could not be innate. The world would be much better off, if that were the case, but alas, natural language does not work this way.
[31] But this does not surprise me. I do not believe that language has a function or that its nature is determined by its function(s), but if I was forced to adopt such a point of view, I believe a strong case can be made for the claim that the function of language is to make communication impossible or at least as difficult as possible (and that case is much stronger than the case for the claim that the function of language is communication and that communication co-determines the nature of language). See [Odijk 1993].

language, and each language will show only a proper subset of the potential, but the overall structure of the closed class items and their properties is identical across languages.

The major part of the conceptual system does not lead to predefined concepts, but it has the ability to form new concepts and it uses the predetermined conceptual properties to characterize new concepts. In the grammatical module, new basic grammatical units can be created and such units will belong to any of the open class syntactic categories (noun, verb, adjective, maybe adpositions), and this can be done during one's whole life (as with concepts).[32] It is very fortunate that the grammatical module has the ability to create new basic grammatical units: if it would not have this ability, we would be able to come up with new concepts but not be able to express them (or perhaps only by descriptions)! In our theory, the fact that the grammatical module is able to do this just as the conceptual module is no accident: the grammatical module is, after all, in our view identically structured as the conceptual module but simply fed with different input.[33]

A remarkable fact about human natural language is that it is so effortlessly and automatically acquired by human children, who otherwise have very limited cognitive abilities. How can this fact be explained? If the grammatical module is identical in structure to the conceptual-intentional module, this property may be derived from the latter module. And indeed, though this is not so often emphasized, the acquisition of a C-I module also goes in a fully automatic manner, again apparently without any effort. Both systems need input in order to develop appropriately: in the case of the C-I-system it uses input from all senses, while the grammar module only uses input from the SGF modules. It is also predicted that the critical period for such acquisition will be identical for both modules, though both depend on input from outside passed on by other modules that have to develop, at least partially, first (e.g. vision for the C-I module, and processing speech input (SGF modules, part of PF) for the grammatical module). It is difficult to really assess whether this prediction is correct, but it appears plausible.

The number of basic grammatical units in a language system is pretty large. One must assume that at least several tens of thousands of individual words are present in any (mature) individual and then we are not even mentioning idiomatic expressions and other types of multi-word expressions. A remarkable fact of the language module is that it apparently is able to access these basic grammatical units in an extremely fast and efficient manner. In understanding language, it must find the relevant basic grammatical unit given a phonetic or phonological string, and in language production, it must find the relevant basic grammatical unit out of a set of several tens of thousands on the basis of a

[32] This appears to be true even for people who have been deprived of normal language input in their earlier years. [Jackendoff 2002:242] reports that "Genie" (see [Curtiss 1977]) did acquire new vocabulary right after her discovery, but was unable to develop proper grammar.

[33] [Fitch 2010:411] states that 'children appear to use similar cognitive resources for learning facts as for word meanings', citing [Markson & Bloom 1997], especially for accounting for so-called 'fast mapping'. The latter authors (p.814) also state that 'fast mapping' does not apply to any arbitrary memorization task', although they cannot characterize the exact scope of 'fast mapping'. I tentatively interpret their findings as evidence in favor of my hypothesis.

concept. Under our theory, this fact has a principled explanation: the grammar module has this property because the conceptual module has this property: the conceptual module, after all, must be able to find concepts very quickly in order to appropriately interpret a situation given certain sensory (e.g. audio-visual) input, and act upon it if needed. The development of these systems in the conceptual module may have (and probably has) a long evolutionary history, but the grammatical module, being a copy of the conceptual module, gets these properties for free with no or only little evolutionary history.

As I said, the basic grammatical unit is the correlate in the grammatical module of the concept in the conceptual module. Though it is not relevant for the theory being developed here, there may even be a default relation between concepts and basic grammatical units (semantic bootstrapping hypothesis). But, as we will see in many other correlates as well, this is just a default relation and therefore one that can be overruled by form aspects quite easily, especially since the modules are autonomous.

Thus some concepts do not correspond to basic grammatical units but to larger units (e.g. compounds with (partially) idiosyncratic meaning, idiomatic expressions and other types of multi-word expressions). Some words are also used without reference to any meaning, either because they simply have no meaning (but are e.g. only used in idiomatic expressions, e.g. *brui* in Dutch *ergens de brui aan geven* 'quit with something'), or because they have a meaning that is not relevant in the pertinent construction. The English verb *do*, when used as an auxiliary is a good example: *do* as a main verb clearly has a meaning, it imposes semantic restrictions on its arguments (basically an animate argument and an action argument), but it loses these properties when used as an auxiliary: it has no meaning or at best identity as meaning, no semantic restrictions on the following VP or on its subject (cf. *does it rain*, *it does not become better*, etc,). The auxiliary is used in a variety of constructions (yes-no questions, wh-questions, certain types of topicalization, negative sentences, emphatic constructions) for which it is not clear at all that they have some meaning aspect in common that could be ascribed to *do*. In addition, each of these constructions can also occur without *do* (depending on the syntactic context), so *do* is not essential to convey this common meaning (if that would exist to begin with). And of course, the standard analysis of these constructions due to [Chomsky 1957] is that *do* as an auxiliary is just a meaningless syntactic means to ensure that an affix does not end up without a stem to attach to. Basic grammatical units thus can be used without any meaning. But there are also examples of larger units without any meaning. An expression such as *How do you do* in British English can hardly be said to have a meaning, though obviously it has an important pragmatic function.

We will now consider several other conceptual categories and properties and their correlates in the grammatical module, with the aim of providing linguistic empirical evidence in favor of our main thesis, thus giving our main thesis at least some plausibility.

We start with a table presented by [Pinker 1984], which provides a pretty good correspondence between semantic notions and their grammatical correlates. Pinker presented this table to support his hypothesis of Semantic Bootstrapping in language acquisition by children.

Grammatical Element	Semantic Inductive basis
Syntactic Categories	
Noun	Name of person or thing
Verb	Action or change of state
Adjective	Attribute
Preposition	spatial relation, path or direction
Sentence	main proposition
Grammatical Functions	
Subject	agent of action; cause of causal event; subject of an attribution of location, state or circumstance; argument with autonomous reference
Object or Object2	Patient or Theme
Oblique	Source, Goal, location, instrument
Complement	proposition serving as an argument within another proposition
Topic	Discourse topic distinct from the arguments of the main predicate
Focus	Discourse Focus
Cases	
Nominative or Ergative	Agent of Transitive action
Accusative or Absolutive	Patient or Theme of transitive action
Nominative or Absolutive	Actor of intransitive action
Dative	Goal or Beneficiary
Instrumental	Instrument
etc.	
Grammatical Features	
Tense	Relative Times of event, speech act and reference point
Aspect	Durativity
Number	Number
Human	Humanness
Animate	Animacy
Tree Configurations	
Sister of X	Argument of X
Sister of X' (Aunt of X)	Restrictive modifier of X
Sister of X'' (Great-aunt of X)	Nonrestrictive modifier of X

Table 1. Pinker's table of correspondence for semantic bootstrapping

What concerns me here is not the semantic bootstrapping hypothesis, and not even whether the exact mapping that Pinker proposes is the right one (one may debate many of his choices), but what is significant is that such a table with a very plausible (almost perfectly) one-to-one correspondence between semantic notions and grammatical correlates can be made at all.
There is no such a one-to-one correspondence just because grammatical properties must have a meaning. Many of the grammatical properties mentioned by Pinker do not have any meaning at all (e.g. noun, subject, nominative, etc.). And even if some grammatical property does have meaning, there is no reason why there should be exactly one correlate in the grammar for such a meaning. This can be illustrated for many languages, but German provides a particularly nice example. Note that German has a grammatical property *plural*, which corresponds to a concept property PLURAL (see below for more details). So there is exactly one correlate in the grammar for the conceptual property PLURAL. But now let us look at the PF-component, how this grammatical property is concretely realized. My hypothesis does not say anything about the relation between the grammatical property and its concrete realization as sequences of phones. In fact, what one finds here is highly variable per language. German represents a pretty complex situation for the realization of plural, with 5 different suffixes, 3 of which can occur with or without Umlaut, so in total 8 different ways of realizing plural (and we ignore here cases such as Latin-inspired plurals)..

The situation can also be reverse, i.e. many different distinctions in the grammatical module, and only few in concrete morpho-phonology. Again, there is high variability per language, and again a German example illustrates an extreme. Adjectives in German, following the analysis in traditional grammars of German [Duden 1998: 281-3] have among their grammatical properties *case* (4 values), *number* (2 values), *gender* (3 values), *strength* (3 values). The property *gender* is neutralized when *number=plural*, so looking at number and gender alone leads to 4 different combination Combined with the other two properties, we get to 4 * 4 * 3 = 48 different combinations. However, if we look at the actual forms, there are only 5 suffixes (*er, e, en, es, em)* to realize these 48 combinations. So again we see a big mismatch, in contrast to what we see in the relation between the C-I and the grammatical module where there is always exactly one correlate in the grammatical module for each property in the C-I module.[34]

So let me make very clear what kind of evidence we need in favor of my hypothesis. It is not enough to show that there is some correspondence between conceptual categories and syntactic categories. The correspondence must be more precise: for each conceptual

[34] Of course, one might wonder why so many (48) distinctions must be made in German for adjectives if the concrete forms provide only evidence for 5 forms. The reason is that the concrete forms are not the only source of evidence for such distinctions. The grammatical system has internal rules and principles as well, which may give rise to more distinctions than one sees at the surface. Thus, in many theories it is assumed that noun phrases have a property (*abstract*) *case*, even in languages that do not show any overt (*concrete*) case distinctions. In the German example discussed, evidence for the distinctions in adjectives also derive from rules of agreement and government that state that nouns and (attributive) adjectives must agree in certain features, and that adjectives take on different forms depending on the governing determiner.

category we must find a corresponding grammatical category provided that the concrete forms combined with grammar-internal principles or rules (e.g. of agreement and government) supply sufficient evidence to make such a distinction. Furthermore, it is not sufficient to only show such a close correspondence: in fact it must be shown that each conceptual category postulated can be justified independently of language (e.g. by showing that nonlinguistic entities such as babies and our closest relatives the primates have this distinction; or by showing that the conceptual category is needed to explain phenomena from nonlinguistic input to humans (vision, touch, hearing of sounds (not speech), taste, etc.), or at least (though this is much weaker evidence) evidence that the conceptual category exists independently of the correlating grammatical category.

We have indications for a very close correspondence between plausible conceptual categories and grammatical categories, as we have seen above. And we have some evidence that some conceptual categories are independent of language, in particular for the categories NUMBER, DEIXIS, and the INDIVIDUAL/CLASS distinction for objects, as we will show in the next section. The close correspondence found and some initial evidence for its basis outside of language contributes to the plausibility of the thesis and invites one to investigate language-independent evidence for the other postulated conceptual categories. And that is exactly why a research programme in which these questions are actually investigated in detail makes sense.

5 Grammatical Properties

We will now discuss several individual properties in more detail.

Let's start with the NUMBER concept as a concrete example It has been argued by [Dehaene 1999:24-25; 66; see also Pinker 2007:138-141; Fitch 2010:152] that primates and children before they attain language have a concept of NUMBER.[35] More specifically, this concept of NUMBER is very precise for small numbers (1 – 3) ('small-exact', [Fitch 201:152]) and approximate for larger numbers ('large – approximate') Significantly, the grammatical category *number,* in all languages that distinguish it, shows exactly the same distinctions: language distinguishes singular-dual-trial (corresponding to 1-2-3), and paucal-greater paucal-plural (corresponding to FEW-INTERMEDIATE – MANY) [Corbett 2000]. Of course each individual language only distinguishes a subset of these categories, or does not have the category number at all. The grammatical system makes such distinctions only when there is input that requires making these distinctions. Thus, Japanese forms[36] do not provide any evidence for the grammatical category *number*, many languages have just a bipartite distinction between *singular* and *plural*, some languages in addition also distinguish a *dual*. The category *trial* is reported to occur. A category *quadral* has not been observed in any language

[35] Sometimes called NUMEROSITY. See [Pinker 2007:141]

[36] Japanese does have a suffix *tachi* that can occasionally be used to express plurality, but it is only compatible with animate nouns, is not used systematically and no grammatical rule (e.g. agreement rules) need refer to it. I assume that it is a normal derivational suffix.

(alleged cases are properly analyzed as paucal or greater paucal [Corbett 2000]). A distinction between *paucal* vs. *plural* has been claimed to exist in a number of languages including Hopi, Warlpiri, Arabic, and some Slavic languages. Obviously, the study of the grammatical category number is not finished and there are several controversies about the proper analysis, but the description here (based on [Corbett 2010] appears to be a good approximation, and shows a correspondence with the conceptual distinctions that is in my view too striking to be accidental.[37]

Note that there is no a priori reason at all why the grammatical category *number* should be as indicated. We can contrast it with the number words. In number words, most are created by perfectly general rules, but there is some idiosyncrasy, e.g. in English there are idiosyncratic words for 1..13, 15, 20..50, 100, 1000, 1000000. English could have expressed the grammatical category *number* in such a way, with special categories for the ones for which there are idiosyncratic names, e.g. with a suffix identical to the first syllable of the number word (e.g. *bookthir* would mean 'thirteen books', *bookeight* would mean 'eight books', etc.). But it doesn't. And no language uses such a system, and, I claim, for principled reasons, since the human C-I system, and therefore the grammatical system, is not structured in this way.

Of course, the human C-I system can form concepts corresponding to numbers, but whether it does and to which extent is not innately predetermined, but determined by cultural environment and/or nonlinguistic human intelligence. Similarly, in the grammar module the existence of words for specific numbers, and how many of them are distinguished differs from language to language (and indeed, many languages from technologically less-developed cultures are reported to have only a few words for numbers). The basic distinctions between 'small exact' and 'large approximate', however, are determined by the structure of the conceptual module, hence also of the grammatical module, and these distinctions (or a subset of these distinctions) will be found in any language that has the grammatical category number at all.
One might suggest that the distinctions in the C-I module are in fact caused by the grammatical system, so that it is not the case that the grammatical distinctions originate from distinctions made in the C-I module, but rather the other way around. However,

---

[37] In fact, the system is probably more complex. For example, it is likely that the different categories are interrelated somehow., perhaps in a hierarchy. Absence of the feature corresponds to no number distinctions. The category *number first branches into singular* and a more abstract category *2+*, which then splits into *dual* and *3+*, etc. Such a structure would be able account for the fact that e.g. *dual* only occurs when *plural* is also distinguished (cf. [Greenberg 1963]'s Universal 34 ("No language has a trial number unless it has a dual. No language has a dual unless it has a plural"), and can account for the fact that the semantics of *plural* differs slightly between languages with a *dual* (>=3) and languages without a dual (>=2). In fact, it is possible that for each grammatical property a range of possible values is defined and options on how to subdivide these values, derived from its conceptual correlate. How the actual range of values is actually subdivided in syntax then depends on input from concrete morphology. This will make it possible for grammatical features to be ambiguous or vague with regard to certain conceptual categories despite being modeled after them.

this cannot be true since crucially the tests referred to above involve language-less beings such as animals and children before they attain language.[Dehaene 1997:66; Pinker 2007:138; Fitch 2010:152]

It is also important to realize that though the grammatical category number is "inspired", so to say, by the conceptual notion of NUMBER, and though it often has a meaning that directly corresponds to the semantic distinctions made for the conceptual category NUMBER, the grammatical category number (1) is not identical to the corresponding conceptual category; (2) is often used grammatically without any reference to any meaning (e.g. in verbal and adjectival forms when it arises from agreement processes); and (3) even on nouns or pronouns does not always have the expected meaning, e.g. the French pronoun *vous* and the Russian pronoun *vy* (both 'you', polite form,) are grammatically plural but can be used to address a single person (showing reverence to him/her). The same is true for German *Sie*. The independence of the grammatical and conceptual categories is also clear from the fact that in Russian the numeral for 1 (*odin*) can appear in grammatical plural (with inherently plural nouns such as *chasy* 'watch' (lit. 'hours'), or *ochki* 'glasses'). But even for nouns, it is dubious to assign the grammatical category a directly corresponding meaning. Phenomena with fractional numerals are very interesting in this context. The fraction 2,5 (using the Dutch notational conventions) can be expressed in Dutch in multiple ways: (1) by adding *eneenhalf* lit. 'and a half' to the word for 2 (*tweeëneenhalf*); (2) by reading the digits and symbols more or less literally (cf. English *two point five*); in Dutch this is then *twee komma vijf* lit. 'two comma five'; and (3) by pronouncing 2 ½, in Dutch *twee een tweede* lit. '*two one second*'. Now look at what happens to nouns these numerals modify:

(3) Expressing *2.5 days* in Dutch

a.	Tweeëneenhalve	dag / *dagen	'two and a half	day / *days'
b.	Twee komma vijf	*dag / dagen	'two comma five	*day / days'
c.	Twee een tweede	dag / *dagen	'two one second	day / *days'

So in two cases, the grammatical category *singular* must be used, and in the remaining case the grammatical category *plural* is required. This clearly suggests that the number category on the noun is used here as a purely grammatical category, and the relation with the concept category NUMBER is at least not as direct as one might expect.
Despite these mismatches, in most cases the grammatical categories singular and plural do correspond to the conceptual categories SINGULAR and PLURAL. Though it is completely independent of the thesis being developed in this paper, there may even be a default way of expressing a conceptual category in the grammatical system. Such a default mapping is of course very useful for learning the language (see Pinker's Bootstrapping Hypothesis), but the default mapping can be overruled by actual form input. Deviations from the default mapping form a large part of the idiosyncratic aspects of a language.

DEIXIS is a second example. Pointing to an object with one's arm or with one's gaze is recognized as a way to focus attention to this object by some animals (dogs, to some degree) but not by others (e.g. cats). [Hurford 2003:47] states that 'primates more closely related to humans are better at following the human gaze than those less closely related' (basing himself on [Itakura 1996]).[38] Furthermore, according to him spontaneous pointing has been observed in captive common chimpanzees (without any language training) and in free-ranging orangutans.[39] Small children understand and use pointing before their first birthday and use an elaborate system at about the age of 1 ([Bates 1976] cited in [Lieberman 1984:87]), long before they show any signs of understanding or producing language. We may thus safely assume that DEIXIS is a conceptual category independent of language. Our hypothesis predicts that language will show deixis as a grammatical property, and indeed deictic pronouns exist in every language.

[Bickerton 2009:80], in discussing attempts to teach languages to chimpanzees and bonobos, observes, that, in all these attempts bonobos and chimpanzees apparently were able to make the distinction between naming individual objects (usually animate beings) and naming classes of objects. He states "apparently", because this ability is never discussed anywhere in the literature on teaching apes languages. It is apparently taken for granted, though the distinction is hardly trivial. [Jackendoff 2002:239] reports on research by [Macnamara 1982], which shows that very young children also appreciate this distinction. The grammatical distinction between *common nouns* and *proper nouns* is a direct correlate to this distinction, so if Bickerton, Macnamara and Jackendoff are right, we have evidence for a conceptual correlate to the distinction between common nouns and proper nouns in non-humans and pre-linguistic humans.

The grammatical category *tense* might be another grammatical category, for which there is some evidence that conceptual correlates are non-linguistic. [Fitch 2010:151] discusses data that 'support the idea that other vertebrates have memory capacities, allowing "mental time travel" like our own', but the details of the conceptual subcategories of TIME are not known, so that no convincing support for a correlate between conceptual categories in the area of TIME and grammatical categories related to *tense* can be provided in the current state of understanding.

For some examples of grammatical categories we have weaker evidence for their basis in conceptual categories. For these we have evidence for the existence of the postulated conceptual categories independent of their grammatical correlates (but not independent of language as a whole).

A case in point is the grammatical category *person*, which distinguishes 3 values: *first*, *second* and *third* person. These correspond to the conceptual categories SPEAKER, ADDRESSEE, and OTHER THAN SPEAKER or ADDRESSEE. Evidence that the

---

[38] [Fitch 2010:159-160] notes that dogs are much better at 'gaze following' than chimpanzees or wolves. He suggests that 'gaze following' might be a capacity that all have but that in most species 'gaze prohibition' (the unwillingness to look at another's eyes) might prohibit full development of the capacity

[39] Though it is much less developed than in young (human) children, according to [Jackendoff 2002:242], citing [Povinelli *et al.* 2000].

conceptual categories exist independent of the correlating grammatical categories can be derived, inter alia, from languages that manifest the conceptual category PERSON even if they do not distinguish the grammatical category person. For example, Japanese, as many languages, distinguishes a regular series of demonstrative pronouns in three variants (near the SPEAKER), (near the ADDRESSEE), near neither SPEAKER nor ADDRESSEE):

(4) Japanese Demonstrative Pronouns
  d. Kore sore are 'this one, that one, yonder one'
  e. Kono sono ano 'this that yonder'
  f. Koko soko ako 'here there yonder'

even though there is not the slightest evidence for the *grammatical* category of *person* in Japanese.[40] Again, some languages do not distinguish *person* as a grammatical category at all (e.g. Japanese), but the ones that do always show the tripartite distinction as indicated. The grammatical and conceptual categories must be clearly distinguished from each other, there is no perfect correspondence between the grammatical values and the corresponding conceptual categories (e.g. German *Sie* and Spanish *Usted* are grammatically *third person* but are used to address the ADDRESSEE; Dutch *u* fluctuates between second and third person).

A further example is *aspect*. This grammatical category is clearly present in e.g. the Slavic languages. For a language such as Dutch no such grammatical category can be justified. However, the conceptual correlate ASPECT does play a role in Dutch sentences (see e.g. [Verkuyl 1972]), thus providing evidence for the existence of the conceptual category independent of the grammatical category.

There are many more parallels, but I currently do not have strong evidence for the independence of the postulated conceptual categories of language, and this is one of the types of evidence that should be searched for in the research programme being proposed here. For example, many linguist have felt an intuition that the distinction between parts-of-speech such nouns and verbs must be grounded in semantic types such as (class of) OBJECT and ACTION. Many linguists even go so far as to define nouns and verbs in this way (which is completely wrong in my view)! But the intuition felt naturally falls into place in the hypothesis I develop here: noun is the grammatical category that is the correlate of the conceptual category OBJECT CLASS, and verb is the correlate of the conceptual category ACTION CLASS. Again, there is limited variation here. For example, some languages make a systematic grammatical distinction between *active verbs* and *stative verbs* (as justified by different conjugational patterns, different combinatory properties, etc).[41]

---

[40] Japanese verbs do not inflect for person; reflexives and reciprocals do not show person distinctions; and arguably Japanese does not even have personal pronouns, and even if it has them, they do not distinguish the grammatical category *person*

[41] E.g. Lakhota http://en.wikipedia.org/wiki/Lakota_language

Similar correlates can be found for adjectives (PROPERTY), and adposition (LOCATIVE RELATION).

Of course, also here the grammatical categories develop independently and autonomously, so there are many mismatches between the grammatical categories and the conceptual categories. Nouns e.g. can denote virtually any concept, including (classes of) OBJECTS, PERSONS, ANIMALS, but also EVENTS, ACTIONS, STATES, PROPERTIES, SPACES, PERIODS, and many others.

As we stated before, the C-I system has some combinatorial potential, and it also has restrictions on how concepts can be combined. These are generally known as semantic selection restrictions, and such restrictions specify that e.g. a property such as COLOR cannot be applied to abstract concepts; that an ACTION such as expressed by the verb *to say* needs as an agent a HUMAN, etc. etc. In the grammar, there is a perfect correlate, syntactic selection restrictions, there defined as restrictions on combining certain grammatical categories such as syntactic categories (PoS), case features, etc.

We see a direct correlate of an element from the conceptual system in an element of the grammatical module. It is interesting to see here that the element in the conceptual system is not a concept itself; it is rather a restriction on the combination of concepts. Again here, we assume there is a default relation between semantic selection restrictions and syntactic selection restrictions, which here must be defined by decomposing the semantic selection restrictions into its composing parts (restrictions on arguments expressed in the form of semantic types), mapped to their grammatical correlates, after which these are composed into a syntactic selection restriction.

It is occasionally claimed that syntactic selection does not exist, and can be derived in full from semantic selection restrictions. I believe that there is plenty of evidence against this hypothesis in general and also in specific cases (cf. e.g. [Odijk 1997]), and to my knowledge, no theory that really has investigated the matter adopts this hypothesis.

There are mismatches here between the conceptual system and the grammatical system, at least according to some theories. It is sometimes claimed that verbs do not syntactically select for subjects, even though many verbs do impose semantic restrictions on the argument expressed as the subject. This is controversial, and not accepted in all theories, and if it is real, it should be explained from independent factors, since the theory exposed here predicts (at least by default) a systematic correlation. [Carstairs-McCarthy 1999] indeed asked the question why so many languages show a major syntactic split between subject and predicate (VP) constituents, and suggests that it might be related to the asymmetry that also exists in the syllable (with the structure [C[V C]] rather than [C

V C]). This is an interesting option, which would fit in well with what I later propose about structure building (see section 7).[42]

The C-I module must have ways of combining arguments with predicates. Often, arguments can be combined with predicates in multiple ways, or the C-I module must have ways to systematically derive new concepts with different combinatorial potential. The correlate of this in the grammatical system is *voice*, and *verb alternation* (e.g. *middle*, *inchoative* and *causative* variants, *intransitive* and *transitive* variants, etc. See [Levin 1995]). Various forms of *reflexivisation* can perhaps also be considered in this way if the analysis by [Reinhart & Reuland 1993] which postulates reflexive marking of verbs even in languages that do not overtly mark this is correct.

Perhaps the differences between nominative-accusative case systems and absolutive-ergative systems, if they are not just PF-differences in concrete case realization, can also be seen as correlates of different ways of combining arguments with predicates. However, this requires a deeper analysis of these phenomena than I can provide here. See [Marantz 1981] for some relevant discussion.

For several grammatical properties, it is not so evident that they can be considered a correlate of a category from the C-I module. It either requires additional clarification, or is truly problematic.
For example, in the C-I module we have SEMANTIC ROLEs, also called THEMATIC ROLEs, such as AGENT, PATIENT, SOURCE etc. Grammatical *case* (*nominative*, *accusative*, etc.) might be considered a plausible correlate of these semantic roles (which sometimes are even called (deep) cases [Fillmore 1968]). However, case is limited to nominal categories (semantic roles are not), and there is an alternative correlate: grammatical relations such as subject, object, etc. Some theories do not recognize grammatical relations as primitive notions in the grammatical module but replace them by structural configurations. Anyway, grammatical relations or structural configurations are much more plausible correlates to semantic roles, which leaves it unclear what the grammatical category *case* is a correlate of. [Pesetsky 2010] argues that *case* is just an affixal variant of the major syntactic categories for words (N,V, A, P), which, if correct, would solve the problem of the correlate of *case*,

The grammatical module for human language allows phrases to occur in other positions than one would expect given the fact that predicate-argument relations must be determined in local configurations. The prime example is unbounded dependencies, analyzed in many theories as created by a movement operation, or by linking a filler using gap-threading to an abstract element that is local to the predicate that the filler is an argument of. One may wonder what element in the C-I module this grammatical phenomenon (however one analyzes it) is a correlate of.

[42] [Jackendoff 2002:353 fn 11] discusses this and suggests that the organization of Topic and Comment in information structure might be alternative source. However, in my view, all three phenomena might arise from a common cause (see section 7).

Chomsky has argued that this property should be analyzed as a special case of *Merge*. *Merge* is the major operation in syntax to combine syntactic objects into new syntactic objects. Movement is analyzed by Chomsky as involving Merge of a syntactic object SO1 with a syntactic object SO2 where SO1 happens to be contained in SO2 (it is therefore called *Internal Merge*). One might argue that this phenomenon therefore does not need any explanation over the explanation for Merge itself. However, I am not sure that the reduction of movement to Internal Merge is complete. I will come back on this below in section 7.

Quantifiers binding variables in the conceptual system form a more plausible candidate for providing the source for unbounded dependencies. After all in such conceptual representations one has to make a link between a scope marker and a variable that can be far away from each other. As soon as some mechanism is in place in the grammatical module, it can be used not only for unbounded dependencies between a quantifier and variable (as e.g. in wh-movement), but also for different types of dependencies (e.g. A-movement/A-binding), due to the autonomy of the grammatical module. Of course, this presupposes that the human conceptual system uses variables in its representations, but that is exactly what has been suggested by [Chomsky 2005:10 fn 29].

6 Open Issues

There are several properties of the grammatical module for which I cannot at this point provide a plausible element in the conceptual module that they are a correlate of.
Let us start with the most important ones. The grammatical module must have rules and principles for accounting for agreement and structural case assignment. In some analyses, structural case assignment is a special case of agreement, which would reduce the problem slightly. Maybe, but this is highly speculative, agreement is a correlate of multiple variables occurring with multiple predicates in the C-I –module. Agreement is typically in terms of so-called φ-features (such as person, number, gender, etc) and case, and complexes of these features are assumed to correlate to variables in the C-I module [Reuland 2011:35]. A form such as (5) from Russian thus almost perfectly corresponds to a first order logic representation of its semantics such as *interesting'(x) & book'(x)* where the variables in the logical structure correspond to the φ-features (same features, same variable):

(5) interesnuju knigu
interesting+3-f-sg-acc book+3-f-sg-acc

A second problem is formed by such categories as *word* v. *clitic* v. *affix*, Of course, in many cases such categories can be kept out of the grammatical module and be attributed to the PF-module, where they naturally belong. But the problem is that esp. *clitics* in many cases do not only have properties that make them different from words with regard to PF-phenomena, but that they have special syntactic properties as well (e.g. a different distribution that cannot be accounted for by PF-factors alone).

Bound anaphors –to the extent that they cannot be analyzed as instances of voice - might be a correlate of bound variables in the C-I module, but what are bound pronouns then a correlate of, and why is there a distinction between anaphors and pronouns to begin with? I have no answer, but [Reuland 2011] makes an attempt to explain the difference between pronominals and anaphors from a combination of linguistic and non-linguistic factors, including φ-features and agreement.

The operations on grammatical structures are all subject to locality constraints (subjacency, ECP effects, local domains for binding, relativized minimality etc.). The locality constraints differ a bit from phenomenon to phenomenon. These locality constraints cannot be accounted for on the basis of corresponding constraints in the C-I module. I will suggest a different route to an explanation in section 7, though it is certainly not clear that that route will lead to an explanation.

7 Recursion

A final property of the grammatical module that does not follow from what I said so far is the central property [FHC 2002] ascribe to FLN, viz. recursion. Recursion does not follow from the CCH, and an independent explanation is required. This is certainly not a weakness of my theory, since as far as I can see nobody has an account of where recursion came from and everybody must stipulate that it arose at a certain point in the development of modern humans. [FHC 2002] also just stipulate that it arose at a certain point. I would like to sketch one option for accounting for the rise of recursion that could be pursued further, though again, it is highly speculative and tentative.

Let us first observe that Chomsky's point of view of recursion as belonging to FLN is somewhat confusing. After all, recursion does not only occur in language but also in the counting faculty, and in music and dance. This is acknowledged by [Chomsky 2010:53]: "one may raise the factual question of whether the basic properties of language, notably recursive generation, are unique to the language faculty or are found elsewhere. […] We know that it is not". [Chomsky 2004] (see also [Chomsky 2010:53]) states that "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." He goes on to suggest that "one possibility is [that] the natural numbers result from a simple constraint on the language faculty", which would make recursion in arithmetic "parasitic on the language faculty". [Jackendoff & Pinker 2005 217-218] suggest that recursion also plays a role in certain types of visual patterns. [Chomsky 2010:53] discusses recursion in visual patterns, apparently countering the critique by [Jackendoff & Pinker 2005], however, without mentioning them. He states that recursive operations applied to visual arrays will result in a discrete infinity of visual patterns, but he also claims that "this is simply a special case of arithmetic" concludes that it "tells us nothing new about recursion beyond language"

Recursion also occurs in man-made artificial languages such as programming languages, which Chomsky considers as clearly derivative on recursion in natural language. So how specific to natural language is recursion in fact? Chomsky seems to suggest that recursion originates in language (actually even more narrowly, in the narrow syntax module), and that all other uses are derivative. But how can a part of an autonomous and otherwise encapsulated module such as the narrow syntax module suddenly become available to completely other modules of the brain? Chomsky notes this problem and related ones ("dissociation with lesions and diversity of localization"), but he qualifies them as "apparent" and goes on to state that "the significance of such phenomena is unclear for many reasons (including the issue of possession vs. use of the capacity)." And he concludes that "there may be something to these speculations, perhaps along the lines just indicated". But I am not so sure that the problems with this point of view can be dismissed so easily.

Second, though the recursive property of grammatical systems is real, I do not believe it is a very essential aspect. If we would have had natural language that restricts sentences to a finite length (e.g. maximally 10 or 15 words; or a limited degree of recursion (e.g. a maximum recursion depth of 2), language could still have all the roles it has and has had in the past. In fact, one might even argue that we would be better off then, since overly complex and stylistically infelicitous sentences will then occur much less often, facilitating the clear formulation and expression of one's thoughts and improving communication via natural language.

The domain of numbers and counting is probably the most basic recursive system in human cognition. It has been defined or formulated in various ways, always in recursive terms, e.g. as recursive application of the successor operation (*add 1*), recursive application of set inclusion [Whitehead & Russell 1972], or recursive application of *Merge* to a single argument [Chomsky 2004, see above and Chomsky 2005:6]. But to my knowledge, almost all or even all natural languages use a **finite** grammatical subsystem without recursion to form expressions for numbers (and of course, they therefore can cover only a finite number of numbers). Even many artificial notational systems for numbers (e.g. the Roman number system) define only a finite number of expressions. It is only with the invention of the artificial language based on the place-value system in which numbers are written as a sequence of digit symbols using the position of each digit as a crucial factor in determining its value and crucially using the symbol *0,* that a recursive system arose that allows one to notate any number. And with the invention of the so-called 'scientific notation', very large numbers could also be notated in a relatively compact manner. Natural language itself did not come up with a recursive solution, but the finite system generally functioned perfectly well for all practical purposes, though occasionally new words had to be invented for higher numbers not covered yet when the cultural, technical and economic developments in a society required so.

In addition, it has been claimed [Everett 2005] that some natural languages actually are finite (Pirahã), which, if correct, shows that recursion is not a central and necessary property of language. Everett's claims have risen to some controversy, and many of his factual claims and especially his explanation for them have been (correctly, in my view) rejected, in particular (but not exclusively) by [Nevin et al. 2009]. However, as far as I can see, his claim that Pirahã does not have indefinite embedding still stands, (though it does have embedding of a depth of one level, contrary to Everett's claims).

Recursion is in fact an emergent property that arises in certain types of combinatorial systems if certain additional conditions are met. And, as I will argue below, the combinatorial system, more specifically *structure building* (with or without recursion) is a much more central notion for (multiple modules within) the language faculty (and other cognitive systems). Recursion is just a special side effect that arises under certain circumstances in combinatorial systems. So, the emphasis that e.g. FCH put on recursion[43] is, in my view, misleading because it focuses on a less central aspect of language.

If the 'certain circumstances' that I mention do not hold in some construction, there will be no or only limited recursion. [Evans and Levinson 2009:442-443] cite the example of *Kayardild,* a language in which due to a constraint on doubling case and the ways subordinate clauses must be formed, recursion depth is limited to 1 in certain constructions.[44]

As stated above, the property that is – in my view – much more central is the possibility to build structure. Chomsky uses the operation *Merge* as the central operation of this combinatorial system. I believe that Chomsky's description of *Merge* as the core of the combinatorial system is not incorrect but may again be somewhat misleading, for two reasons. First, Chomsky restricts the application of Merge to (narrow) syntax. It is possible that this is just an artifact of Chomsky working on syntax only, but if it has more principled grounds, I submit this to be incorrect, Second, [Chomsky 1999] describes *Merge* as an operation that takes two syntactic objects SO1 and SO2 as input and yields a new syntactic object SO3={SO1,SO2}. This is misleading because no two syntactic objects can be combined into a new syntactic object by the operation *Merge* alone! And correctly so, since just combining two arbitrary syntactic objects (let's just take single words for simplicity) usually does not lead to a well-formed syntactic object (cf. **the the*, **John John*, **annoy very*, etc.). A *Merge* operation can only lead to a new syntactic object in combination with other principles and properties of the input syntactic objects, in particular the combinatorial potential of the input syntactic objects. For example, properties of syntactic selection of complements (e.g. *destroy* must take a NP complement, *eat* optionally takes an NP-complement, etc), or modification potential (e.g.

---

[43] For example, [HCF 2020:1573] state "FLN comprises only the core computational mechanisms of **recursion** as they appear in narrow syntax"; [FCH 2005:182] state "a significant piece of the linguistic machinery entails **recursive operations** "

[44] But of course, this does not mean that Kayardild uses completely different grammatical formalisms in comparison to other languages that show unlimited recursion depth in comparable constructions, as [Evans & Levinson 2009:443] appear to suggest.

very can modify adjectival phrases but not verbs), etc. I believe that Chomsky agrees with this, but this aspect is almost never made explicit by Chomsky.[45,46]

I claim that structure building plays a role in multiple independent modules of the language faculty (not only in syntax but also in morphology, phonology, and orthography). It also plays a role to account for the human ability to count (as also suggested by Chomsky), as well as in music [Lerdahl & Jackendoff 1983][47], dance and vision. I assume that for all these aspects there are autonomous modules. Structure building and recursion also play a role in processing artificial languages used in logic, mathematics and in programming.[48] Chomsky claims that all manifestations of recursion are derivative of language, and *Merge* is a genetically determined property of language and unique to it [Chomsky 2010:53]. However, if structure building plays a role in all of these cognitive capacities, structure building cannot be a property of any of the individual modules (hence it cannot be specific to language, or (narrow) syntax, contra [HCF 2002], [Chomsky 2004], [Chomsky 2010]). In fact, if structure building plays a role in all these domains, it must be an *independent module*. I propose that the capacity to build structure is part of an independent module specific to humans but not to language (or any of the other modules or faculties mentioned above). I will argue below that structure building plays a crucial role in alleviating the restrictions that the short term memory module imposes on humans somewhat, thus enabling them to process more information in a faster way than other primates. Structure building is part of an independent module, it takes two objects and a combinatorial instruction from a single other module, and yields a new structure containing the two objects.

[45] Or maybe he believes that such properties can be derived from other (non-syntactic) properties, which would explain their absence in syntax, but again, to my knowledge, he is not explicit about that. Other researchers (e.g. [Stabler 1997] in the Minimalist framework have made this aspect explicit, e.g. by using features to represent such combinatorial potential, and in other grammatical frameworks (Categorial Grammar, HPSG) it deservedly forms a crucial and central aspect.

[46] [Chomsky 2005:6] does mention that 'for an LI to enter into a computation,[…], it must have some property permitting this operation. […] an LI has a feature that permits it to be merged'. In my view, the relation between Merge and the feature is less direct: structure building takes two objects and a combinatorial instruction or principle, and this combinatorial instruction may be formulated in terms of a feature on LIs. Chomsky appears to leave the role of the combinatorial instruction implicit.

[47] [Chomsky 2011] appears to accept this for music, though he believes this is a derivative of language. [Fitch 2010:121] explicitly states that music 'has a form of syntax (a set of rules for combining these [notes, JO] into larger hierarchical structures of essentially unbounded complexity'

[48] In these cases one must clearly distinguish between our scientific understanding of the concept of 'recursion' (which is irrelevant here), and the unconscious capacity of humans to produce and analyze an indefinite number of expressions from such artificial languages. A programmer can perfectly use a programming language that allows for an unbounded number of expressions even if he/she does not master the scientific concept of 'recursion'. The latter **is** needed, of course, for designers of such languages, for builders of compilers and interpreters for such languages, and for programmers who want to program recursive functions, but that is irrelevant in this context.

7.1.1 Short Term Memory

Let me elaborate on the role of structure building in the Short Term Memory module.

Consider the following example:

(6) qplswkrgcdxvbtrs

Suppose that a person is presented this example for a short period (let's say .5 seconds), and then it is hidden again. The task for the person is to reproduce the example. This is an extremely simple task. Any computer can do this. I can teach anyone who has never written a computer programme before in a few minutes how this can be programmed. However, humans, the most intelligent beings on earth, fail on this task!

An even more interesting example is the following

(7) aigmlsnoa

Under the same task, again humans score pretty badly. However, the amazing thing is that if we present an example containing *exactly the same characters*, but in a different order (as in (8)), humans score much better if not perfectly:

(8) mangaliso

How can this be? It cannot be that the test person "knows"[49] the string "mangaliso", since it is not a word of English.[50] So the reason for the significant difference in performance of humans when presented with (7) v. (8) must lie elsewhere.

The explanation I propose for this is as follows. First, humans score badly on (6) because the string has to go through a module of the human brain that is called the *Short Term Memory* (STM). The STM can contain only a very limited number of objects [Miller 1956].The string in (4) contains 19 characters, and these simply do not fit all in the STM. One way to correctly repeat this string is then to make sure the string ends up in *Long Term Memory*, but the duration of the presentation of the string is too short to get the string into LTM. In fact, getting a list of arbitrary items into LTM requires a lot of effort and is pretty difficult for humans. So this is no option either. Hence, humans cannot reproduce this example.[51]

---

[49] And even if this is the case, it would still require an explanation why "knowing" the word helps in processing a sequence of characters.

[50] The word does occur as a name in certain areas in South Africa, and it may have a meaning in isiZulu.

[51] This proposed explanation of course does not originate from me. Many others have proposed it, starting with [Miller 1956]. See [Fitch 2010:99-100] for a recent discussion in the context of phonology. Here Fitch claims that structure building in phonology does not show self-embedding, in contrast to syntax. This may be true, though one might argue that a morpheme can consist of an indefinite number of syllables, only limited by memory constraints, which would require iteration (weakly equivalent to a trivial

If this is correct, why doesn't the same story hold for example (8)? The reason is, in my view, that people who speak a language and can read have internalized rules on how characters can be combined in their language (*orthotactic rules*). And when a sequence of characters satisfies these rules, the characters can be combined in a single structure, e.g. for a syllable. After that, only the reference to the syllable needs to be stored in STM. For an English speaker (and for speakers of many other languages), the example in (8) can be analyzed as a sequence of 4 syllables (*man*, *ga*, *li*, *so*). The structure building operation can take a subsequence of the example and orthotactic rules (e.g. describing the structure of syllables) as input, and if the rule description is met, it builds a single structure for the whole subsequence. In this way, the STM needs to contain only 4 (references to) syllables (which fits easily in STM) instead of 9 (references to) characters (which is pretty hard). Of course, people can still reproduce all characters in this task, so the structures for the syllables should somehow also store all individual characters. And they do, but the individual characters in the structure representing the syllable are not stored in STM, but in a different memory that I will call Working Memory (WM), which is accessible only via elements in STM.

Methods for alleviating the limits of STM of this kind were already described by [Miller 1956], who talked about it as organizing the input sequence into *chunks*: *grouping input events and apply a name to the group, and then remember the new name rather than the original input events* [Miller 1956:11]. Structure building is a generic method to achieve this effect.

Of course, the examples so far involve characters, and reading, and one might argue that that is not real language but just a derivative of language. Surely true, but the same principles apply in the case of phonology (and in fact the rules and processes of (alphabetic) orthography are to a large extent modeled after phonology). So in phonology sequences of phonemes that satisfy *phonotactic rules* can be combined in a single structure for a syllable, exactly in the same way as in the case of orthography.[52,53]

Independently, a metrical structure is built up and imposed on input strings. In fact, building up a metrical structure can be used for many types of objects, and imposing

---

form of (right-linear) recursion, see [Chomsky 2005:5, fn 12]). Long words from a foreign language can be processed, though with some difficulty, witness example (9) for people who do not speak German or (partially artificially) long names of Welsh villages. I do not believe that structure building in syntax is qualitatively different: (as Fitch appears to suggest); only, in syntax the output of structure building often happens to be able to form input to it again, which results in recursion.

[52] Of course, in many cases the input is not discrete but continuous, as in handwriting and speech. But I assume (as does everybody, as far as I know) that independent modules map these continuous signals into sequences of discrete (orthographic or phonetic) symbols before they enter STM.
[53] I assume something similar will hold for gestures, but this will have to be confirmed by experts.

metrical structure is often used as a means to remember otherwise arbitrary combinations of symbols (such as telephone and credit card numbers, pin codes, etc.)

Next, also in morphology the same principles apply. A native speaker of German can analyze a productively formed compound such as (9):

(9) Warentrennstabregalboden

not only as consisting of 8 syllables, but also as consisting of 5 morphemes (*Waren, trenn, stab, regal, Boden*), or actually as a combination of two words, (*Warentrennstabregal* and *Boden*), the first one of which itself consists of two words (and so forth), since the morphemes and words match combinatorial rules of German compounding.

If an encountered string is actually a lexical item stored in the mental lexicon, the sequence of (references to) morphemes can be replaced by a single reference to (a copy of) the lexical item.

And similarly, a sequence of references to lexical items can be combined into a phrase, and a sequence of words can be replaced in STM by a single reference to a phrase, provided that the combinatorial principle or rule allows this. This is basically *Merge* as used by Chomsky in syntax, though Chomsky is never or seldom explicit about the requirement about the combinatorial principle or rule that allows this. If the sequence of words cannot be combined into a structure, the same effects arise as we saw for arbitrary sequences of characters, as e.g. in an arbitrary list of digit words (10), names (11), or words (12):

(10) Seven four three nine eight six seven five five two
(11) John Peter Mary George Ann Ray Martin Elvis Mick Steven David Derek Alice
(12) Get by of tired the sitting beginning was very on sister bank her Alice to

If anything, with ordinary words the task is probably even more difficult than with digit words, because there are many more non-digit words than digit words.

Example (12) is the most interesting one, because if we take a string with *exactly the same words* in a different order as in (13):

(13) Alice was beginning to get very tired of sitting by her sister on the bank

reproducing the sentence (and in fact even understanding it) is much easier for humans.

Presenting a large number of arbitrary sequences of words and asking subjects to reproduce them is generally not very pleasant for these subjects. At best, it is considered by them a hard intellectual challenge. But every four-year old daughter insists that a hundred or so such strings (a story) are presented to her every evening, and she actually enjoys it and is very disappointed if she is not told such a story! But of course, she does not want arbitrary sequences of words, but sequences of words that match the combinatorial rules of the syntax of her native language. And just an arbitrary sequence of sentences will not do: it must be a sequence that meets several requirements, in particular requirements of discourse coherence, so that the sentences can be integrated in a structure also at the discourse level. See [Levinson & Evans 2010] and references there, and [Koschmann 2010], who argue that also recursion can arise in discourse.

Summarizing: when two elements from module X enter the STM, the structure building module can combine them in a single structure, and replace them in STM by a single reference to this structure, provided that there is a rule or principle specific to module X that allows the elements to form such a structure. In this way the STM is relieved. The elements that are part of the structure are still stored, but not in STM but in Working Memory. Working Memory cannot be used directly to relieve STM, since (by hypothesis) elements in working memory can only be accessed via elements in STM.

Though I discussed use of the structure building module in terms of analysis of input, which might be interpreted as a performance factor, this is no way essential. The structure building module is part of competence, but of course it can be used during performance (analyzing or producing expressions).

Though the structure building mechanism relieves the STM, one can create structures where the structure building mechanism does not succeed in reducing the contents of STM. This happens in multiple self embedding ('parenthetical' embedding) constructions, which become very difficult to process after 3 or 4 center embeddings. [Chomsky & Miller 1963: 286], who illustrate it with the example

(14) The rat the cat the dog chased killed ate the malt.

of which they state that it is "perfectly grammatical" though "surely confusing and improbable".

Working Memory is also limited in size, and this might be the locus of one type of locality condition on syntactic processes such as subjacency.[54] If the 'size' of Working Memory is defined in more abstract structural terms, it might also be the locus for locality conditions such as conditions on phases [Chomsky 2008] and on relativized minimality [Rizzi 1990]. However, if Working Memory is the locus for such locality constraints, one would predict that linguistic locality constraints are specific instantiations of domain-independent locality constraints, which therefore may also show up in other linguistic modules than syntax such as orthography, phonology and morphology, and in nonlinguistic modules such as music, dance and artificial languages (in logic, mathematics, programming). Whether that is the case is not clear to me; usually the locality constraints in linguistic modules other than syntax are even stricter than in

[54] [Marcus 1980] attempts to derive locality constraints from the design of a natural language parser that has a limited buffer (5 positions with a sliding window of 3 positions). He also (p. 197, fn 11) suggests that the difficulty of processing multiply center-embedded structure might be due to the limited buffer size.

syntax; the facts of nonlinguistic modules are not known to me. However, [Katz & Pesetsky 2011] and [Pesetsky in prep.] argue explicitly that structure building operations used in music are (at an abstract kevel) identical to structure building operations in language, including head movement ([Katz & Pesetsky 2011] and phrasal movement [Pesetsky in prep], and they do so under the slogan 'Same recipe, different ingredients', which fits in perfectly in the proposal developed in this paper.

Under the hypothesis put forward here, the grammatical component does not contain structure building operations itself, but must make use of the external structure building module. Since the structure building module is not limited to specific domains, it follows that the nature of the structure must be identical (at an abstract level) across all domains (syntax, morphology, phonology, etc.), including the determination of the label of the resulting structure. The label may of course itself be domain specific but is created in a way that is not domain-specific. This creates a whole new domain of investigation, i.e. are indeed all structures from various domains identical at an abstract level? Several researchers have argued in favor of this, e.g. [Kaye et al. 1990], [Harris 1994], [Carstairs-McCarthy 1999] and others argue for similarity in structure between syllables and syntax, and we already mentioned Katz and Pesetsky's claims for parallels between syntactic structure and the structure of music.

As to the determination of the label of a newly created structure, Chomsky indeed makes attempts, in his minimalistic framework, to derive its nature from general principles that perhaps are not specific to syntax. He assumes that the label of the resulting structure, if it is assigned at all, is identical to one of the labels of the composing parts. That is a strong hypothesis, possibly too strong.[55] This assumption is shared by many grammatical frameworks but usually restricted to so-called *head features.* As far as I can see, Chomsky always restricts attention to features such as syntactic category, φ-features, and *case*: these are indeed considered head features in most grammatical frameworks. However, features to specify the syntactic combinatory potential (subcategorization features) are generally not considered to be head features, and are never explicitly considered by Chomsky in this context. Other grammatical frameworks have different conventions for labeling a newly created structure with such features (e.g. the HPSG SUBCAT principle, [Pollard & Sag 1987:11]). So I believe that Chomsky's attempts are in the right direction, but need considerable refinement. Deriving such principles of labeling in a sufficiently abstract manner so that also labels of phonological, orthographical structure, etc. are determined by the same principles is even further away, but again, I believe, they form an interesting but quite challenging avenue of research.

---

[55] [Katz & Pesetsky 2011:20] also mention all kinds of reasons why this thesis is probably too strong. The examples they mention, however, are valid for Chomsky's usual assumptions, but they involve properties such as the phonological form, declension and conjugation class features that in the model sketched here would appear only in concrete form modules, hence not in the grammatical component, and would therefore not pose a problem for labeling structures in syntax.

Chomsky has also argued that movement is a special case of Merge (called Internal Merge). If we translate this to the current assumptions, it would imply that internal structure building is also an operation available to other modules than syntax or even linguistic modules. I have no idea whether there is any evidence for this, and it appears unlikely, since movement is generally considered to be highly specific to the syntax of human natural language. This makes it less plausible that movement can be considered a special case of Structure building. Furthermore, the proposal that movement is a special case of structure building faces a number of problems that are maybe technical in nature, but might point to a more fundamental problem. After all, Internal Merge creates a copy, but it does not establish the relation between the two copies. Therefore I do not see how the generation of a sentence such as (15) can be avoided:

(15) John said that left

(intended meaning: *John said that John left)*
After all, it can be derived starting from

(16) John said that John left

Spell-out will now erase and delete the second occurrence of *John* since they are (accidentally) identical (same lexical item, same features: nominative case, singular, etc.).

(17) John said that ~~John~~ left

In order to derive movement from Internal Merge, one must therefore somehow be able to distinguish between intended and accidental identity

This problem has been observed, e.g. by [Chomsky 1995:227] (see also [Chomsky 2005: 12]). He suggests that two lexical items *l* and *l'* should be marked as distinct if they enter the derivation via different applications of *Select*, and he contrasts it to copies created by Internal Merge (which **are** identical, see p.251-252) But this does not avoid the need to mark somehow which lexical items are identical and which are not, and thus the reduction of movement to Internal Merge is incomplete. [Frampton 2004] suggests that Internal Merge just adds a link to an SO already contained in the structure, in effect creating *token identity* between target and source in movement. Assuming token identity requires adaptation of the spell-out algorithm (just deleting an element would result in complete deletion of the element), which Frampton also provides. So maybe, the reduction of Move to Internal Merge can be saved by assuming token identity, but Chomsky has never been explicit about this and appears not to assume token identity in [Chomsky 1995: 251-252], cf. "the element α now appears **twice** in Σ'" (p. 252) .

Finally, if structure building is applicable to objects from any domain, it will also be applicable to objects from the C-I-module. This has three potential effects. First, people can, using their structure building module, form indefinitely complex conceptual structures, and thus entertain much more complex thoughts than animals even independently of language. Second, complex conceptual structures might be built that cannot be expressed by structured grammatical constructions, and can only be expressed by associating such a complex conceptual construction with a single basic grammatical unit (kinship terms, and the analysis of *kill* as *cause-to-become-not-alive* come to mind). If this is correct, language will contain basic grammatical units for complex concepts that animals cannot have (since they have no structure building). Finally, and even more speculatively, if building complex structures for forms has to go through STM since they have direct relations to input-output systems, and if this is not necessary for building conceptual structures which can take place outside of STM, conceptual structures will not be subject to the same locality constraints as e.g. syntactic structures. This might be a source for the explanation of why overt movement obeys subjacency but covert movement does not (and why other locality constraints that are formulated in terms of the structures themselves and not in terms of the memory space they occupy such as relativized minimality do hold in conceptual structures). Needless to say, these are all very tentative and speculative ideas, for which it is not clear at all they can be upheld.

8 So, how did language originate and evolve?

Assuming what I have proposed in the preceding sections, how did language actually originate and evolve? This is a difficult question, which requires even more speculative and tentative assumptions than in the preceding section, but I will make an attempt anyway.

Two independent changes are required from our direct predecessor to humans with full human linguistic capacities, in accordance the proposed theory. The third change is dependent on the appearance of the grammatical module and (by hypothesis) a fully automatic consequence of it. I have no idea which of these changes came first, but I will tentatively assume that addition of the grammatical module (as a copy of the C-I module) comes first.

This change must have been caused by some mutation in our DNA, and the effect that it had must at first have been at least not detrimental from an evolutionary point of view. I assume that initially this change was evolutionary neutral, i.e. it does not yield any benefits but also no significant negative effects. Since genes, especially regulatory genes, usually do not have just one but multiple effects in different parts of the body (pleiotropy), it is also possible that the change had other beneficial effects on reproductive adaptability. This will allow the mutation to spread through a community, though slowly and perhaps not among all members.

Next, within at least one community of the hominids in which at least some members have the mutation, someone must have come up with a first mapping between a sound and a meaning. The community members that have the mutation have the capacity to do so (since the grammatical module links the SGF and C-I components), but this does not mean that this capacity is immediately used. Fortunately, mouth-produced sounds and gestures are made by these hominids for social interaction, expressing emotions etc. and this input will not only find its way to the social interaction and emotional modules, but also (via the grammatical component) to the C-I component (since this is fully automatic and cannot be controlled). This must have led at a certain point, at least for one very bright member, to the realization that such sounds can be used for expressing concepts, and when this bright member is in an appropriate social group, this idea and its realization can spread throughout the group. In this way, we get a community that can use some initial words. I assume that the number of words grows, so that more and more thoughts can be expressed, and we get a sort of very very primitive proto-language.[56]

In a community that uses such expressions, the children will receive regular proto-language input so that their grammatical modules start to develop. Therefore, very soon (possibly just one generation) after a community uses protolanguage to a sufficient degree, a real but extremely simple language can arise, much in the same way as has been described by Bickerton in various works for creolization of pidgins, and similar to creolization of Nicaraguan pidgin sign language into Nicaraguan Sign Language ([Bickerton 1990] [Kegl 2002]). A grammaticalization process can start since the children have a grammatical module that is stimulated (gets the right input in an appropriate social environment, etc.).

However, the 'language' that we are talking about here is still very very simple. Since there is no structure building yet, short term memory imposes strong constraints on possible expressions: words can only be very short (one sound, two sounds), they cannot be internally structured morphologically and sentences can also only be very very short (2, maybe 3 words).

The second change, the capacity to build structure and thus alleviate the STM, must have caused a real 'big bang': it gives the potential to process multiple sounds in a syllable, multiple syllables in a morpheme, multiple morphemes in a word, multiple words in a phrase and multiple phrases in a sentence. The PF component, which so far probably was just a very minor module, could now develop in full, and each relevant module can now be used at a much larger capacity. With that development, real human natural language as we know it today started. (see [Chomsky 2005:5] for a similar view).

---

[56] [Chomsky 2010:54-55] appears to suggest that the grammatical component arose (long) before any link with the external sound system was made. Such a scenario is excluded in the thesis being developed here, since the grammatical component as conceived here can develop only when it receives input from the SGF modules. But Chomsky probably has recursion in mind, and that is a possibility allowed by the thesis here. See below.

From that moment on, ‘glossogeny’[57] could also start, *inter alia* as a consequence of mismatches between the abstract forms of grammatical module and the concrete forms in PF and the actual sounds: for example, phonological processes may cause a different input to the grammatical module, which has to adapt itself to it (e.g. a stress shift, followed by reduction of final syllables, leading to lack of evidence for abstract inflectional categories, as happened in many Germanic languages). On the other hand, certain innately determined properties of the grammatical module are so strong that they will override actual speech input (as e.g. in regularization of inflectional paradigms).

The order of changes might also have been the other way around. In that case, structure building may have been applied to the C-I module before language arose, which would be compatible with the scenario that [Chomsky 2010:54-55] sketches. Other typically human capacities dependent on structure building such as music, dance and counting may have originated before language, and language arose almost immediately once the grammatical module was in place, initially as a proto-language in the Bickertonian sense, but shortly (possibly just one generation) after that as a full human natural language.

The theory proposed here assumes that the capacity for human natural language arose only once. This seems natural, since the odds of such a complex system to arise multiple times in such a, from an evolutionary point of view, short period are extremely low. However, the theory proposed here is perfectly compatible with multiple origins of individual languages. For example, a group of hominids who did not get the idea of using words themselves, may have adopted the idea from seeing other groups using (proto)language and may have developed their own language, with, of course, its own idiosyncratic sound-meaning mappings. In fact, the example of the Nicaraguan Sign Language clearly shows that this is a realistic possibility.

Finally, currently all (normal) humans have language. Perhaps, having language brought such a big reproductive advantage, that none of the hominids without language survived. But surely, such advantages could only manifest themselves as soon as language had come into existence.
Needless to say, all these remarks are highly speculative and tentative, but at least the theory proposed here provides all the ingredients for the hominid brain to develop into the human brain with the full capacity for human natural language.

9 Concluding Remarks

In this paper I sketched a tentative and speculative account of the modules in the brains of humans that are involved in language, and how they developed from the modules in the brain of our direct predecessor. The account is based on the CC-theory, consisting of three very tentative

[57] i.e., historical language change [Fitch 2010:90], citing [Hurford 1990] as the one who first coined this term.

and speculative hypotheses that together account for the origin of the major aspects of natural language. The core hypothesis (which I will call the Conceptual Copy Hypothesis or CC-Hypothesis) states that a very small change in the genes of our ancestors had the effect that a second copy of the conceptual-intentional (C-I) component develops: this small change at the genotype level was argued to have dramatic consequences at the phenotype level: this new copy of the C-I component starts to function as the grammatical component and many properties of the grammatical component are derived from assuming this origin. These include the existence and nature of the basic grammatical units, the fact these are not atomic but have properties, the combinatorial nature of the grammatical component, the fact that open class and closed class word types exist, the precise properties for each word type, and also the word types themselves. The hypothesis also predicts that a large lexicon (at least several tens of thousands of entries) can be processed efficiently in the grammatical component, and it even makes (intensional) predictions on what kind of mechanism is used for that. Finally it also accounts for the fact that the grammatical system is such a acquisition-eager component The hypothesis also shines a new light on the relation between grammar and semantics, in particular it offers a completely new perspective on the intuition of many linguists that grammatical categories are somehow grounded in semantic categories, but it does so while maintaining the independence and the autonomous character of the grammatical component. It also predicts that there is a universal set of grammatical properties, from which each language uses a subset, and provides a new source of evidence for determining the exact nature of this set. The basic idea, if correct provides a solution for the paradox that language evolution must have involved just one or a few very small changes while at the same time the grammatical component is a very rich and intricate system that is to a large extent specific to humans and to language.

The second hypothesis states that this new component makes a link with the already existing system to generate and interpret mouth-produced sounds, gestures and facial expressions in use in primates and our direct ancestor for emotive calls and social interaction (the SGF modules). It thus accounts for the fact that the primary media for language are speech and gesturing, and, together with independent facts and assumptions, for the fact that natural language uses symbolic reference rather than indexical or iconic reference.

The third hypothesis states that integrating two elements into a structure (structure building) is an operation in an independent module, unique to humans but not to language. Recursion can arise from this structure building under specific conditions. It thus accounts for the fact that structure building (and in many cases also recursion) can be used for objects from many faculties and modules, both linguistic ones (orthography, phonology, morphology, syntax, and discourse grammar) and non-linguistic ones (counting (mathematics), music, dance, vision, thinking, artificial languages in logic, mathematics and programming).

The account I gave is framed in a multi-component analysis of language, and is completely functional in nature, i.e. it specifies which modules are assumed to exist, what their role is, and how they relate to other modules, but it does not state anything about the exact implementation of these modules in the human brain. With regard to that, the account given is compatible with many different variants, ranging from an account purely in terms of patterns of brain activity to an account purely in terms of dedicated physiological structures, and anything in between.

Let me conclude with some methodological remarks. First, in the theory that I propose in this paper, I try to explain the properties of the grammatical module, a module that we have at least some understanding of, from properties of the conceptual module, which, unfortunately, we have hardly any knowledge of. That makes the investigation of this hypothesis a very hard job, most of which still has to start. Second, most of what we (think to) know about the conceptual module, we know though our investigation of language. That of course introduces a serious danger of circularity. It is therefore essential that for each conceptual category postulated it is shown that it can be justified independent of language, e.g. by showing that it plays a role in nonlinguistic beings such as primates or other animals, in prelinguistic beings such as very young children, and at the very least that there is evidence for a conceptual category independent of the correlating grammatical category.

As stated before, the CC-theory is tentative and highly speculative. However, even if the theory I propose here would turn out to be completely wrong, I believe it is interesting to study. The theory postulates one very small change on the genotype level, causing a big change at the phenotype level, and it describes the origin of language as a consequence of a difference in the configuration of modules in the brain. With these properties, this theory, in my view, has the properties that the correct theory of the origin of language should have, and therefore it is interesting to study this theory even if it would turn out to be completely wrong.

10 References

[Bates 1976] Bates, E. (1976). *Language and context: The acquisition of pragmatics*. New York: Academic Press

[Bickerton 1990] Bickerton, Derek (1990), *Language and Species.* Chicago: University of Chicago Press.

[Bickerton 2009] Bickerton, Derek (2009), *Adam's Tongue. How Humans Made Language, How Language Made Humans*. Hill & Wang, New York.

[Bolk 1926] Bolk, L. 1926. *Das Problem der Menschwerdung*. Fischer, Jena.

[Carstairs-McCarthy 1999] Carstairs-McCarthy, A. (1999), *The Origins of Complex Language.* Oxford: Oxford University Press.

[Chomsky 1957] Chomsky, N. (1957), *Syntactic Structures.* The Hague: Mouton.

[Chomsky 2004] Chomsky, N. (2004), 'Biolinguistics and Human Capacity', lecture delivered at MTA, Budapest, May 17, 2004 http://www.chomsky.info/talks/20040517.htm. Also appeared as Chapter 7 of Noam Chomsky, *Language and Mind, Third Edition*, Cambridge University Press pp 173-185.

[Chomsky 2005] Chomsky, Noam (2005), 'On Phases', unpublished manuscript, MIT.

[Chomsky 2010] Chomsky, N. (2010), 'Some simple evo devo theses: how true might they be for language?', in [Larson et al. 2010], pp. 45-62.

[Chomsky 2011] Chomsky, N. (2011), Lecture on Linguistics, Leiden, the Netherlands, March 14, 2011. http://media.fsw.leidenuniv.nl/getmedia.php?id=5119

[Chomsky & Miller 1963] Chomsky, Noam and George Miller (1963). Introduction to the formal analysis of natural languages. In Luce, R., Bush, R. and Galanter, E. (eds.) *Handbook of Mathematical Psychology, Vol 2.* New York: Wiley. 269-323.

[Corbett 1991] Corbett, Greville (1991), *Gender*, Cambridge Text Books in Linguistics, Cambridge University Press.

[Corbett 2000] Corbett, Greville (2000), *Number.* Cambridge Text Books in Linguistics, Cambridge University Press.

[Christiansen & Kirby 2003] Christiansen, Morten H. and Simon Kirby, eds. (2003), *Language Evolution.* Oxford: Oxford University Press.

[Curtiss 1977] Curtiss, S (1977). *Genie: a psycholinguistic study of a modern-day "wild child"*. Boston: Academic Press.

[Deacon 1997] Deacon, Terence W. (1997), *The Symbolic Species*: *Co-evolution of Language and the Brain*. New York: W.W. Norton & Company,

[Dehaene 1999] Dehaene, Stanislas (1999), *The Number Sense*. London. Penguin Books (original edition 1997).

[Duden 1998] Duden (1998) Band 4 Grammatik der gegenwartigen Deutschen Sprache 6. ed. Mannheim: Duden Verlag.

[Evans & Levinson 2009] Evans, Nicolas & Stephen C. Levinson (2009), 'The myth of language universals. Language diversity and its importance for cognitive science' BEHAVIORAL AND BRAIN SCIENCES (2009) 32, 429–492. doi:10.1017/S0140525X0999094X

[Everett 2005] Everett, D. L. (2005) 'Cultural constraints on grammar and cognition in Pirahã. Another look at the design features of human language'. *Current Anthropology* 46:621–46.

[Fillmore 1968] Fillmore, Charles J. (1968) "The Case for Case". In Bach and Harms (Ed.): *Universals in Linguistic Theory*. New York: Holt, Rinehart, and Winston, 1-88.

[Fitch 2010] Fitch, W. Tecumseh (2010), The *Evolution of Language*. Cambridge: Cambridge University Press.

[Frampton 2004] Frampton, John (2004) 'Copies, Traces, Occurrences, and all that. Evidence from Bulgarian multiple wh-phenomena'. Unpublished paper http://www.uni-leipzig.de/~muellerg/frampton04.pdf

[Gould 1977] Gould, Stephen Jay (1977), "short-list of 25 characters reprinted", in Ontogeny *and Phylogeny*, Cambridge, Massachusetts: The Belknap Press of Harvard University Press, p. 357. http://www.sjgarchive.org/library/text/ontogeny/p0357.htm

Gould, Stephen Jay, & Eldredge, Niles (1977). "Punctuated equilibria: the tempo and mode of evolution reconsidered." *Paleobiology* **3** (2): 115-151.

[Greenberg 1963] Greenberg, J. (1963), "Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements," *Universals of Language*, London: MIT Press, pp. 110-113.

[Harris 1994] Harris, John (1994), *English Sound Structure*. Oxford: Blackwell.

[Hauser 1996] Hauser, Marc (1996), *The Evolution of Communication*. Cambridge Mass., MIT Press.

[HCF 2002] Hauser, M.D., N. Chomsky & W.T. Fitch (2002), 'The Faculty of Language: what is it, who has it, and how did it evolve', *Science* 298, 1569-1579.

[Hauser & Fitch 2003] Hauser, Marc D. & W. Tecumseh Fitch (2003), 'What Are the Uniquely Human Components of the Language Faculty?', in [Christiansen & Kirby 2003], pp.158-181;

[Hurford 1990] Hurford, J (1990), 'Nativist and functional explanations in language acquisition', in *Logical Issues in Language Acquisition,* ed. I. M. Roca, pp 85-136. Dordrecht: Foris Publications.

[Hurford 2003] Hurford, James R. (2003) 'The Language Mosaic and its Evolution', in [Christiansen & Kirby 2003], pp.38-57.

[Itakura 1996] Itakura, S. (1996), 'An exploratory study on gaze-monitoring in nonhuman primates', *Japanese Psychological Research* 38(3):174-80.

[Jackendoff 1992] Jackendoff, Ray (1992), *Languages of the Mind*. Cambridge, Mass., MIT Press,

[Jackendoff 1994] Jackendoff, Ray (1994), *Patterns in the Mind. Language and Human Nature*. New York: Basic Books.

[Jackendoff 1997] Jackendoff, Ray (1997), *The Architecture of the Language Faculty*, Linguistic Inquiry Monograph 28, Cambridge, Mass.: MIT Press.

[Jackendoff 2002] Jackendoff, Ray (2002), *Foundations of Language. Brain Meaning, Grammar, Evolution.* Oxford, Oxford University Press.

[Jackendoff & Pinker 2005] Ray Jackendoff & S. Pinker (2005), 'The nature of the language faculty and its implications for evolution of language (Reply to Fitch, Hauser, and Chomsky)', *Cognition* 97, pp 211-225

[Katz & Pesetsky 2011] Katz, Jonah & David Pesetsky (2011), 'The Identity Thesis for Language and Music', unpublished paper, Version 2.1, revised January 2011. http://ling.auf.net/lingbuzz?_s=wduLK0tIeLoKRVHS&_k=Rfc_6Rz7P-wfS6s9&3 ,

[Kaye et al. 1990] Kaye, Jonathan, J. Lowenstamm & J-R Vergnaud (1990),'Constituent Structure and Government in Phonology', *Phonology* 7: 193-231

[Kegl 2002] Kegl, J. (2002), 'Language emergence in a language-ready brain: Acquisition Issues', in *Language Acquisition in Signed Languages*, ed. G. Morgan and B. Woll, pp. 207-254 Cambridge: Cambridge University Press..

[Koschmann 2010] T. Koschmann (2010), 'On the universality of recursion', *Lingua* 120, pp. 2691-2694.

[Koster 1982] Koster, Jan (1982), Karel, Darwin en Chomsky: over taal en evolutie, *Hollands Maandblad* Vol. 24, No. 415-416, p. 3-21

[Larson et al 2010] Larson, R.K., V. Déprez, and H. Yamakido (2010), *The Evolution of Human Language – Biolinguistic Perspectives,* Cambridge: Cambridge University Press.

[Lerdahl & Jackendoff 1983] Lerdahl, Fred & Ray Jackendoff (1983), *A Generative Theory of Tonal Music*, Cambridge, Mass.: MIT Press.

[Levin 1995] Levin, Beth (1995), *English Verb Classes and Alternations. A Preliminary Investigation.* Chicago: University of Chicago Press.

[Levinson & Evans 2010] Stephen C. Levinson & N. Evans (2010), 'Time for a sea-change in linguistics: Response to comments on 'The Myth of Language Universals'', *Lingua* 120, pp. 2733-2758.

[Lieberman 1984] Lieberman, Philip (1984) *The Biology and Evolution of Language.* Cambridge, Mass. Harvard University Press.

[Macnamara 1982] Macnamara, John (1982), *Names for Things.* Cambridge, Mass: MIT Press.

[Marantz 1981] Marantz, A. (1981), *On the Nature of Grammatical relations*. PhD, MIT.

[Marcus 1980] Marcus, M. P. (1980). *A Theory of Syntactic Recognition for Natural Language.* Cambridge, MA: MIT Press.

[Markson & Bloom 1997] Markson, L. & P. Bloom (1997), 'Evidence against a dedicated system for word learning in children', *Nature* 385, pp. 813-815. [pdf]

[Mayr 2002] Mayr, Ernst, *What Evolution is,* London: Butler and Tanner.

[Miller 1956] Miller, G. A. (1956). "The magical number seven, plus or minus two: Some limits on our capacity for processing information". *Psychological Review* **63** (2): 81–97. doi:10.1037/h0043158. PMID 13310704. (pdf)

[Miller & Chomsky 1963] Miller, George and Noam Chomsky (1963). Finitary models of language users. In Luce, R.; Bush, R. and Galanter, E. (eds.) *Handbook of Mathematical Psychology, Vol 2.* New York: Wiley. 419-93.

[Mitchell et al. 2008] Mitchell, T., *et al* 'Predicting Human Brain Activity Associated with the Meanings of Nouns', *Science* 320, 1191, pp.1191-1195.

[Musso et al 2003] Musso, M. *et al* (2003), 'Broca's area and the language instinct', *Nature Neuroscience* 6(7), pp. 774-781.

[Nevins et al. 2009] Nevins, Andrew, David Pesetsky, & Cilene Rodrigues (2009),'Pirahã Exceptionality: A Reassessment', *Language* 85(2), pp. 355-404.

[Odijk1993] Odijk, J. (1993), 'Natural Language, Communication, and Man-Machine Interfaces, unpublished paper, IPO, Eindhoven.

[Odijk 1997] Odijk, J. (1997),'C-selection and S-selection', *Linguistic Inquiry* 28(2), pp. 365-371.

[Peirce 1978] Peirce, Charles S. (1978), *Collected Papers. Volume II: Elements of Logic*. Cambridge, Mass., Belknapp Press.

[Pallier et al. 2011] Pallier C,, Devauchelle A-D, Dehaene S (2011) 'Cortical representation of the constituent structure of sentences', Proc Natl Acad Sci USA 108:2522–2527.

[Pesetsky 2010] Pesetsky, David (2010), 'Russian Case Morphology and the Syntactic Categories', unpublished manuscript, Oct 10, 2010.
http://ling.auf.net/lingbuzz?_s=wduLK0tIeLoKRVHS&_k=XEXHl06Zm77j_GT3&5

[Pesetsky in prep.] Pesetsky, David (in prep.), 'Same Recipe, Different Ingredients. Music Syntax is Language Syntax', unpublished manuscript.

[Pinker 1984] Pinker, S. (1984). *Language learnability and language development.* Harvard UP. : Semantic Bootstrapping

[Pinker 1994] Pinker, S. (1994), *The Language Instinct.* New York: HarperPerennial.

[Pinker 2007] Pinker, S. (2007), *The Stuff of Thought. Language as a Window into Human Nature*. London, Viking Penguin.

[Pinker & Bloom 1990] Pinker, Steven & Paul Bloom (1990), 'Natural Language and Natural Selection', *Behavioral and Brain Sciences* 13:707-26.

[Pinker & Jackendoff 2005] Steven Pinker & R. Jackendoff (2004), 'The faculty of language: what's special about it?' *Cognition* 95, pp.201-236

[Pollard & Sag 1987] Pollard, Carl & Ivan A. Sag (1987), *Information-based Syntax and Semantics.* Stanford, CSLI Lecture Notes13.

[Povinelli et al 2000] Povinelli, Daniel J., Jesse M. Bering and Steve Giambrone (2002), 'Towards a Science of Other Minds: Escaping the Argument by Analogy. *Cognitive Science* 24:509-42.

[Reinhart & Reuland 1993] Reinhart, Tanya & Eric Reuland (1993), 'Reflexivity', *Linguistic Inquiry* 24 657-720.

[Reuland 2011] Reuland, Eric (2011) *Anaphora and Language Design*. Cambridge, Mass: MIT Press.

[Ridley 1999] Ridley, Matt (1999), *Genome. The Autobiography of a Species in 23 Chapters.* London, Fourth Estate.

[Rizzi 1990] Rizzi, Luigi (1990), *Relativized Minimality*. Cambridge Mass.: MIT Press.

[Ronshaugen et al. 2002] Matthew Ronshaugen, N McGinnis & W. McGinnis (2002), 'Hox protein mutation and macroevolution of the insect body plan', *Nature* 415 pp 914-917.

[Selkirk 1982] Selkirk, E. (1982), *The Syntax of Words.* Cambridge, Mass. MIT Press

[Smits 2009] Smits, Rik (2009), *Dageraad. Hoe taal de mens maakte*. Amsterdam: Nieuw Amsterdam Uitgevers.

[Stabler 1992] *The logical approach to syntax : foundations, specifications, and implementations of theories of government and binding*. Cambridge, Mass. MIT Press.

[Stabler 1997] Stabler, E. (1997) 'Derivational Minimalism', in C. Retoré (ed.) *Logical Aspects of Computational Linguistics.* LNAI Lecture Notes, LACL '96, Springer, Nancy, pp. 68-95. http://www.linguistics.ucla.edu/people/stabler/eps-lacl.pdf

[Verkuyl 1972] Verkuyl, H. (1972), *On the compositional nature of the aspects*. Dordrecht: Reidel

[Whitehead & Russell 1972] Whitehead, A.N. & B. Russell (1972/1927). *Principia Mathematica*. Cambridge: Cambridge University Press.

[Wilkins & Wakefield 1995] Wendy K. Wilkins and Jennie Wakefield (1995). Brains evolution and neurolinguistic preconditions. Behavioral and Brain Sciences, 18, pp 161-182 doi:10.1017/S0140525X00037924