The Origins Of Language

Alejandro Fernández

Abstract

The origins of language are reviewed from the perspectives of phonology, phonotactics and information content. A few well-defined evolutionary steps (or language eras) are proposed: monosyllabic words with a single consonant, monosyllabic words with two consonants, monosyllabic words with vowels and finally polysyllabic words. The available evidence for this evolutionary path is reviewed from several perspectives, including comparative linguistics, paleoanthropology and genetics. A number of alternative evolutionary paths in the literature are also discussed. Finally, two artificial languages are proposed to illustrate the expressive power of the earliest language eras: Singlecon and Doublecon.

Figure 1: Cave people in Bhimbetka.

The relative scarcity of modern speculation about prehistoric languages is surprising, given how much interest there is both in academia and among the general public about other aspects of prehistoric life. The most likely reason is lack of direct evidence, which in these often unimaginative times can be a showstopper. But there are many indirect approaches that can give interesting answers.

In this article we will explore how language originated and developed in prehistoric times, combining language reconstruction and information theory with

generous amounts of conjecture to arrive at a consistent view of language evolution. From beyond the mysts of time we will try to build a few defensible positions, or at the very least lay a path to future discovery.

Language Evolution

It is not reasonable to expect that language appeared at some point in the past, perfectly formed with all the elements that it has today.

Back In Time

Furthermore, in historic times there have been significant improvements. As Bakker *et al* (2002) hint, subordinate sentences were an innovation to Herodotus roughly 2,500 years ago (2.5 kya), who wrote endless sentences and often forgot to finish them (technically called anacoluthon).

Going further back into the past, partial reconstructions such as the enlightening work done with Indoeuropean shows that many elements of language that we take for granted are relatively recent additions. For instance, Proto-Indo-European (PIE) developed gender gradually, as explained by Luraghi (2011).

How far back can we go with reconstructions? García Calvo (2003) was able to reconstruct an even earlier prehistoric language, as we will see below. Some researchers have even posited a Proto-Human language, with little success and even less credibility. Past a certain point reconstructions are no use: our interest is not in tracing common roots for all languages, but to investigate how broad language features evolved.

We need some other tools to delve into the darkness of time.

Increasing Information

A very basic criteria that describes a language era is how many different words can be uttered. The number of possible words in a given language correlates strongly with how much information can be conveyed. We will see below an attempt at quantification.

Modern languages can be seen as open-ended: the number of productions is unlimited, as explained in Trask (2003). This however needs not be the case in the past.

We have to suppose that hominin languages started somewhere, by assigning meaning to a few sounds. As the number of possible utterances increased, some methods to combine them appeared, thus also increasing the amount of information conveyed. The process must have conferred a distinct advantage to speakers, and so it has continued until the present time.

A Timeline Proposal

With this preliminary analysis in mind we can create the following timeline, dividing language evolution into a few distinct stages.

Individual Sounds

Our most remote ancestors probably communicated using simple sounds. Many animals use distinct vocal utterances to express different meanings. Elements include: pitch, volume and articulation (the shape of the mouth to generate the sound).

At this point we cannot yet speak about words, since there is no language system. Each sound is uttered individually making its own “sentence”. Where there are no words there can be no proper sentences. And without words or sentences there is no “language” to speak about.

Do animals have a grammar, i.e. do they combine different signs into a coherent whole? More specifically, can chimpanzees talk? Yang (2013) compares the relative frequency of words uttered by a chimpanzee with those from human language, and concludes that chimpanzees do not have a grammar.

We can suppose that early hominins had a way to augment their communication beyond what other primates can do, even without a grammar: perhaps by emitting a bigger range of sounds, or by adding context information with their hands. As Tomasello (2008) refers, pointing is a very human activity. This humble activity requires having a model of someone else’s mind, and is therefore the basis for deixis.

This era is therefore the “ground zero” of language: the common base with animals, with deictic information added on top.

Single Consonants

The first stage of language may have involved single consonants. At this point the sounds from the previous step can be combined into sentences, making them basic building elements in very simplistic sentences. These simple elements to combine might be thought to have been content words: nouns, adjectives or verbs, as proposed by Hurford (2003). Instead we will work from the opposite assumption, that these simple elements were only function words: prepositions, negations, pronouns, quantifiers and adverbs. Why? Simply because of sheer numbers.

The problem of single consonants is that there are not many of them. The number of simple, distinct sounds that can be uttered separately, without combining with others and without significant mixing, is limited. Modern languages can have as much as 87 consonants for Taa, about 80 for Ubykh. A single consonant

word is thus limited to less than 100 words in the best of cases, and probably no further than 30 or so, since many of these consonants can be viewed as combinations of simpler sounds. These combinations may have evolved from consonant groupings, and be therefore modern constructions.

The number of consonants has likely increased with time. We can choose any number from 10 to 20 or so, with the reasonable assurance that at some point in time humans could only profer that number of words. A value of 12 seems certainly achievable, so it will serve as archetypical. We will use the symbol K (cursive ‘K’) to represent any single consonant.

Using content words would have limited utterances greatly, and specially the ability to combine them. Function words on the other hand can be combined readily, creating the first sentences. There are other reasons such as the use of pronouns (‘I’ and ‘you’) as deictic words in a sentence, without which conversations would be really inane and would be limited to point out that there are an entity from one of 12 possible categories out there, at any given time. Also, function words are always in a finite supply anyway, so they seem like a more natural fit for a very restricted set of words.

It is easy to defend a definite advantage for hominins that can combine simple meanings into one coherent whole, as compared with those that could only emit separate sounds. Group coordination becomes possible without having to see each other. Still, it is hard to see how a language of this kind could work. For illustration purposes, in the appendix we have an example of a single-consonant language.

At which point in time did the era of single consonants start? The determination of the era of single consonants will have to involve considerable guesswork, being so far back in time. The advantage of group coordination would definitely have been greater when hominins started to hunt in groups, which may have been around two million years ago. (2 Mya). According to Roach (2016), around this point *Homo erectus* started to hunt and scavenge animals, increasing its meat consumption. But even such a simple language would also have been useful in foraging. An educated guess might thus be that around 1 Mya hominins could profer single consonants and combine them into simple sentences. We can thus identify the rise of *Homo Erectus* with the appearance of proper language.

Double Consonant Syllables

At some point early humans learned to combine several sounds into one syllable. This great advance extends the range of possible words at least by an order of magnitude. Considering the simplest structure of two consonants, or KK , the number of words would be the square of the number of consonants. Adding a second consonant without a vowel may seem a strange choice. It is however a logical extension of the single-consonant words seen above, which does not require additional phonetic entities.

How do languages without vowels work? The rhotic pronunciation of English omits vowels in many words. Nuxalk or Bella Coola is an example of a contemporary language without vowels; you can hear a sample. From our modern perspective, not having vowels does not mean that all consonants are stuck together in an unpronounceable blob. It just means that vowels are not part of the language. Some kind of vocalic sound must still link the two consonants together, and mark syllable boundaries. This allows listeners to distinguish a double-consonant word KK from two single-consonant words in succession $K K$.

We will use the italic $\mathfrak{a}$ (uppercase schwa) to represent a vocalic separator. Bi-consonant syllables will therefore be rendered as $K\mathfrak{a}K$. The evolution so far can be represented as:

$$K \rightarrow K\mathfrak{a}K$$

In practice, $\mathfrak{a}$ can be a neutral vowel or schwa. It can also be a more definite vowel sound, either automatically assigned by prosodic rules. or chosen at random. This would mean that e.g. the variants ‘SAN’, ‘SIN’ or ‘SUN’ are in fact different renditions of the same word ‘S $\mathfrak{a}$ N’. Technically it is an epenthetic vowel, only present to link both consonants.

A two-consonant language might still preserve the original one-consonant functions words K , and add content words combining two consonants: $K\mathfrak{a}K$. A word during this era can therefore be either K or $K\mathfrak{a}K$. 12 consonants would thus yield $144 + 12 = 156$ combinations.

What brought forward this improvement? Calling syllables an *invention* may very well misleading: it is very likely that both preferring and parsing syllables require adaptations in the brains of speakers and listeners to handle the increased complexity. Probably the first words with two consonants were formed by common linguistic mechanisms: compounding and reduplication of function words. The appendix has a hypothetical double-consonant language.

Our reasonable guess for the double consonant era is that it overlapped largely with the emergence of *Homo Sapiens*, at around 200–100 kya. At this point modern hominins were evolving fast, with large brains and impressive advances in technology. An exponential increase in the complexity of their communications would have suited perfectly their new activities, and would be a great help for the the new challenges they were facing.

A note on the European branch of the family is in order. Neanderthals are usually classified as a distinct species from *Homo Sapiens*. Given the now ample evidence that both groups could and did interbreed, as reviewed in Groucutt *et al* (2015), it does not seem unreasonable to classify them as two different subspecies inside *Homo Sapiens*, as many scientists have advocated for decades. The degree of shared evolution is far from clear, but there may have been only minute differences between African and Neanderthal language evolution.

Vowels

At some point in time vowels become first-class citizens, just as meaningful as consonants. A proper vowel is represented by *A* (cursive ‘A’).

$$K \rightarrow K\text{ə}K \rightarrow KAK$$

With vowels the number of possible syllables grows a bit. Let us suppose that there are three possible vowels at some point in time, along with the original 12 consonants. The number of combinations is now: consonant + vowel + consonant, or *KAK*, with 432 combinations; adding the original 12 single-consonant function words *K* yields a total of 444 combinations. Note that function words may also acquire vowels to yield simple structures like *KA*, in order to increase their number to 36; in this case yielding 468 possible words.

The process of vowel acquisition may require larynx, ear and brain adaptations to generate, distinguish and process the new grammatical elements. The incorporation of vowels into consonantic languages may happen similarly to the Indo-European ablaut or more generally as vowel apophony: vowels enter the system as modifiers for existing roots.

Tracing the appearance of vowels in all languages is again a hard issue. At what point did this happen? Vowels may in fact have appeared several times throughout prehistory. Almost all known languages have at least a few vowels, so that we can infer that their appearance happened before the last diaspora at around 70–50 kya. We will suppose that it did not happen much earlier than that.

Monosyllabic Expansion

In this next period the number of words grows in a couple of different ways while they remain monosyllables. The basic problem of a monosyllabic language is how to cram a large number of words into a single syllable. There are a couple of expansions that may have happened at this point in time.

The first strategy is to increase the number of consonants and vowels. The basic inventory of consonants may have grown in time. As a modern example, Khoisan click languages (sample) have added clicks as regular consonants. The phoneme inventory may have grown from 12 consonants and 3 vowels to e.g. 20 consonants and 5 vowels, which yields 2000 combinations while following the same *KAK* structure.

The second strategy is to add tones. At this stage speakers may start distinguishing words by tone as many modern languages such as Bantu (sample) do nowadays. Adding four tones per vowel for the original structure *KAK* yields 1728 combinations.

Of course both of these innovations may have been used in conjunction to achieve even greater vocabularies. This increase of combinations may have enabled the separation of word forms for verbs, nouns and adjectives, effectively establishing word classes as we know them.

As we will see below, tones are a prerequisite to polysyllables, and therefore had to be acquired at this stage.

It is not unreasonable to suppose that all languages were monosyllabic at least until 50 kya, probably later. The new techniques to expand vocabulary sizes were likely there for the transition to the Upper Paleolithic as described by Bar-Yosef (2002), which must have required a broader lexicon of plants, animals, materials, tools and techniques.

Polysyllables

An immediate extension of two consonant words are three consonant words.

$$K \rightarrow K\grave{a}K \rightarrow KAK \rightarrow KAKAK$$

Continuing with the basic structure of 12 consonants and three vowels, there are now the following combinations: 12 *K*, 444 *KAK*, and 15552 *KAKAK*, for a total of 16008 combinations. A gain of an order of magnitude over monosyllabic words, even after the expansion period.

However, and a bit surprisingly, it does not look as if this extension has happened early in prehistoric times. For starters, tones had to be developed first. To see why we need to consider how the listener can tell apart words.

All polysyllabic languages use tones as a distinctive feature. E.g. in English, the only difference in pronunciation between ‘insight’ / nsa t/, ‘incite’ / n sa t/ and ‘inn sight’ / n sa t/ is intonation. Tones allow us to distinguish between function words (which usually have no word accent) and content words, and in many languages also between different content words which have developed different word accents.

In the case of a monosyllabic language, if a function word is ‘KA’ and a content word is ‘KAK’, a new polysyllabic ‘KAKAK’ will sound identical to the combination ‘KA KAK’, unless it develops an accent first ‘Á’ which makes it ‘KAKÁK’ or ‘KÁKAK’, or perhaps even ‘KÁKÁK’ with double word accent.

At what point did polysyllabic words arise? Arab and other semitic languages have trilateral roots consisting of three consonants. It looks like a straightforward extension of the two-consonant syllables seen above. Agmon (2010) shows that for Semitic languages, the transition from two- to three-consonant roots happened during the transition to the Neolithic period, or 15–10 kya.

García Calvo (2003, pp. 90–97) reconstructed a prehistoric language ancestral to Indo-European and Proto-Indo-European with just two consonants per syllable

and without vowels. In contrast, Proto-Indo-European (PIE) already has polysyllabic words and vowels; it is in fact a mostly modern language with suffixes, pronouns and conjugation. PIE was supposedly spoken around 7 kya, which is very close to historic times in many places of the world. A timescale was not given by García Calvo, but we may again deduce that the transition from monosyllabic language to PIE may have happened near the transition to the Neolithic period.

The Sino-Tibetan family might seem a perfect example of the transition to polysyllabic languages in near-historical times, since many languages have monosyllabic morphemes. But Michaud (2012) refers that these monosyllables are often the result of compacting longer words, and therefore not a direct descendent of monosyllabic languages. In the reconstruction of Baxter & Sagart (2014), Old Chinese was already polysyllabic. There are however some hints at monosyllabic Sino-Tibetan ancestors as in Lama (2012), as recently as 4 kya.

Languages probably adopted polysyllabic forms as the necessity arose, given their particular information density configurations. In combination with other monosyllabic techniques they may have easily achieved vocabularies of tens of thousands of words. These newfound combinatorial riches allow for conjugations, declensions and other modern language features. This leaves us just right before the invention of writing, which is the point where Prehistory ends along with our journey.

Time Scales

Figure 1 shows several time scales for language eras as discussed.

Figure 1-a shows how the period under study looks when represented linearly. Hardly anything can be shown since most of the period is occupied by the single consonant era.

Figure 1-b shows the main eras along with the hominin species that was thriving at the moment. The single consonant era has been compressed for easier visualization. Note that Neanderthals are included in *Homo Sapiens*.

Figure 1-c shows an alternate time scale with logarithmic axis. The scale has been expanded to include the apparition of the first hominins, about 5 Mya. The graph ends at the beginning of history around 5 kya. Note that it would not be particularly hard (and it may be tempting) to make all eras approximately equal in the logarithmic axis, which would give a nice picture of constantly accelerating linguistic progress. But the change would break all the known boundaries of cultures and species.

Figure 2: Figure 1: Language time scales. Prepared by the author.

Alternative Timelines

There are many possible objections that can be raised about this timeline. In this section we will examine a few.

Recent Origin

Many scholars support a much later date for the appearance of language, around 100 kya, as reported by Dediu *et al* (2013). This is not supported by evidence.

The hyoid bone is a critical part of our capacity to speak; more specifically it is used for articulation. Martínez *et al* (2008) report that a couple of hyoid bones of hominins from 530 kya are very similar to samples from modern humans. D'Anastasio *et al* (2013) find the hyoid bone in Neanderthals to be structurally very similar to modern samples. Even further, Martínez *et al* (2013) find that 530 kya hominins had a similar auditory range to modern humans.

Barring some theory that explains why non-speaking hominins had similar bones as we use today for speech, these findings are probably enough to discard this timeline. Furthermore, a sample from 3.3 Mya ago is very similar to the hyoid bone of modern chimpanzees, which seems to imply that speech capabilities were developed between 3.3 Mya and 530 kya.

Much Earlier Origin

There is also the opposite belief: that language was already fully formed at least 100 kya. A good review can be found in von Petzinger (2017, pp. 70–71). The supporting evidence is intriguing but certainly not bulletproof: geometric patterns engraved on common items and personal adornment during the Middle Paleolithic.

We have however reviewed how polysyllables appeared much later in the game in some language families, right before or during the Neolithic period. Monosyllabic languages simply do not have the necessary breadth of vocabulary to incorporate features such as gender, declination or even tense. Even in highly analytic languages such as English, a restricted number of words would make it hard to construct a modern language. We have also seen reliable reconstructions of the appearance of such features as gender or declination in proto-languages.

A more gradual appearance of modern language, in accordance with other signs of human progress such as art or technology, appears to be more in line with the existing evidence. Still, this area will need much more study before it can be reliably clarified.

Earlier Introduction Of Vowels

The choice of early words consisting of two consonants may seem arbitrary. It is very well possible that vowels were introduced earlier, perhaps with the first syllables. Words might consist of one consonant plus one vowel; just adding three vowels to a 12-consonant language would extend the number of words from 12 to 36.

But there is some evidence supporting the late appearance of vowels. García Calvo (2003, pp. 87–89) made the argument that prehistoric Indo-European in its earliest incarnation had no vowels: just function words with a single consonant, and content words consisting of two juxtaposed consonants. As we saw before there is no timescale for this language.

On the other hand, there are languages with very different syllable structures, such as Upper Arrernte (sample). Interestingly enough, almost all languages with simple syllable structures. (V or CV) are almost exclusively found in Africa, South America or Pacific Islands, suggesting some ancestral clustering.

Ultimately this question is hard to answer and not terribly significant. It is possible that certain proto-languages have developed vocals first, and they may have added a second consonant to syllables later.

$K \rightarrow KA \rightarrow KAK.$

Compare this path with the one described above:

$K \rightarrow KəK \rightarrow KAK.$

More complex paths are possible: developing vocals first, losing them later in favor of biconsonant words, and regaining them later.

$$K \rightarrow KA \rightarrow K\text{ə}K \rightarrow KAK.$$

In any case, the number of words grows monotonically.

Earlier Monosyllabic Expansion

Our estimation for the monosyllabic expansion is from 50 to 15 kya, due to the divergence of schemes to augment monosyllables found across the world. There is a small problem with this argument. When *Homo Sapiens* left Africa, there was already a sizeable population of Neanderthals occupying Europe and Asia. So it is reasonable to think that these same Neanderthals might have their own languages, which influenced any newcomers. In this case the monosyllabic expansion might have happened much earlier.

It may be an interesting avenue of enquiry. Its first obstacle is again to reconcile the earlier monosyllabic expansion with the late appearance of polysyllables that we have seen in many language families. What happened in between? Unless a satisfactory answer for the extended era is found, this point of view should be put on hold.

Broader Monosyllabic Expansion

For the monosyllabic expansion a number of mechanisms have been proposed above. There are many others that can be seen in modern use, but that are not likely to have been a part of this era. We will review them along with the reason to have discarded them.

The first strategy is to increase the number of consonants.

- Gemination, or reduplication of consonants. These longer versions are usually derived from intermediate consonants, and therefore cannot appear from monosyllables alone.
- Compound consonants: when a combination of consonants becomes a single phoneme. The idea of compound phonemes in Clair (1972) can be expanded to many other consonants. An example can be how the modern sound for ‘ch’, is tellingly shown in IPA as $tʃ$, suggesting it was originated by the combination of t and $ʃ$.
- Prenasalized or palatalized consonants. These seem to come from assimilation of intermediate consonants, although further study is needed to assert it generally.

The second strategy is to add more phonemes per syllable.

- Consonant clusters, or sequences of consonants without vowels. It is possible that the basic *KAK* scheme was supplemented at this point by adding consonants at the beginning or the end, using the same mechanism of compounding that we saw for two-consonant words. These new syllable structures *KKAK* or *KAKK* would yield 5184 combinations each. But as we saw above, there is evidence of languages keeping the old two-consonant words at a later phase, so these structures would be anachronical at this point.
- Diphthongs allow combinations of vowels in a single syllable. Again, probably come from collapsing two separate syllables, in this case with an intermediate aspirant consonant, later lost.

The last strategy is to add modifiers for vowels and consonants.

- Add different contours, as in Chinese (sample). They are a special kind of tone that may vary within a syllable. As Michaud (2012) refers, tones evolved from consonant depletion.
- Add nasalized vowels as in French (sample). As Hombert (1986) explains, nasalized vowels come from assimilation of a nasal consonant into its preceding (or occasionally anteceding) vowel.

All of these innovations, if originated from polysyllabic words as seems likely, would not have increase the range of possible words but just have followed the natural paths of languages.

It is interesting to note that tonal languages have developed mostly in Africa and Southern Asia, while more complex syllable structures are more abundant in Europe and North America. But it does not seem to bear any effect on prehistoric languages.

Language As A Tool

Language has many uses. There can be no doubt that it has gained utility with time.

Information Content

Language becomes more useful as it becomes more expressive. A very basic criteria that describes a language era is how much information can be conveyed in a single sentence, something which can luckily be estimated using information theory.

First we will compute the information carried by a single word. Using the classic Shannon formula, see e.g. Lombardi (2016):

$$I = -\sum p_i \log_2(p_i)$$

If there are N words, all equally likely, the resulting information is

$$I = \log_2(N)$$

But we can do better using Zipf’s law, as characterized in Yang (2013): it is the empirical distribution followed by all known languages. We can suppose for simplicity that prehistoric languages also followed this distribution. The resulting information content in bits per word, generated by this script, is shown in the table 1 below.

Vocabulary size	Bits per word
12	3.1
100	5.3
200	6.0
400	6.6
1000	7.5
10k	9.5
20k	10.1

Table 1: Bits per word depending on vocabulary size. Prepared by the author.

Adding more possibilities to the vocabulary increases the information carried by each word. Another way to convey more information is to combine words together. The information content of a sentence depends on the number of words that can be put together per sentence.

We can estimate the average number of words per sentence using MLU, Mean Length of Utterance. Kemper & Sumner (2001) estimate that young adults have a MLU of around 12 ± 4 . With a vocabulary of around 20k words following a Zipf distribution, each word would carry around 10 bits of information, as we saw above. Therefore a typical adult sentence might carry 120 bits of information, if all words were independent.

What happens with smaller vocabularies? Devescovi *et al* (2005) estimate that children with a vocabulary of 101-200 words have a MLU of around 3.5 words, as shown in figure 2. 200 words following a Zipf distribution would carry around 6 bits per word, yielding a total of 21 bits per sentence. Similarly, 400 words would have a MLU of around 4 words, which at 6.5 bits per word would yield 26 bits per sentence. 1000 words are outside their range of study, but a MLU of 10 words seems a reasonable extrapolation; at 7.5 bits per word each sentence would reach 75 bits. 100 words correspond to a MLU of 3 and 5.3 bits per word, or 16 bits per sentence. On the lower range, a vocabulary of 12 words might have a MLU of 2, which combined with 3 bits per word yields 6 bits per sentence.

Data from both studies is not necessarily compatible, as they use different definitions for MLU. Yet it is reasonable to expect that after a point, increasing

Figure 3: Figure 2: Mean Length Of Utterance In Children. From Devescovi *et al* (2005).

the vocabulary does not entail a corresponding increase in sentence size. Table 2 summarizes these estimations.

Vocabulary size	MLU	Bits per word	Bits per sentence
12	2	3	6
100	3	5.3	16
200	3.5	6	21
400	4	6.5	26
1000	10	7.5	75
20k	12	10	120

Table 2: Bits per sentence depending on vocabulary size. Prepared by the author.

This gross approximation is enough to show that that increasing vocabulary size initially yields large advances in communication, and then tapers off. After around 1000 words gains are incrementally smaller. In fact, information content would grow even more slowly with vocabulary size when considering redundancy in sentences. In general words in a sentence are not independent: ‘house’ is much more likely to follow ‘the’ than ‘far’. Particles and other function words are highly redundant. The values above can only be considered bounding limits. The actual number of bits per sentence is thus much lower than represented, especially as the vocabulary size grows.

We will now delve into how different uses for language are enabled by increasing information content.

Coordination

The most primitive uses of language are related to the present situation: they are in fact deictic, as we saw above. Even with only a handful of function words and nothing else, a language that allows combinations of words can be used to guide and coordinate group activities. A group of hominins may use this limited set of words to *remove ambiguity* from their common undertakings.

During group hunting, a scout can pinpoint the location and amount of prey with gestures and a few words, much more effectively than with the gestures alone. Not much information is needed to express a direction, a relative distance (far or near) and quantity (whether there is much or little prey). Similarly, a group leader can direct the movements of team members with a couple of indications for each.

We will see below a sample language that can fulfill a very basic version of this coordination role with only 12 words.

Description

Still in the immediate present, a more advanced language with content words can be used more effectively to *remove ambiguity* about the current situation. Say that two hominins *A* and *B* are in a certain setting. *A* sees an animal, and wants to communicate it to *B*. Perhaps their common language has an utterance for ‘warning’, not necessarily a word: it may be a simpler sound like a shout, probably combined with a prosodic mark such as an intonation. It will be *A*’s first utterance. Combined with a pointing hand it may be enough to make *B* look and appraise the situation on their own. But then *A* may want to give *B* a more precise idea of the situation, even if *B* is not currently able to see the animal: let us suppose it just fled, or maybe *B* is far away. At some other situation there may be a group of listeners: a shout that lets them know there is a bear nearby will be much more effective than waiting for them to look for it. Their reaction if there is a rabbit instead should be equally swift but completely different.

If their language only has a single word for all animals *A* will utter ‘animal’, which is not very precise: there is a lot of ambiguity. It could be a wolf or a sheep, and both situations will probably call for very different actions. Perhaps there will be words for ‘big’ and ‘small’, so some of the ambiguity can be removed by saying ‘animal small’ or ‘animal big’. But a small wolf and a big sheep still call for different actions. A language with different words for ‘wolf’ and ‘sheep’ will have a distinct advantage for its speakers. Given the number of different animal species, a large vocabulary is required to accurately represent even the most common animal families (i.e. a single word for ‘wolf’, ‘dog’ and ‘fox’). Increasing the number of possible words is thus crucial to remove all ambiguity.

More information can be added by adding a plural form: ‘one wolf’ is a very different situation from ‘wolves’ for our hypothetical hominins. Again, this requires increasing the number of possible words.

Training

Another important use is training. Here too language can help to *remove ambiguity*. Complex techniques require very precise movements, be they used for tool making, gathering food or hunting game. Language can be a powerful teaching aid.

Composite materials used for hafting stones to handles require precise combinations of materials, each in turn gathered from the correct environment. Wadley *et al* (2009) explain how red ochre and tree gum must be mixed in precise proportions and heated near a fire at between 250 °C and 300 °C to obtain a composite glue for hafting stone points to wooden handles. There is only one species *Acacia karroo* which is ideally suited for the process, and the proportion of ochre must be adjusted depending on its iron content. This complex process

carries a lot of ambiguity that has to be removed in the teaching process; it is easy to imagine how training time would have been reduced substantially with verbal aids guiding the process. Given that there are archeological samples of such composite materials from at least 70 kya it seems that language must have been quite advanced by then.

Telling Stories

A more elaborate use of language is storytelling. In this setting the speakers are recounting previous events. Here too *removing ambiguity* is essential.

Storytelling is useful as training: someone listening to hunting stories may learn tips and tricks. But it goes much deeper than that. Stories carry with them mythology, world view, family tree, ethics; in short, the essentials of human culture. A complex language and vocabulary allow for better transmission of information among groups and between generations. Groups with strong oral culture have an advantage over those without, especially in a crisis as described by Minc (1985).

The fact that oral traditions have existed for so long, and even now are still maintained by certain groups, points to some innate capability in humans. Our fascination with hearing stories is present today thinly disguised as books or movies. We still get most of our morals in the form of exemplifying stories.

Stories are possible with a limited set of words: Dr Seuss' The Cat in the Hat contains only 236 different words, and Green Eggs and Ham is limited to 50. But these were stories tailored to suit the limited number of words. Experiments like Basic English with a limited vocabulary (850 words) and simplified grammar rules have had some success among foreign speakers. It would be enlightening to learn if a complex story can be told with an even more limited vocabulary.

Sources Of Evidence

Language does not fossilize. This well-known fact means that all of our knowledge about how language evolved before the invention of writing is pure speculation. Or is it? As a matter of fact, there are several sources of evidence about prehistoric languages.

Physical Remains

Archeological remains are the first and foremost source of evidence about all things prehistoric. We know the lifestyle and capabilities of our ancestors thanks to the remains that they left behind.

We have seen above how hyoid bones seem to imply that hominins could already articulate 530 kya (and no sooner than 3.3 Mya). More samples might help us pinpoint these dates more exactly.

There is also symbolic evidence which might prove even more valuable. Von Petzinger (2009) has found that the same set of symbols is repeated in Upper Paleolithic cave paintings across much of France, and in later publications around the world. Figure 3 shows the latest 32 symbols recurrent across Europe.

There are also cup and ring marks from later periods, and probably many more overlooked marks. Do they represent concepts of some kind, or are they just pretty designs? Are marks more likely to represent functions words, or are they content words? With a more concrete language description in mind it is conceivable that these inscriptions might be decyphered, or at least put in context.

Modern Languages

Today there are about 7,099 languages. spoken around the world. It is to be expected that any features before these languages separated would be shared.

Watching the distribution of certain features can be very instructive. The World Atlas of Language Structures Online (WALS) is a great resource organized by language features. It is entirely possible that the distribution of some features is the direct result of prehistoric migrations.

Language Reconstructions

Indoeuropean is the most widely known and studied archaic language, but many others have been reconstructed. A few protolanguages (like Proto-Indo-European or PIE) have also been postulated with varied success.

Rate of linguistic change is hard to estimate. Nettle (1999) argues that languages change faster in small populations. It remains to be seen how ancient languages can be reconstructed to such a high degree of precision if they were spoken by relatively few people for many centuries.

Reconstructions rarely go back beyond the Neolithic revolution, around 12 kya. Except glorious exceptions like the study by García Calvo (2003), they are of limited value to infer how or when language features evolved. It remains to be seen if the veils of time can be pierced any further.

Genetics

Pinker (2003, pp. 297-331) makes a quite convincing case that language must be enabled not by one but by several different genes. FOXP2 is the most

The Geometric Signs of Ice Age Europe

Figure 4: Figure 3: 32 recurrent symbols across Upper Paleolithic Europe. From von Petzinger (2017), used with permission.¹⁹

famous and studied of all. This gene has the same expression in humans and Neanderthals, from which Dediu *et al* (2013) deduce that both groups had mostly equivalent language capabilities. It is likely that we will be able to trace evolution of speech even further back in time by following how this and other genes have evolved.

Another important area of study is speech disorders. The best way to learn how a machine works is often watching how it malfunctions. Similarly, it is now possible to study how defects in individual genes affect individuals that carry them. So e.g. if a gene that enables polysyllabic speech can be identified, it should be possible to study how people without it (or with a non-functional variant) can communicate. For instance, Vargha-Khadem *et al* (1995) study the implications of a genetic problem in FOXP2 on a large family. Would it be possible to find variants of this and other genes that limit for instance polysyllabic communication?

Genetics may thus serve as empirical test of the timeline that has been sketched in this article.

Previous Studies

While not direct evidence, there is a growing amount of literature about the origins of language. Smillie (1985) models the evolution of early languages on the acquisition of language by children. Hurford (2003) offers a limited view of proto-languages, where at the beginning they would have no function words. In the same volume, Newmeyer (2003) gives an overview of other theories. Bickerton (2016, p. 200) also postulates generic names as the first words, and proposes that protolanguages were similar to pidgin languages with very loose rules of grammar. Pinker (2003, pp. 352-368) favors gradual evolution in speech development, but without giving concrete stages or time frames. He notes that all humans have “identical language abilities”, so that this evolution must have happened before the big out-of-Africa diaspora.

It is my sincere hope that I have overlooked a mountain of literature specifying how languages came to be the complex wonders that we know, feature by feature and with even more fantastical timelines. If it does not exist, it definitely should.

Conclusion

Unraveling the origins of language is a particularly hard task, which was once famously banned by the Linguistic Society of Paris. In this article we have seen a possible evolution path that takes us from animal sounds to the written word. We have also explored briefly how it fits with known evidence and how to refine or refute it with more research.

The whole endeavour may seem idle speculation. After all it may seem impossible to refute most of what has been said here. But it is my belief that there is fruitful research to be done. A theory only needs to be consistent with the known facts, and it is partially validated when it can explain new phenomena.

Learned discussion about very early phonetics, phonotactics and grammar may take us a long way towards a deeper understanding of our ancestors.

Appendix

It becomes much easier to visualize language types with an example. In the appendix we will explore several hypothetical languages to illustrate the different eras of language evolution.

Singlecon

The first language has just words that consist of a single consonant. We will call this hypothetical primitive language ‘Singlecon’.

Vocabulary

In Singlecon we only have 12 words, that correspond to the 12 consonants:

- one preposition: ‘to’ (‘R’),
- one negation: ‘no’ (‘P’),
- two non-numeral quantifiers: ‘much’ (‘K’) and ‘little’ (‘D’),
- two numerals that double as pronouns: ‘one’ or ‘me’ (‘V’) and ‘two’ or ‘you’ (‘N’),
- and six adverbs: ‘down’ (‘L’), ‘up’ (‘S’), ‘here’ (‘M’), ‘near’ (‘T’), ‘out’ (‘H’) and ‘far’ (‘F’).

Note that the language consists only of function words. There are no content words: nouns, verbs or adjectives.

Samples

A few sentences in Singlecon follow. They show the value of combining words for a group of foragers or hunters.

- ‘T L.’ means ‘near down.’: the objective is near but they need to go down.
- ‘K K.’ means ‘much much.’: there is a lot of what the speaker is interested in.
- ‘V M. N H.’ means ‘one here. two out.’: the speaker will remain here while the listener goes away.
- ‘V R. N M.’ means ‘one to. two here.’: the speaker will go to the objective; the listener must stay here.

- ‘K S. D M.’ means ‘much up. little here.’: there is a large quantity of what the speaker is seeking going up, while there is little to be found here.

Grammar

Even with such a simple language a grammar would be needed, to convey different meanings depending on the combination. For instance, ‘near much’ would mean that the objective is very near, while ‘much near’ would mean that there is a lot of what the speaker is seeking nearby.

Utility

As the examples show, Singlecon can be used by its hypothetical speakers to coordinate during a group hunt, or to guide their gathering activities. In the uses we saw above, it would serve for coordination for very limited description.

The astute reader will realize that Singlecon almost passes the ten word test in Bickerton (2009, p. 24): it is useful with just a dozen words.

Doublecon

Our next language is an evolution of Singlecon that allows for monosyllabic words with two consonants and a schwa, represented as $K\text{ə}K$. It keeps the same set of single-consonant function words, while the new combinations are all content words. We will call it Doublecon.

Vocabulary

Doublecon has a maximum of 144 content words, that can work as nouns, verbs or adjectives depending on the situation. Content words are formed with two consonants by reduplication and compounding of function words.

- The combination of ‘down’ (‘L’) and ‘to’ (‘R’) is ‘LəR’ and refers to something that moves on the ground such as a small animal.
- ‘up’ (‘S’) + ‘to’ (‘R’) is ‘SəR’: a big animal.
- ‘near’ (‘T’) + ‘much’ (‘K’), ‘TəK’: family.
- ‘near’ (‘T’) combined with itself is ‘TəT’ which means ‘good’.
- ‘to’ + ‘down’, ‘RəL’: trail.
- ‘to’ + ‘near’, ‘RəT’: run.

Not all combinations are used by actual words; of the 144 possibilities for $K\text{ə}K$ only about 100 are valid.

Even when words are formed as compounds there is considerable arbitrariness in the process. The combination of ‘down’ with ‘to’ (‘LəR’) is small animal, while ‘to’ + ‘down’ (‘RəL’) is trail. It might perfectly be the other way around, so it is purely a matter of convention.

Prosody

There are three possible sentence intonations, indicated here by the sentence ending sign:

- neutral, marked with a dot ‘.’,
- interrogative: marked with an interrogation mark ‘?’,
- and exclamative: marked with an exclamation mark ‘!’.

There are no other possible intonations such as our comma ‘,’. Sentences may follow each other closely, with a pause between them.

Samples

A few sentences in Doublecon follow.

- ‘TəK M. TəK TəT.’: ‘family here. family good’, the good family has arrived.
- ‘V R SəR. N M.’: ‘me to big animal. you here.’, I will go to the big animals, you stay here.
- ‘SəR K S.’: ‘big animal much up’, there is much big game going up.

Note that it is very important to distinguish between two single-consonant words and one syllable: ‘S R’ from ‘SəR’. In the first case there is a small pause between ‘S’ and ‘R’, similar to the intersyllabic transition in modern languages, as in English ‘pos · ter’. In the case of ‘SəR’ the schwa provides vocal continuity.

Grammar

Even at this level, a simplistic grammar is needed to combine content and function words.

The copula is straightforward juxtaposition: ‘V TəK’ means ‘I am family’.

Some words can be both nouns or verbs, such as ‘RəT’: run, or ‘RəL’: trail. Nouns are always the first word, followed by verbs or adjectives: ‘SəR RəT’ means ‘big animal runs’. Another possibility is juxtaposition of two nouns: ‘RəL SəR’ means ‘trail of a big animal’. Verbs can also appear on their own or with function words: ‘RəT K!’ means ‘run a lot!’ (remember that ‘K’ means ‘much’), and so with adjectives: ‘TəT K’ means ‘very good’.

Utility

With just about 100 words Doublecon has many more uses than Singlecon, as can be gathered from the examples we have seen. In the list of uses above it would serve for coordination and description, and for very basic training. It might start to be used for story telling, although it would be still unsatisfactory.

References

- Agmon, N. (2010). Materials and Language: Pre-Semitic Root Structure Change Concomitant with Transition to Agriculture. *Brill's Journal of Afroasiatic Languages and Linguistics*, 2(1), 23-79. PDF.
- Bakker, E. J., de Jong, I. J., & van Wees, H. (Eds.). (2002). *Brill's companion to Herodotus*. Brill. Book
- Bar-Yosef, O. (2002). The Upper Paleolithic revolution. *Annual Review of Anthropology*, 31(1), 363-393. PDF.
- Baxter, W. H., & Sagart, L. (2014). *Old Chinese: a new reconstruction*, 318. Oxford University Press. Book.
- Bickerton, D. (2009). *Adam's tongue: how humans made language, how language made humans*. Hill and Wang.
- Bickerton, D. (2016). *Roots of language*. Language Science Press. PDF.
- Clair, R. N. S. (1972). Compound phonological segments. *Lingua*, 29, 120-127. Web.
- Dediu, D., & Levinson, S. C. (2013). On the antiquity of language: the reinterpretation of Neandertal linguistic capacities and its consequences. *Frontiers in psychology*, 4. Web.
- Devescovi, A., Caselli, M. C., Marchione, D., Pasqualetti, P., Reilly, J., & Bates, E. (2005). A crosslinguistic study of the relationship between grammar and lexical development. *Journal of Child Language*, 32(4), 759-786. PDF.
- D'Anastasio, R., Wroe, S., Tuniz, C., Mancini, L., Cesana, D. T., Dreossi, D., ... & Capasso, L. (2013). Micro-biomechanics of the Kebara 2 hyoid and its implications for speech in Neanderthals. *PloS one*, 8(12), e82261. Web.
- García Calvo, A (2003). ~Es~ Estudio de Gramática Prehistórica (Spanish). Ed. Lucina. Ref.
- Groucutt, H. S., Petraglia, M. D., Bailey, G., Scerri, E. M., Parton, A., Clark-Balzan, L., ... & Breeze, P. S. (2015). Rethinking the dispersal of Homo sapiens out of Africa. *Evolutionary Anthropology: Issues, News, and Reviews*, 24(4), 149-164. PDF.
- Hombert, J. M. (1986). The development of nasalized vowels in the Teke language group (Bantu). *The phonological representation of suprasegmentals: Studies on African languages offered to John M. Stewart on his 60th birthday*, 359-379. PDF.
- Hurford, J. R. (2003). The language mosaic and its evolution. *Studies in the Evolution of Language*, 3, 38-57. Web.
- Kemper, S. & Sumner, A. (2001). The structure of verbal abilities in young and older adults. *Psychology and Aging*, 16, 312-322. Ref, PDF.
- Lama, Z. Q. F. (2012). Subgrouping of Nisoic (Yi) languages: A study from the perspectives of shared innovation and phylogenetic estimation. PDF.
- Lombardi, O., Holik, F., & Vanni, L. (2016). What is Shannon informa-

- tion? *Synthese*, 193(7), 1983-2012. PDF.
- Luraghi, S. (2011). The origin of the Proto-Indo-European gender system: Typological considerations. *Folia linguistica*, 45(2), 435-463. PDF.
 - Martínez, I., Arsuaga, J. L., Quam, R., Carretero, J. M., Gracia, A., & Rodríguez, L. (2008). Human hyoid bones from the middle Pleistocene site of the Sima de los Huesos (Sierra de Atapuerca, Spain). *Journal of human evolution*, 54(1), 118-124. PDF.
 - Martínez, I., Rosa, M., Quam, R., Jarabo, P., Lorenzo, C., Bonmatí, A., ... & Arsuaga, J. L. (2013). Communicative capacities in Middle Pleistocene humans from the Sierra de Atapuerca in Spain. *Quaternary International*, 295, 94-101. PDF.
 - Michaud, A. (2012). Monosyllabicization: patterns of evolution in Asian languages. *Monosyllables: from phonology to typology*, 115-130. PDF.
 - Minc, L. D. (1986). Scarcity and survival: the role of oral tradition in mediating subsistence crises. *Journal of Anthropological Archaeology*, 5(1), 39-113. PDF.
 - Nettle, D. (1999). Is the rate of linguistic change constant? *Lingua*, 108(2-3), 119-136. PDF.
 - Newmeyer, F. J. (2003). What can the field of linguistics tell us about the origins of language?. *Studies In The Evolution Of Language*, 3, 58-76. Book.
 - Pinker, S. (2003). *The language instinct: How the mind creates language*. Penguin UK. Book.
 - Roach, N. T., Hatala, K. G., Ostrofsky, K. R., Villmoare, B., Reeves, J. S., Du, A., ... & Richmond, B. G. (2016). Pleistocene footprints show intensive use of lake margin habitats by *Homo erectus* groups. *Scientific reports*, 6. Web.
 - Smillie, D. (1985). Sociobiology and human culture. In *Sociobiology and epistemology* (pp. 75-95). Springer Netherlands. Book.
 - Tomasello, M. (2008). Why don't apes point?. *Trends In Linguistics Studies And Monographs*, 197, 375. PDF.
 - Trask, R. L. (2003). *Language: the basics*, 5. Routledge. Book.
 - Vargha-Khadem, F., Watkins, K., Alcock, K., Fletcher, P., & Passingham, R. (1995). Praxic and nonverbal cognitive deficits in a large family with a genetically transmitted speech and language disorder. *Proceedings of the National Academy of Sciences*, 92(3), 930-933. PDF.
 - Von Petzinger, G. (2009). *Making the abstract concrete: The place of geometric signs in French Upper Paleolithic parietal art* (Doctoral dissertation). PDF.
 - Von Petzinger, G. (2017). *The First Signs: Unlocking the Mysteries of the World's Oldest Symbols*. Simon and Schuster. Book.
 - Wadley, L., Hodgskiss, T., & Grant, M. (2009). Implications for complex cognition from the hafting of tools with compound adhesives in the Middle Stone Age, South Africa. *Proceedings of the National Academy of Sciences*, 106(24), 9590-9594. Web.
 - Yang, C. (2013). Ontogeny and phylogeny of language. *Proceedings of the*

National Academy of Sciences, 110(16), 6324-6327. Web.