PHYLOGENY OF LANGUAGE, MIGRATIONS OUT OF AFRICA
AND LANGUAGE CLASSIFICATION

JACQUES COULARDEAU¹

ABSTRACT. Phylogeny of Language, Migrations out of Africa, and Language Classification. From Guillaume’s 1958-1960 lectures on his three area theory I will move to Teyssier’s language phylogeny and his three articulations, then to the migrations out of Africa, the three regions they occupied and the three families of languages they spoke: Semitic, isolating and agglutinative / synthetic-analytical languages, respectively built on the first, second and third phylogenic articulations. Each migration must have left Africa when language there had reached the concerned articulation. We have to re-evaluate Guillaume’s theory accordingly.

Keywords. Out of Africa, Semitic, Isolating, Turkic, Synthetic/Analytical


Cuvinte cheie: În afara Africii, Semitic, Izolant, turcic, sintetic/analitic

Gustave Guillaume was an inspired genial discoverer and his theory of the three areas is essential to understand the invention of language by Homo Sapiens, hence the phylogeny of language. This theory is clearly explored in his last years of teaching (1958-1960) and the concerned lessons have been published in volume 13 of the said lessons.

¹ Contract-holding lecturer, Synopsis Paie, Nice; has taught at University of California Davis, six Paris Universities, University of Lille 3 and University of Perpignan; two PhDs in Germanic Linguistics and Foreign Language Didactics; Assistant Ivan EVE, Graduate student University Paris 1 Panthéon Sorbonne. E-mail: dondaine@orange.fr
The three areas are:

1. **The primary area** (p. 288): what he calls “Nostratic” languages. All agglutinative languages and all isolating languages, in fact all languages that are neither Hamito-Semitic nor Indo-European;

2. **The secondary area** (p. 288): what he calls the Hamito-Semitic languages;


Since 1960 the fields of archaeology, anthropology and linguistic phylogeny have evolved at a tremendous speed and they have been accelerating exponentially over the last fifteen or twenty years with the new dating technologies based on carbon and other elements. It is obvious we cannot state any shortcoming in Guillaume’s approach in 1960 in the light of what he could not know in his time. But today we have to consider the real state of science in these fields to reexamine the question of the invention and development of articulated language by Homo Sapiens, the migrations of Homo Sapiens out of Africa and the production of the three vast families of existing languages by the phylogeny of language and these migrations. We are not speaking of Greenberg’s dubitatively debated hypothesis about the unique African origin of language because we today know that the whole humanity has come from one single nest in Africa even if it has to be seen as slightly vaster than just the territory covering Ethiopia; Sudan and Kenya, be it only because an early migration to southern Africa came first and because we can consider the sub-Sahara and sub-Sahel savanna area (known as the Sudan running from the Indian Ocean to the Atlantic Ocean) as the natural expansion area of Homo Sapiens. The evidence can be found in the fact that all humans today carry some genetic heritage from Neanderthals who originated in the Middle East (and Denisovans who originated in Central Asia) except those who never left Africa or never had genetic contact with non-African humans mostly in post-Ice-Age times. It is agreed that the peak of the last Ice Age was around 20,000 BCE and that it started receding around 15,000 BCE with a short relapse known as the Younger Dryas or Big Freeze from 10,800 to 9,500 BCE.

It is high time we, as linguists and psychomechanicians, examine what we know on the subject of the migrations of Homo Sapiens out of Africa, reexamine what we can know on the phylogeny of language and try to set the two side by side and see what hypotheses can be brought up to light.

But once again and before getting into the main topic, let me repeat that the ternary approach of Guillaume was creative and inspiring in its ternary nature itself. He dared challenge André Martinet’s binary approach (the double articulation theory) that was being turned into a dogma by Noam Chomsky as soon as *Syntactic Structures* (1957) was published.
I. Phylogeny of language and the three articulations

I borrow the concept of three articulations from Jacques Teyssier’s teaching in 1972-1973 that he had devised for his global approach of Germanic languages and he had taught for example in his last Master’s lectures published post-mortem by Rule Syntactica. I here expand it to language in general. I also borrow the double tree metaphor or roots and stems and expand it into a triple metaphor of roots, stems and fronds.

The first articulation is that of the ROOT. It articulates vowels and consonants. This is the first step Homo Sapiens took on the long road to the invention of articulated language. This is possible because Homo Sapiens is the first hominin species\(^2\) that can produce many vowels and a lot of consonants because of the mutations that he went through in order to become the bipedal long distance fast runner he had to become when he came out of the forest and evolved in the savanna. This genetic developmental phase requires a deep larynx as the breathing pump, great laryngeal and articulatory coordination with extensive innervation and the development of the Broca zone of the brain as the coordinating center of complex activities. We must keep in mind bipedal long distance fast running requires the coordination of the diaphragm, the lungs, the breathing apparatus (larynx, glottis, tongue and mouth, plus the nasal cavity), the heart, and of course all the limbs as well as the whole body for inclination and orientation. Homo Sapiens thus became able to produce a great number of calls when he discovered and developed after more or less non-motivated experimentation in rotating the vowels and the consonants to produce many clusters. He was furthermore able to attach to each one of them some referential meaning thus producing Meaningful Vocal Units (MVU). This last step was possible because of his brain development (Neanderthals had an 11% deficit in brain EQ mass\(^3\)) that enabled him to conceptualize in a six step procedure:

1. to sense
2. to perceive
3. to identify or recognize, hence to name (first articulation MVUs)
4. to experiment
5. to speculate
6. to conceptualize

---

\(^2\) Hominid – the group consisting of all modern and extinct Great Apes (that is, modern humans, chimpanzees, gorillas and orangutans plus all their immediate ancestors). Hominin – the group consisting of modern humans, extinct human species and all our immediate ancestors (including members of the genera Homo, Australopithecus, Paranthropus and Ardipithecus).


\(^3\) encephalization quotient: "brain mass against body mass.”

The first two are borrowed from Bertrand Russell. The first three are borrowed from Jeff Hawkins and the last three are borrowed from L.S. Vygotsky. I will insist here on what Guillaume repeats several times in his last lessons: language and the mind develop simultaneously, dialectically, each one causing the development of the other by its own development caused by the development of the other. This idea was borrowed from the philosopher Henri Delacroix. At the level of the first articulation the only possibility is the concatenation of semantically meaningful units that are neither nominal nor verbal, hence a purely semantic concatenation finding some communicational meaning in the communicational situation and the use of communicational means like body language and supra-segmental expressive means.

Languages can develop on this basis only by using the vowels as the discursive means to build a full discourse from consonantal roots, thus endowing each root with all possible nominal or verbal elements needed in discourse to build a full communicational utterance. Consonants are the basis of the roots because they cannot be uttered without vowels and vowels are the discursive means because they provide consonants with some rather opportunistic, I mean non-semantically-motivated, articulatory power. This will bring the question of the communicational matrix later on. Such languages are Semitic languages in which the vowels are used to provide the roots in discourse with nominal and verbal categories and all types of specifications (gender, number, extension, etc; tense, mode, aspect, etc).

The second articulation is that of the STEM. A stem is a categorized MVU. When it is spatially categorized it becomes a noun and when it is temporally categorized it becomes a verb. These stems are invariable and they can be associated as complementary units giving rise to a syncrctic categorial architecture. Then in discourse such languages can build full utterances on the basis of the communicational matrix we will deal with later. Such languages are isolating languages (which is a better name than character languages since many of these languages do not use characters as their writing system since characters are only a writing characteristic: we must keep in mind writing has only existed for some 5,000 years whereas language has been developing over more than 250,000 years). Note this syncrctic categorial architecture is the first level of syntax because syntax requires these categories which are connected to space and time. We are dealing here with the first stage of syntax integrated in langue and not provided by discourse.

The third articulation is that of the FROND. The nominal and verbal stems are endowed with all syntactic marks. The nominal stems are thus endowed with gender, number, extension and functional case. The verbal stems are endowed with tense, mode, aspect and diathesis. The two can then build a rich syntactic architecture. There are two levels in this third articulation. First the whole syntax is carried by both the nouns and the verbs with agreement.
rules between the two. This type is known as agglutinative languages (holophrastic languages, which are extremely rare, are one particular type of agglutinative languages). These languages are the vast family of agglutinative languages also known as Turkic languages. This family contains many subfamilies.

Then at a second level the various marks are externalized, what is called “déflectivité” [deflectionization4] in psychomechanics, meaning that some of the flections nouns and verbs carry can be extracted from the nouns and the verbs and expressed outside either by special tool-words or by order-rules. All determiners and prepositions are such deflectionizing elements for nouns, and auxiliaries are the same for verbs. Guillaume insists on the evolution of these languages towards a synaptic case for the noun, a case that is no longer marked on nouns and that is reduced to direct and oblique for pronouns though then in the utterance the place or the use of prepositions will indicate the particular case the noun carries as indicated by these deflectionizing elements that are langue elements because prepositions or determiners are the marks of langue systems. These languages are essentially two types generally covered by one name, Indo-European, though there are two subfamilies from one parent language, the Indo-European and the Indo-Aryan subfamilies of what we could call the Iranian family whose modern descendant is Farsi.

I generally call these two third-articulation families, FROND1 for the agglutinative family and FROND2 for the Indo-European, Indo-Aryan and Iranian family. They are synthetic-analytical languages covering a range from synthetic languages in which nouns and verbs still carry various flections without being agglutinative to analytical languages that have deflectionized some or maybe most if not all yet abstract operations and langue systems.

We can see this third articulation case endows fronds with a maximum of langue syntax, reducing discourse syntax to a minimum. I will come to the communicational matrix later on.

II. Archaeology and the migrations out of Africa

I will not consider here the two migrations inside Africa. First the migration to Southern Africa that will produce click languages: clicks are nothing but consonants of a certain type and all languages or communicational exchanges use clicks now and then even if they are not consonants. One common click in English or French is generally spelled “tss” or “tss-tss.”5 And

---


second the migration known as the Bantu migration that will take place later after the Ice Age mostly. Note Fula language(s) as well as Dogon language(s) are of different origins.

I follow the most advanced archaeological reconstructions of Homo Sapiens’ emergence like Sally McBrearty’s and I take a time line that starts around 250,000 BCE if not 300,000 BCE. I do not recognize the appellation “Homo Sapiens Sapiens” that is based on a fake genetic mutation that would have produced the ability to write, since all ancient Mesoamerican civilizations (Maya, Aztec, etc) who had rich and complex writing systems did not possess the mutation that was supposed to produce the ability to write, and I will add that many peoples who had the mutation did not invent writing that remained a minority art because many civilizations did not practice it and remained till a very recent period totally oral, and writing was introduced in these civilizations by Europeans mostly. We are the direct descendents of the Homo Sapiens that emerged from Homo Faber (or Homo Ergaster) in Africa sometime between 300,000 and 250,000 BCE. The human species has not basically changed in its genetic affiliation since then and the development of the mind, the intellectual capacity, language, culture, religion, science and philosophy is the result of the mutations that brought to Homo Sapiens the bipedal long distance fast running competence as I have already said.

a-c. 150,000 BCE. This first migration went down the Nile river to Egypt and then west along the Mediterranean coast to Northern Africa and the Sahara, and with an eastern temporary option to the Levant up to 80,000 BCE when they come back to Egypt to only move again to the Levant around 35,000 BCE. This migration will reach Crete around 130,000 BCE. We can note that all these areas speak Semitic languages.

---

b- C. 100,000 BCE. This second migration crossed the red Sea from the Horn of Africa and followed the southern Arabian corridor, a coastal route along the Yemen-Oman littorals to cross the Persian Gulf. From there they moved north and met with the Denisovans which endowed them with the gene necessary for them to live in the Himalayas. They expanded in the whole of eastern Asia as far as Japan, Korea, South East Asia and of course Siberia, Tibet and the Himalayas, Myanmar and Thailand. The Indian subcontinent is debated. The older population (before the Indo-Aryans who will arrive after the Ice Age) is either from this migration and Dravidian languages would be of the same isolating family or from the next one and then Dravidian languages would be agglutinative.

All these areas speak isolating languages of different types. The use of characters to write these languages is a Chinese invention and does not concern most of these languages who have been using many different writing systems most of them syllabic or alphabetical. Note these languages are not necessarily monosyllabic in their words.

c- Starting c. 75,000 BCE. The first wave of the third migration crossed the same way as the previous one. Then from Pakistan they moved up into Central Asia as far as the Urals and Siberia, and from the Urals they reached Finland and Lapland (or Laponia).

---

They also moved west under the Caspian sea, then to Anatolia and through the Caucasus, and from there to the whole of Europe they reached around 45,000 BCE. These languages are agglutinative and can be known under the generic term of Turkic. That's where the debate about Dravidian languages is possible if we consider Dravidian language not as isolating languages but as agglutinative languages, following in that Professor Doctor Alfred Toth. Although I personally finds his arguments on agglutinative languages very weak since many of the criteria he uses are not agglutinative at all.

The second wave of this migration came somewhat later c. 50,000 BCE along the same route though they went up the Persian Gulf and established themselves on the Iranian plateau where they stayed a long time to come down from it, but only after the Ice Age hence after 15,000 BCE, east to Pakistan, Afghanistan and beyond India where they pushed the Dravidians south, on one hand, and west on the other hand into Mesopotamia and from there to Anatolia. Note though the Greek migration did not come from the Indo-European Hittites who were in Anatolia because of great differences between Greek and Hittite, though Greek is closely related to Indo-European Armenian that is to the east of the Hittite empire. Then from Greece they moved west to the whole Mediterranean in competition with the Semitic Phoenicians. From this migration came the Roman Empire. But from the Middle East this migration also crossed the Caucasus and spread into the whole of Europe. These languages are Indo-Aryan languages in the east and Indo-European languages in the west.

---

III. Crossing linguistic phylogeny and archaeology.

We have practically done the whole work already, but we have to state an important principle here. Migrating groups are always linguistically conservative in the fact they keep the level of development they have reached when they migrate and they develop linguistically on the basis of this level of development. This is true of all migrating people and still today.

Then if the first migration produced Semitic languages it must correspond to the first articulation that is the only one capable of producing these Semitic languages. That implies Homo Sapiens had reached that first articulation somewhere around 170,000 BCE and had stabilized at least for a while at this level, so that the migrating people left then and developed their languages from this acquired articulation. They invented or Homo Sapiens in the nest had already developed the discursive use of vowels to specify the consonantal roots (remember consonants cannot be pronounced without vowels) in discursive categories and various syntactic functions to build the discourse. But then where does the syntax of this discourse come from since it can't come from the langue of the languages? We are going to see in a moment what I call the communicational matrix. Note these languages develop some syntactic tools that are embedded in langue as systems and they integrate in langue, in the form of systems, the syntactic elements they convey. Note the
use of the vowels can be architecturally regular and thus it can build langue systems both for nominal elements and for verbal elements. But the matrix is outside. These elements are not deflectionized but they are in a way flectionized into the langue of these languages. Discursive means are thus reconstructed into some kind of langue syntactic systems.

If the second migration corresponds to isolating languages that implies these people left the African nest when the phylogeny of language had reached the second articulation. They constructed then their languages on the basis of invariable categorized nominal and verbal elements. Here too they may have created syntactic tools in their discourse and then integrated these tools in langue by building various systems. Here again they do not deflectionize langue elements from words but they flectionize discursive elements into langue as abstract tool systems. Many of these languages are tonal languages which is not surprising since tones are an easy way to flectionize invariable elements in discourse.

If the third migration corresponds to agglutinative and synthetic analytical languages it must correspond to the third articulation. That means the people migrating then left the African nest when it had reached that third articulation and then they developed their languages on that basis. I state two waves because we cannot see in anyway agglutinative languages evolving into synthetic-analytical languages. In fact the languages of each big family seem to be durable which means their communicational power seems to be sustainable and thus make the languages durable.

That leads us to the communicational matrix. This is a complex question. First of all Homo Sapiens when he developed language and any child when he is confronted to language as soon as the 24th week of pregnancy and to communication as soon as he is projected into the real world find themselves in a communicational situation that becomes the matrix of this communication. This need to communicate will develop language and language in its development will enrich the communicational situation and translate this situational matrix into langue via communicational discourse. We must understand the stake both for Homo Sapiens and for a new-born baby is survival. Without that communication Homo Sapiens would not have been able to compensate his frailty and a new-born would not be able to adapt to the surrounding world. A feral child that is raised by animals outside human communication is not able to adapt to human society and he is not even able to learn how to communicate in any extensive way.

This communicational matrix developed in the mind of Homo Sapiens and builds in the mind of the child the simple three basic ternary (and quadruple) relations:
First we have the relation of the cry/call seen as the commanding action from the caller both agent and source to the callee, the call’s and the callee’s goal. The call is its own theme though it implies another expected theme: what will satisfy the need expressed by the call that will bring the callee in the shape of a carer/feeder/loving adult. What is implied is in fact the second level, that of the expected response.

**RECTUM 1 (Agent/Source) – RECTUM 2 (Goal) – REGIS**

Second we have the responding action of the Carer/Feeder/Loving adult who is the Agent and Source of this response bringing the Care/Food (Fodder)/Love to the Goal of the response, and the ternary matrix of the call becomes the quadruple matrix of the response.

**RECTUM 1 (Agent/Source) – RECTUM 2 (Theme) – RECTUM 3 (Goal) – REGIS**

And this quadruple matrix of the response leads to a second ternary relation, that of the satisfaction of the need behind the initial call in which the Agent/Goal performs the action that satisfies his need with the theme that fulfills this satisfaction.

**RECTUM 1 (Agent/Goal) – RECTUM 2 (Theme) – REGIS**

This matrix contains the basic structures of all syntactic utterances constructed around an action. No need to any genetic black box for Homo Sapiens to develop it and for a baby to learn it and from it for both Homo

<table>
<thead>
<tr>
<th>COMMUNICATIONAL MATRIX CALL</th>
<th>Action</th>
<th>Goal (Theme)</th>
<th>Theme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agent/Source (Agent/Goal)</td>
<td>CALL-ER</td>
<td>TO &amp; FOR Goal &amp; Theme CALL-EE</td>
<td>FOR Theme (Car-ER &amp; Care)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NEW-BORN MATRIX FIRST CRY or CRIES</th>
<th>Action</th>
<th>Goal (Theme)</th>
<th>Theme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agent/Source Baby</td>
<td>CRY(ies)</td>
<td>TO &amp; FOR Goal &amp; Theme Car-ER</td>
<td>FOR Theme Care, Food, Love</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NEW-BORN'S NEED</th>
<th>Action</th>
<th>Goal (Theme)</th>
<th>Theme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baby's need</td>
<td>CRY(ies)</td>
<td>TO &amp; FOR Goal &amp; Theme Car-ER Feed-ER Loving adult</td>
<td>FOR Theme Care Food Love</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SYNTACTIC MATRIX EAT</th>
<th>Action</th>
<th>Goal (Theme)</th>
<th>Theme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agent/Goal John Eat-ER</td>
<td>EAT-s</td>
<td>TO &amp; FOR Goal &amp; Theme Car-ER Feed-ER Loving adult</td>
<td>Bread Eat-EN (Food/Fodder)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FEED</th>
<th>Action</th>
<th>Goal (Theme)</th>
<th>Theme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agent/Source Paul Feed-ER</td>
<td>FEED-s</td>
<td>Goal (TO) John F-ED</td>
<td>Theme (WITH) Food/Fodder</td>
</tr>
</tbody>
</table>
Sapiens and the baby to develop language. Complex utterances are nothing but the repeating process that integrate one basic structure into another as many times as necessary to get the level of complexity required by the targeted meaning.

This leads then to the spanning of the full communicational syntax over langue, discourse and the residual communicational situation. The last element is constant and covers body language and other supra-segmental expressive elements. But the proportion of langue syntax and discourse syntax varies with the articulation of the language concerned.

### Conclusion

I do not pretend to see a cause-effect situation in this parallelism between the migrations out of Africa (plus the two post-Ice-Age migrations from the Iranian Plateau) and the linguistic families determined by the phylogenic approach of language I presented. But the parallelism between these two phenomena, one purely archaeological and historical, the other purely linguistic, brings up a question and we cannot evade it. Is there a relation between the two, is one the cause of the other, and which one is if any? What are the consequences of this parallelism on the linguistic theory itself? Doesn’t it make deflectionization the real end of the process producing languages that are more and more analytical and yet more and more synaptic and thus synthetic?

The main hypothesis I come to thought is the phylogenic order of languages. I insist here it is different from the one set by Guillaume in 1958-1960 but one thing is sure: there is no hierarchy of value and communicational power.

<table>
<thead>
<tr>
<th>My appellation</th>
<th>My 3 families</th>
<th>Guillaume’s 3 families</th>
<th>Guillaume’s appellation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st articulation</td>
<td>Semitic languages</td>
<td>Nostratic languages: isolating and agglutinative languages</td>
<td>Aire Prime</td>
</tr>
<tr>
<td>2nd articulation</td>
<td>Isolating languages</td>
<td>Semitic languages</td>
<td>Aire Seconde</td>
</tr>
<tr>
<td>3rd articulation</td>
<td>Agglutinative languages &amp; synthetic-analytical languages</td>
<td>Indo-European Languages</td>
<td>Aire Tercer</td>
</tr>
</tbody>
</table>
No language is inferior or superior, though they are anterior or posterior, but any language is able to fulfill the communicational demands any situation brings up and any language will evolve within its own articulatory architecture to satisfy these communication demands. As Paul Radin⁹ has often said and written, all languages even the most “primitive” have their philosophers.

For us, psychomechanicians, this is a tremendous challenge that requires all our attention because we cannot ignore what modern science is bringing up and consequently we have to devise the theory that will account for the linguistic phenomena that are brought up by this modern science. Guillaume and his theory are not at stake here, except as the method that might enable us to analyze and understand the enormous mass of new knowledge in our field. Only linguists, mainly psychomechanician linguists, definitely all cognitive linguists are the only people who can decipher the “Talking Stones” of the rock art of the Cosos in California just the same way as Egyptian hieroglyphs and Maya glyphs have been deciphered because they are a communicational ritualistic code supporting a discourse and a language.

BIBLIOGRAPHY

Chomsky, Noam, *Syntactic Structures*, Mouton de Gruyter, Berlin, 1957
Hawkins, Jeff, & Blakeslee, Sandra, *On Intelligence*, St. Martin’s Griffin; Reprint edition (August 1, 2005), New York, NY.
Russell, Bertrand, *Lectures on the Analysis of Mind*, “lectures both in London and Peking, and one lecture, that on Desire, has been published in the Athenaeum,” 1921, available at http://www.gutenberg.org/files/2529/2529-h/2529-h.htm

⁹ Paul Radin (1883-1959), American linguist, anthropologist, ethnographer: ethnoogy of religion and mythology; Native Americans of California and the Great Lakes.
