From Holophrase to Syntax: Intonation and the Victory of Voice over Gesture

Teresa Bejarano †
tebefer@us.es

ABSTRACT

In the origin of syntax, primitive, holophrastic signs had to be weakened (original, drastic 'bleaching') and to lose their previous status of whole message. The original syntax was probably thema/rhema syntax. The earliest themas repeat the hearer's message: the speaker embeds the hearer's message in his own message. In this way a holophrase could be weakened, and turn into a part of a syntactic combination. This pregrammatical, interpersonal 'recursive embedding' is embodied in sensorimotor processes. The upper level is embodied in the intonation; the lower level, in the articulatory-phonetic word. This decoupling of intonation and articulatory pattern—i.e. the emergence of intonation capable of comprising more than one word—facilitated the weakening of previous holophrases and the genesis of syntax. In time, that facilitation determined the preeminence of voice over gesture, regardless of whether or not that preeminence existed before syntax.

Keywords: bleaching (semantic weakening), embodied cognition, holophrastic sign; intonation, recursive mind.

In the origin of syntax primitive, holophrastic signs had to be weakened (original, drastic 'bleaching') and to lose their previous status of whole message. The original, pregrammatical syntax was probably the predication in response to the hearer's previous message. This previous message is embedded in the speaker's message. In this way, the hearer's holophrastic sign could be weakened and turn into a part of a syntactic composition. This is developed in Section 1.

Section 2 focuses on a more concrete issue. In original syntax, the hearer's holophrastic message is only repeated at articulatory level, not at intonation

† University of Sevilla, Spain.

level. The intonational unit is already able to comprise two words. This involvement with audiomotor resources could mitigate the difficulties of original syntax.
Section 3 extracts a consequence from Section 2. I begin by acknowledging the similarity between verbal sign and pantomimic gesture: both involve motor imitation. It is even likely that gestures were as important as voice for a long period of time. However, when the syntax arose, intonation would have given the voice the decisive edge.

1. Thema/rhema-syntax and the hearer's previous message. Interpersonal recursive embedding

The original syntax was probably thema/rhema-syntax (Aitchison, 1998; Givón, 1979; Hurford 2007 and 2012, p. 187, and chapter 9; Jackendoff, 2003; Tomasello, 1999). Since they were coined by the Prague Circle of the Twenties, the concepts of thema and rhema have been frequently reformulated. In this way, differences have been–more or less consistently–established between three dimensions–given/new and topic/comment, and also background/focus (see Hinterwimmer, 2011). But, since these reformulations refer to the grammaticalization level peculiar to our present-day language, this paper is not going to deal with them. I take the predication in response to the hearer's message as the basic type of this syntax. The *thema* mirrors what is known by the hearer; therefore, in the basic type, it mirrors the particular combination of knowledge and ignorance previously expressed by the hearer. The *rhema* adds what the hearer does not know, i.e. what the limited knowledge of the hearer needs in order to become similar to that of the speaker.

Thus, the previous message uttered by the hearer expresses a degree of knowledge different from the speaker's. In present-day language, previous messages revealing a particular combination of knowledge and ignorance or mistake can be questions or wrong predications, but also requests for something unavailable, or calls to someone absent. Those first two cases–we could say–would lead us into a vicious circle if we projected them onto the historical origin of syntax. By contrast, previous messages in the form of calls or requests could be expressed in an entirely holophrastic language.

But the key question is: Did 'parts of sentence' exist in the original holophrastic language? Heine and Kuteva (2007, p. 300) say that «at layer 1 (holophrastic, i.e., one-word utterances) there was only one kind of category, namely nouns». However, they acknowledge that «a category 'noun' does not make much sense unless contrasted with other kinds of word categories», and conclude: «What exactly the reconstruction of a category 'noun' means with reference to layer 1 is a question that is open to further research». In my view, the original holophrastic sign is as close to (or as far from) the meaning 'give me', as it is to the meaning 'hammer', for example, and, therefore, it is neither of those two meanings. I will return to this central issue in a moment.

Now, let us look at the first syntax of children. Of course, a large part of this first syntax will not provide any clues to the historical origin. A lot of a child's utterances and of the messages directed at him are aimed at helping the child to learn grammaticalized language. For example, the mother says 'Look who's there!, 'There!', and, then she, or the child, adds, 'Let's go!', and finally the mother recapitulates 'Let's go there!'. That is, after 'There!' and 'Let's go!', we arrive at the grammatical composition: 'Let's go there!' However, in the historical origin there would have been no need for the conversion of those two messages into one syntactic composition, since in this case the second message is unambiguously understood: Here, the automatic inference of the link between the two messages is enough. Obviously, utterances like these cannot throw any light on the historical origin of syntax. However, other kinds of utterances of children are more promising.

Let us focus on the following example (which I really observed). An adult and a child are playing with wooden blocks. Later, the adult requests the child, who has the box with the blocks in his hands, 'Give me more blocks! More!'. The child, who still has the box in his hands, sees that it is empty and says 'More, no'.

In the initial, pregrammatical syntax of children, the rhema often consists of a particular kind of 'metalinguistic' (Horn, 1989) or 'metarepresentational' (Wilson, 2000) negation: the negation does not refer to a predicate but to the previous message of the hearer. In the example, the 'no' serving as rhema resembles the negation used for rejections of someone else's requests or invitations (Dimroth, 2010), but with that 'no' what is rejected is the false belief involved in the speech act of the hearer.[1]

[1] Children well under the age of 4 can perceive wrong beliefs of other individuals. Mainly from Onishi and Baillargeon, 2005, this claim is no longer a fantasy for the researchers in the 'Theory of Mind'.

In this example the child repeats the hearer's message. If the hearer's previous message were not repeated, the negation would be interpreted as a refusal of the request. In other words, here the only way for the child to indicate–or, at least, suggest–that the 'no' is uttered to correct or update the belief involved in the hearer's previous message is to repeat the core of this message. On the contrary, in the case of B rejecting a previous request of A, or in the case of B supporting or continuing ('Let us go!') the previous message ('There!'), there is no need to go beyond holophrases.

In short, only when 1) the belief involved in the hearer's previous message must be corrected or updated, and 2) there is not yet any grammaticalization–when both conditions concur–, the repetition of the hearer's previous message is compulsory. In order to see the effect of grammaticalization, we can think of the questions. Thanks to our sophisticated grammar resources, our questions designate exactly the unknown we are asking about. In this manner, the answer will unambiguously be interpreted as the unveiling of that unknown, although only the rhema need be present. But we are interested in the first condition–in the original need of syntax. In a child, the repetition of a previous message and the correction of the hearer's false belief tend to co-relate, however inexact the co-relation. (Children will also use 'More, no' to reject the command/invitation to eat more pap, although in this case a rejecting holophrase would have been enough.) But my point is that in the historical origin, when there had never been any previous combination of words, such a repetition would show clearly the novelty of the process–the correction of the hearer's belief.

Let's see again the essential difference: The holophrase in which the hearer's previous message is supported or rejected *versus* the original, indispensable syntax (whose first part is a copy of the hearer's message). In the first case, the previous message has already become a favourable or unfavourable element of the current reality. In the second one, the previous message continues to be just a message.

Returning to the analysis of the wooden-block example, the 'more' uttered by the first speaker (a 'more' that could be expressed in holophrastic language) becomes the 'more' which functions as the thema for the second speaker. Both 'more' have the same articulatory pattern, meaning and concrete referent, and both involve the false belief that there is still some left of what has been requested. However they are also different. The second 'more' does not work as a request. Instead, it is the platform to which the 'no' will be attached.

In my view, the second 'more', which functions as a thema, would not be simply part of a message. In the original syntax, that 'more' still maintains its whole-message status, since it mirrors the hearer's message. Admittedly, that whole-message status is now embedded in another message, but that does not mean that it has disappeared.

In other words, there is recursive embedding of someone else's message in one's own message. ('Message inside Message': 'X inside X'.)[2] Certainly this embedding is not a grammatical, Chomskyan recursion.[3] However it can be included in Corballis' (2011) 'recursive mind', if we add that primitive recursions would be interpersonal. Thus, pregrammatical, original syntax would be an early stage of the recursive mind.[4]

Certainly the classic 'embedding of one message in another' is grammatical reported speech. But this is really an embedding of sentences, and must have been a much later conquest. In reported speech neither the speaker, nor the place nor the time of the original message are present. Consequently, these elements must be provided through linguistic means. More importantly, this classic embedding requires both syntactic subordination and the meaning 'say/tell', which most probably appeared at a much later time than the point when humans started saying things. By contrast, the predications of answer–the 'More, no' example–do not have such demanding requirements. In short, the kind of 'embedding of one message in another' considered up to now as paradigmatic could in no way have been the original one. Despite these differences, both kinds–this is what mainly interests me–share three crucial features: interpersonal repetition, syntactic advance and the weakening of the embedded message. Let us see the weakening. The predication 'The Earth is flat' that is reported in 'The ancients said that the Earth was flat' is seen by the speaker as a mere mental state–a mental state of somebody else.[5] In 'More, no',

[2] Levinson (2013, p. 154) also connects recursion and pragmatics, but in a different way. Thus, the alternative explanation that Legate et al. (2014, p. 524) propose in their criticism against Levinson cannot be applied to my examples.

[3] Even 'recursive Merge'– «the most simple recursion», according to Hurford (2012, p. 51), although other authors, as Bickerton (2009, p. 6), do not accept this redefinition of recursion–requires more than two words. Anyway, there is an advantage (which has always fascinated me) for the older definition of recursion. «This shares an essential ingredient with some uniquely developed human trait–language about language, thought about thought, converse about each other's thoughts»: Hurford (2012, p. 575). Or, in a more general formulation, recursive processes are crucial for human mind: Corballis (2011).

[4] But the earliest recursion could be prelinguistic (Bejarano, submitted).

[5] It is thanks to this loss of referential strength that the whole composition can be true. This weakening–a Fregean point–has been reformulated in an evolutionary perspective (Gilbert, 1991; Jerison, 1988).

the 'more' is also weakened: it has stopped working as a request, and it is seen by the speaker as a mere repetition of the hearer's message.

The step from holophrastic language to syntax must have been very difficult. The meaning of words had to be integrated into an upper-level unit and contribute to a new meaning which had never been learnt before. This is a familiar issue–Humboldt's "infinite use of finite means ", Chomsky (1965)–, which can be associated with "the expanded working-memory " (Coolidge & Wynn, 2005). But, in my view, we must also look at a second, more basic source of difficulties: Holophrastic signs, in order to become true words, have to suffer a semantic weakening. If the genuine word (the noun 'hammer', for example) can serve as efficiently for any function–to request, narrate, ask, say how you say it, etc.–, it is because its meaning no longer includes its function. By contrast, the holophrastic sign was inseparable from its function and its strength. The request was not the meaning of a use of that sign, but it was the meaning of the sign itself. The loss of that strength and that self-sufficiency–the gain in versatility–is tantamount to the origin of syntax. In other words, 'semantic weakening' or 'bleaching'–a central concept in grammaticalization studies–can be extrapolated to the very origin of syntax, i.e. to the original, interpersonal recursion. The original, drastic bleaching would have been the crucial difficulty. Schafer and colleagues (2013) have found that «words that can be associated with many words are underrepresented in the comprehension vocabulary when a weak central coherence can be attributed to the patient». In a similar way, the change from comprehension of holophrastic sign to comprehension of genuine word would require an enormous increase of central coherence.[6]

Obviously, there are two hypotheses about this issue. On the one hand, many authors–not only Fodor (1975) and his extreme innatism, but also Givón (2002) or Hurford (2003) – accept that, previously to language learning, thought already possesses a compositional, syntax-like format. On the other hand, the original syntax would derive from the bleaching of holophrases that is achieved in the dialogue.

According to this second hypothesis, prelinguistic perception and thought lack syntax-like compositionality. It is indisputable that within a prelinguistic perception there are agents, action, quality. However within a perception none

[6] The difference between words and holophrastic signs can be reinforced if we accept–cf. Taylor (2012)–that the most properly linguistic part of the meaning of a word is constituted by the past episodic links of this with other words.

of these features would be addressed separately. That is, at the conscious perception level the scene would be captured as a unity, not as a composition of features, each receiving separate attention. Of course, the construction of perception integrates many features and each of these features is captured by different groups of neurons. Some features, such as line direction or colour, are captured in the first stages of visual processing. (In these features we know for sure that a 'perceptual binding' takes place and it is thanks to it that we do not perceive 'red' and 'square shaped' separately but a 'red square'.) By contrast, other features–which can well be called 'concepts'–are captured at a later processing level of the scene. However, my point is that neither of these features is attended to in the way the successive meanings within a syntactic composition are attended to.

In addition, according to the second hypothesis, these features are not the original root of words, in spite of the fact that they are present in the animal's mind. Primitive holophrastic signs had no relation to these features (nor had they any relation to the 'parts of the sentence'), because their aim was not to mirror reality. They were only used for calling or requesting. Certainly, after interpersonal recursion and consequent bleaching, meanings would have come nearer to those features, and, as a result, these features could progressively be separated from the perceived whole.[7] For example, a holophrastic sign whose only original function was to act as a vocative to call a particular individual, after the genesis of syntax, would have become a word for that individual, in increasingly closer resemblance to the prelinguistic 'concept, or recognition pattern, of that individual'. However, according to the second hypothesis, we cannot presume that that close resemblance was in the origins of language. *Separately attendable* concepts did not exist in prelinguistic minds. Instead, their genesis took place at a later stage and was mediated by interpersonal communication. It is thanks to these concepts that we can 'decompose and recombine'– «we disassemble the world, and then create alternative versions of reality» (Tattersall, 2009, p. 586). Certainly, syntactic combinations produce new wholes, some of which can in the end be similar to prelinguistic perceptual

[7] This separation might have also been generated at the level of features captured in the first stages of visual processing. "Artificially produced geometric forms (75000 bp) serve as an externally derived supernormal correlate of the patterns processed by the early visual system leading to a proto-aesthetic pleasure (or perceptual fluency not determined by familiarity)" (Hodgson, 2014). See also Arden et al. (2014; I emphasise p. 1845 – "Do not mention any of the body parts").

.

wholes. However the new whole, even if it mirrors a trivial scene, has been "created " by human skill, is an embedding, upper-order unity.

In other words, in holophrastic language there was no neutral meaning which could subsequently be used either to represent the world ('declarative' use), or to try to change it ('imperative' use). Instead, there was request intonation or call intonation within which phonetic articulation would have progressively developed to make more specific the object of the request or call. The road to neutral, versatile meaning was long and arduous.

To be more exact, instead of two opposite hypotheses it would be better to say that the second one makes the way suggested by the first one more indirect and complex. 'Concepts similar to those of an animal mind + communicative signs representing no concept at all + bleaching > Syntax': the two intermediate steps are exclusive to the second hypothesis. On one hand, this hypothesis may seem less parsimonious, on the other, this hypothesis would better correspond with the enormous difference between human and non-human primates.

To make a more concrete contrast between the hypotheses, let us concentrate on denials. According to the first hypothesis (more specifically, according to a non-innatist version of this hypothesis), denials might have originated from the surprise caused by the absence of an expected element (Davidson, 1982, p. 318): the surprise would have led to the denial of the wrong expectation or prediction held previously. Accordingly, the origin of denials would have been intrapersonal. On the other hand, the second hypothesis supports the increasingly accepted idea that the surprise or 'prediction error' is dealt with using a much less costly resource than a denial: the incorporation of the latest information into the perceptive content will suffice. There is no need to deny the wrong prediction, because this disappears when the latest information arrives. Consequently, in the light of the second hypothesis, denials would have originated in interpersonal communication. In the beginning, i.e., before syntax, only received requests would be denied. Later, in concurrence with the genesis of syntax, the speaker would deny the hearer's wrong beliefs: denial as rhema. In time, true negative statements can become mere representations of reality.[8]

[8] Certainly many exemplars of this type of statements become mere representations of reality. However, if the interpersonal origin of the type is forgotten, the Bergsonian question arises: where does the characteristic negated come from?

I do not believe we have the resources today to reject either the existence of prelinguistic syntax or the opposite. The only thing that we can do in a case such as this is 'wager and see'.[9] Thus, I wager on the original bleaching and the interpersonal genesis of syntax.

The question then arises as to whether there was anything that could mitigate those difficulties in the original bleaching. Looking for an answer, Section 2 will make the above proposal more concrete. 'Embedding your holophrastic message not only in my message but also in my intonation': This is the new schema.

2. The two levels of embedding and their respective audiomotor resources: Embodied embedding

At the articulatory level, the two 'more' are identical. But, apart from that identity, there is an intonation difference between them. One of the functions of intonation, both now and–even more importantly–before grammaticalization, is to indicate the end of the message. At the holophrastic stage, the end of the message coincides, obviously, with the end of the word: the intonation unit and the articulatory-semantic unit run parallel. But later, with the emergence of syntax, the intonation pattern becomes capable of comprising two or more words.

[9] When did "declarative" communication emerge in evolution? Given that I do not consider the cries of vervets to be declarative communication, let me be more precise: When–where in the bridge between our nonhuman ancestors 7 million years ago and modern human–did declarative pointing arise? Please note that, in the case of an early emergence of declarative pointing, declarative communication would already have been deeply established when the first articulated holophrases–undoubtedly, very close to human language–began and would therefore have most likely impregnated (against the second hypothesis) these holophrases. Could this be a path for the second hypothesis soon to become a falsifiable hypothesis? I am inclined to accept the three following claims. 1) The white of eye is adaptive because it facilitates the understanding of pointing gestures as communicative. 2) The declarative pointing did not arise later than the imperative one. In my view, chimpanzees can realize that a caged–or physically disabled–companion desires an object, without them understanding as communicative the–really communicative–unlearned gesture of the companion of getting as close as possible to the desired object. So, the true understanding of pointing would appear in evolution simultaneously with the declarative use of gestures. 3) The origin of the white of the eye could be soon clarified. In short, I suggest linking the first hypothesis with an early date for the white of the eye and the second one with a late date. In Bejarano (2011, chapters 3, 4, 6) and Bejarano (submitted), I propose that the true understanding of pointing gestures, since it involves a demanding requisite, arose extremely late (in Neanderthal, if not in Sapiens). If, on the contrary, paleogenetics discovers that the white of the eye had already appeared earlier, then that result would be unfavourable regarding the second hypothesis. It is true that this suggestion depends on other theoretical crossroads and, consequently, it does not really lead to the falsifiability of the second hypothesis. However, it can be seen as a step in that direction.

In 'More, no', the intonation comprises the whole of the two-word message, whereas the one-word message of the hearer is only repeated at articulatory level and not at intonation level. Let us say it more concretely. In 'More, no', 'more' is identical to the hearer's previous message, and remains a complete message. But at the same time, that 'more' occupies the first part of the intonation pattern and therefore has become a part of a message. In other words, the two levels of that recursive embedding respectively have recourse to two different audiomotor resources. The upper-order level of embedding is represented in the intonation, and the lower-order level, in the phonetic articulation of the word.

Thus, original syntactic recursion would have been not only interpersonal, but also embodied. 'Embodied cognition' is a multifaceted approach, and some of its elements are controversial. But there is a wide acceptation of its oldest, quasi-Piagetian nucleus. It is in this sense that I use 'embodied'. The recursive embedding would be intrinsically associated with those two sensorimotor patterns, and in this way it would be facilitated, both in children (see Keitel et al., 2013) and in historical genesis.

If my proposal is correct, then the decoupling of intonation and articulated-semantic pattern, i.e., the emergence of intonation capable of comprising more than one word, is a crucial milestone.[10] We do not know when the decoupling took place. But there is little doubt that it had a firm basis.

The evolutionary root of intonation probably consists of primate cries, in which the continuum of intonation intensity mirrors the continuum of emotional intensity. Later, after the genesis of syntax, intonation certainly becomes less rooted and less emotional. But in the beginning, its production and reception are disconnected from learning. On the other hand, articulatory patterns depend on imitative learning of new motor sequences, and we know that neither syllabic abilities, nor–in the general motor field–the imitation of new sequences (which is the only useful motor imitation) are developed in non-human primates. In short, there is a strong contrast between intonation and articulation. Side by side, however, as was to be expected, antecedents of articulation are finally beginning to appear in non-human primates (vs. songbirds: Rosselló, 2014), some facial expressions with a speech-like rhythm

[10] Obviously, this decoupling does not mean a disconnection. The stressed syllable of the rhema becomes the peak of the intonation of the sentence. More important, as a reviewer wrote, "tone languages show that the division of audiomotor resources is partly influenced by culture. Still, it is probably right to say that there's a strong underlying basis for the division".

(Ghazanfar & Takahashi, 2014), or antecedents of voiceless consonants (Lameira et al., 2014).

We must also take into account the hemispherical specialization of the brain. Articulatory, learned patterns are mostly associated with the left hemisphere: Broca's discovery has been qualified over time, but its essential truth remains. By contrast, intonation (or "the system involving recognition of pitch contours ") does not seem to have such a strong hemispherical preference (Peretz & Hyde, 2003). This must have facilitated–or rather, consolidated–the crucial decoupling.

Both in children and in history, the articulatory-phonetic pattern would have been the first and foremost cause for the decoupling of articulation and intonation. With regard to articulatory sequences, de Boer and Zuidema (2010) emphasize a very interesting distinction: 'true segmentation in the producer' vs. (more primitive) 'superficial combinatorial structure'. They characterize the superficial one as «combinatorial structure that can be observed by an outside observer in a system of signals, but that is not actively used by the agents using the signals» (Boer & Zuidema, 2010, p. 144). Very probably, this distinction is valid both for primaeval language and for children. Sequences of babbling, and even the earliest holophrases, would be 'superficial combinatorial structures'. Returning to our argument, the 'true' learning of articulatory-phonetic sequences is very different from (originally emotional, holistic) intonation, and this difference could have triggered the decoupling of intonation and articulated pattern. Nowadays (but, in my view, not in the very historical origin!) protodeclaratives (pointing + holophrase without any innate intonation), which only communicate in order to learn (Begus & Southgate, 2012) or teach words, facilitate that decoupling in children.

I am focusing on the consequences of this decoupling. «The division of labor between prosody and segmental sounds may shed an important light on the evolution of language» (Chafe, 2000, p. 253). This Section has suggested one consequence of that 'division of labor': original syntactic recursion would have been an embodied, facilitated recursion. Now, from that consequence I am going to extract another consequence.

3. Intonation, syntax and the victory of voice over gesture

The articulatory patterns of words are produced by motor imitation of models on the part of the learner. The same motor imitation involved in the learning of words also takes place in pretending (communicative pantomime and children's symbolic play): In pretending, since the reproduction is executed without the proper object or context, any action becomes a genuine motor imitation. This similarity between words and iconic gestures is in fact a very strong one. Only in these two types of patterns is motor imitation maximally implemented. In technical tasks, although they have been learnt from a model, other factors besides motor imitation intervene. Goal-imitation (aka 'mere emulation'), commonly associated with animals, also occurs in humans. More precisely, it has been observed that accuracy in genuine motor imitation decreases when there are objects, or, stated in the reverse, the temptation to turn to mere emulation grows stronger when the movement is acting on objects (Gattis et al., 1998). Moreover, the size, shape, weight, etc. of objects impose constrictions on the degrees of freedom of the motor system regardless of the model. In short, it is only in imitations performed without the handling of objects that the motor pattern is really under the control of the model. Thus, articulatory-phonetic imitation and communicative pantomime are very closely related. It is likely that gestures were as important as voice for a long period of time, or perhaps even more important (see Arbib 2012; Corballis, 2002; Donald, 1991). Also, although pointing is different from iconic gestures, handedness (left hemisphere) in toddlers' pointing is connected to this issue (Vauclair & Cochet, 2013).

However, when the need for syntax arose, intonation would have given the voice the decisive edge. Intonation and articulated pattern, which are heard at exactly the same time and in the same sequential linearity, nevertheless may not coincide in their respective units, which, I have suggested, facilitates the embedding of a whole in a new whole, and, consequently, syntax. Deprived of that edge, gestures have eventually become merely a minor, subordinated means.[11]

The rest of the advantages traditionally invoked in favour of the voice do not seem at all decisive. If vocal language can take place in the dark, gestural language can claim to be able to take place in silence, which can sometimes be

[11] But «they can perform many of the functions of natural language» (Goldin-Meadow, 1999, p. 420), even in the thought itself (Cartmill et al., 2012; Warburton et al., 2013).

useful. Busy hands–that is, hands that cannot perform other tasks – are certainly a disadvantage for any gestural language, but only in extended productions, which are unlikely until relatively recent historical times. In addition, analogy does not necessarily have to have been a limiting factor for gestures. Continuous use would have made them progressively less analogical (Tomasello, 2008, p. 223; Smith & Höfler, in press).

In short, intonation would be the major cause responsible for the overwhelming success of the voice. The emergence of spatial verbal deictics–'there', for example–is a sign of this success. At first these deictics would have been unnecessary; a pointing gesture can fulfil their function. However, they came into being.[12] Most probably the reason is this: spatial indications had to be verbalized if they were to be integrated into a syntactical combination. This happens quite soon in children. Certainly children, in the very beginning of their post-holophrastic stage, produce some one-word messages which, accompanied by a pointing gesture towards an object, designate a quality or fact about the object: Here, there are not yet any verbal deictics. But children soon begin to include verbal deictics, mainly in contrastive functions. In short, syntax provokes a strong preference for the vocal channel.

Given the growing primacy of voice, language only by gesture has been relegated to communication with foreigners or the deaf. True languages of the deaf, which are predominantly manual, use another commonality, besides their shared relationship with motor imitation. This second commonality is that one can see one's own hands (more easily than any other part of one's own body), and one can hear one's own voice.[13] This self-perceptiveness greatly facilitates imitative learning, since the model and the reproduction adopt a common perceptual format for the learner.[14] However, despite these two commonalities of vocal language and manual language of the deaf, there are significant differences. In manual languages of the deaf, the prosody is certainly produced not only by face and head, but also by hands. At the end of an intonational phrase, the hands carry phrase-final lengthening (Sandler et al., 2014, p. 254). However even this manual prosody is unable to embed the sequence of gestures-words. On the contrary, intonation can embed the sequence of words. «Intonation is spoken, and that connects it directly with the stream of speech» (Bolinger, 1983, p. 70). For that reason, oral language and manual language

[12] "At layer 4": Heine & Kuteva (2007).
[13] During her learning, Helen Keller also perceived the results of her movements.
[14] Is there another consequence of self-perceptiveness? See Bejarano (2011, chapter 9).

are differently efficient in the process of providing the upper order, in which the different meanings must be embedded, with embodiment. In vocal language, the facilitation of syntax is more efficient.

Obviously, recursion and syntax are not unknown to the deaf. Indeed, deaf people have used their linguistic abilities in the socio-cultural process of generating new, independently developed languages (Senghas et al., 2004; Sandler et al., 2014). In more general terms, recursion, which was interpersonal and embodied in origin, subsequently became intrapersonal (to use the Vygotskian term) and began to rely on historical developments–grammaticalization or writing. But my point is that in the origin of syntax–in the original bleaching–, not only the repetition of the hearer's message but also the two different audiomotor resources of vocal language were decisive resources.

ACKNOWLEDGEMENTS

I am very grateful to an anonymous reviewer whose comments have helped me in a crucial way.

REFERENCES

Aitchison, J. (1998). *The Seeds of Speech*. Cambridge: Cambridge University Press.

Arden, R. et al. (in press). Genes Influence Young Children's Human Figure Drawings and Their Association With Intelligence a Decade Later. Psychological Science.

Begus, K. & Southgate, V. (2012). Infant pointing serves an interrogative function. Developmental Science, 15, 611-617

Bejarano, T. (2011). *Becoming human: From pointing to syntax*. Amsterdam: Benjamins.

Bejarano, T. (submitted). In Your Shoes When You Look at Me: Clear-cut human uniqueness.

Bickerton, D. (2009). Syntax for Non-Syntacticians. In D., Bickerton & Szathmáry (eds.), *Biological Foundations and Origin of Syntax*. Cambridge: MIT Press, 3–14.

Bolinger, D. (1983). Intonation and Gesture. *American Speech, 58,* 156–174

Cartmill, E., Beilock, S. & Goldin-Meadow, S. (2012). A word in the hand: Action, gesture, and mental representation in human evolution. *Philosophical Transactions of the Royal Society, B, 367*, 129–143

Chafe, W. (2002). A memoir. *Historiographia Linguistica, 29*, 245–261.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge: MIT Press.

Coolidge, F. & Wynn, T. (2005). Working memory, its executive functions, and the emergence of modern thinking. *Cambridge Archaeological Journal, 15*, 5-26.

Corballis, M. (2002). *From Hand to Mouth*. Princeton: Princeton University Press.

Corballis, M. (2011). *The Recursive Mind: The Origins of Human Language, Thought and Civilization*. Princeton: Princeton University Press.

Davidson, D. (1982). Rational animals. *Dialectica, 36*, 317-327.

de Boer, B. & Zuidema, W. (2010). Multi-agent simulations of the evolution of combinatorial phonology. *Adaptive Behavior, 18*, 141–154.

Dimroth, C. (2010). The acquisition of negation. In L. R. Horn (ed.), *The Expression of Negation*. de Gruyter, 39-72.

Donald, M. (1991). *Origins of Human Mind. Three Stages in the Evolution of Culture and Cognition*. Cambridge: Harvard University Press.

Fodor, J. (1975). *The Language of Thought*. Cambridge: Harvard University Press.

Gattis, M., Bekkering, H. & Wohlschläger, A. (1998). When actions are carved at the joints. *Behavioral and Brain Sciences, 21*, 691–692.

Ghazanfar, A. & Takahashi, D. (in press). Facial Expressions and the Evolution of the Speech Rhythm. *Journal of Cognitive Neuroscience*.

Gilbert, D. (1991). How mental systems believe. *American Psychology, 46*, 107–119

Givón, T. (1979). *On understanding grammar*. New York: Academic Press.

Givón, T. (2002). The visual information-processing system as an evolutionary precursor of human language. In T. Givón & B. F. Malle (Eds.), *The Evolution of Language out of Pre-language*. Amsterdam: Benjamins, 3-50.

Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends in Cognitive Sciences, 3*, 419–429

Heine, B. & Kuteva, T. (2007). *Genesis of Grammar*. Oxford: Oxford University Press.

Hinterwimmer, S. (2012). Information Structure and Truth-Conditional Semantics. *Semantics.* 1875–1908.

Hodgson, D. (2014). Decoding the Blombos Engravings. *Cambridge Archaeological Journal, 24*, 57-69.

Horn, L. (1989). *A Natural History of Negation*. Chicago: Chicago University Press.

Hurford, J. (2003). The neural basis of Predicate-Argument Structure. *Behavioral and Brain Sciences, 26*, 261-283.

Hurford, J. (2007). The origin of noun phrases. *Lingua, 117*, 527–542.

Hurford, J. (2012). *The origins of grammar*. Oxford: Oxford University Press.

Jackendoff, R. (2003). *Foundations of Language*. Oxford: Oxford University Press.

Jerison, H. (1988). Evolutionary neurology and the origin of language. In M. Landsberg (Ed.), *The genesis of language*. de Gruyter, 3-9.

Keitel, A. *et al.* (2013). Perception of conversations: The importance of semantics and intonation in children's development. *Journal of Experimental Child Psychology, 116*, 264–277.

Lameira, A., Maddieson, I. & Zuberbühler, K. (in press). Primate feedstock for the evolution of consonants. *Trends in Cognitive Sciences.*

Legate, J., Pesetsky, D. & Yang, C. (2014). Recursive misrepresentations: A reply to Levinson. *Language, 90*, 515-528.

Levinson, S. (2013). Recursion in Pragmatics. *Language, 89*, 149-162.

Onishi, K. & Baillargeon, R. (2005). Do 15-Month-Old Infants Understand False Beliefs? *Science, 308*, 255–258.

Peretz, I. & Hyde, K. (2003). What is specific to music processing? *Trends in Cognitive Sciences, 7*, 362–367.

Rosselló, J. (2014). On the separate origin of vowels and consonants. In B. de Boer, & T. Verhoef (Eds) *Evolang X - Workshop on Signals, Speech and Signs*, 29-33 http://ai.vub.ac.be/~bart/ProceedingsSignalsSpeechSigns (accessed 16 July 2014).

Sandler, W., Aronoff, M., Padden, C. & Meir, I. (2014). Language emergence. In N. Enfield, P. Kockelman & J. Sidnell (Eds.), *The Cambridge Handbook of Linguistic Anthropology*. Cambridge: Cambridge University Press, 246-278.

Senghas, A., Kita, S. & Özyürek, A. (2004). Children creating core properties of language: evidence from an emerging sign language in Nicaragua. *Science, 305*, 1779–1782.

Schafer, G., Williams, T. & Smith, P. (2013). Which Words are Hard for Autistic Children to Learn? *Mind & Language, 28*, 661–698.

Smith, A. & Höfler, S. (in press). Metaphor as the driving force of language evolution. In *Proceedings of EVOLANG 2014*.

Tattersall, I. (2009). Becoming Modern *Homo sapiens*. *Evolution: Education & Outreach, 2*, 584-589.

Taylor, J. (2012). *The Mental Corpus: How language is represented in the mind*. Cambridge: Oxford University Press.

Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge: Harvard University Press.

Tomasello, M. (2008). *Origins of Human Communication*. Cambridge: MIT Press.

Vauclair, J. & Cochet, H. (2013). Hand preference for pointing in toddlers. *Developmental Psychobiology, 55*, 757–765.

Warburton, E. et al. (2013). The Cognitive Benefits of Movement Reduction. *Psychological Science, 24*, 1732–1739.

Wilson, D. (2000). Metarepresentation in Linguistic Communication. In D. Sperber (Ed.), *Metarepresentations*. Oxford: Oxford University Press, 411–441.