

---

**Linguistic theory and fieldwork in  
interaction: the case of Pirahã**  
Eugénie Stapert

---

Proceedings of Conference on  
**Language Documentation & Linguistic Theory**

Edited by Peter K. Austin, Oliver Bond & David Nathan

7-8 December 2007 School of Oriental and African Studies, University of London

School of Oriental and African Studies  
Thornhaugh Street, Russell Square  
London WC1H 0XG  
United Kingdom

Department of Linguistics:  
<http://www.soas.ac.uk/academics/departments/linguistics>

Hans Rausing Endangered Languages Project:  
<http://www.hrelp.org>  
[elap@soas.ac.uk](mailto:elap@soas.ac.uk)

© 2007 Eugénie Stapert

No part of this publication may be reproduced, stored in a retrieval system, or transmitted, on any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the author(s) of that part of the publication, except as permitted by UK copyright law.

ISBN: 978-0-7286-0382-0

This paper can be cited as:

Eugénie Stapert. 2007. Linguistic theory and fieldwork in interaction: the case of Pirahã. In Peter K. Austin, Oliver Bond & David Nathan (eds) *Proceedings of Conference on Language Documentation and Linguistic Theory*. London: SOAS. pp. 235-242.

or:

Eugénie Stapert. 2007. Linguistic theory and fieldwork in interaction: the case of Pirahã. In Peter K. Austin, Oliver Bond & David Nathan (eds) *Proceedings of Conference on Language Documentation and Linguistic Theory*. London: SOAS. [www.hrelp.org/eprints/ldlt\\_27.pdf](http://www.hrelp.org/eprints/ldlt_27.pdf)

# **Linguistic theory and fieldwork in interaction: the case of Pirahã**

EUGÉNIE STAPERT

*University of Manchester*

## 1. THE CYCLE OF THEORY AND DATA<sup>1</sup>

To live is to learn about the world. Consciously or unconsciously we learn something new every day, whether how to interpret a traffic light, how to work a computer, or about principles of human interaction.

In this process of understanding we intuitively apply the natural cycle of theory and data: we observe phenomena, formulate hypotheses about their interpretation and test these hypotheses in the real world. We come to generalizations and theories if our hypothesis is corroborated, and modify our hypothesis when counterevidence is found.

Formulation of scientific theory should be guided by exactly the same principles. Science is not more than a cultivated way of learning in which interaction between theory and data is as important as in everyday survival. This might seem self-evident, but in practice it is not always recognized as such. Sometimes too much attention is paid to data and generalizations are missed out on, but more frequently theory is assigned major importance even though it is based and tested on little data.

This is understandable since generalizations and theory provide a solution to a problem, whereas data constitute the problem we are trying to solve. In addition, data are not always unlimitedly accessible which may impede thorough testing of the theory. Therefore it is no surprise that sometimes we see patterns where in fact they might not exist or do not care too much about little data.

Although understandable, it does not eliminate the serious consequences a disruption of this cycle could lead to. Within linguistics as much as in other disciplines theoretical concepts should be based on patterns and hypotheses, which have emerged from data. However, when a theory loses its connection with data it becomes a coherent whole on the basis of its internal logic rather than reflecting real language facts. A possible danger is that theoretical concepts that originated as a representation OF the language have now become a pattern FOR the language. Data are forced into the framework of the theory to support its universal truth even if a different interpretation could sometimes explain the phenomenon more efficiently and faithfully (cf., Occams razor). Although undesirable in any area of research, such a theoretical discussion can become a serious issue when it concerns fundamental assumptions about human language.

In this paper I will illustrate this with a case study concerning the debate about recursion in Pirahã. In this debate it appears that disagreements based on purely theoretical convictions are impossible to solve unless new data breaks the circle of argumentation and proves or disproves our assumptions.

---

<sup>1</sup> I thank Ted Gibson, Michael Frank and Nigel Vincent for useful comments on this paper.

Section 2 will explain the main theoretical argument about recursion and Pirahã and how new data are needed in order to come to a conclusion. Section 3 provides results from my own fieldwork on Pirahã, which aimed to supply these necessary facts. It will focus on one experiment on dependent clauses in particular. After a description of the procedure and a discussion of the results section 4 will give a conclusion.

## 2. RECURSION AND PIRAHÃ

### 2.1. *Recursion*

The concept of RECURSION has entered linguistic theory from mathematics and plays an important role especially in formal language theories (Tomalin 2006). In mathematics and logic recursion refers to a formula that calls itself (Tomalin 2006), i.e. the entity itself recurs in its own explanation e.g.  $f = f(x)$  or  $n = (n + 1)$ . This can be translated into linguistic terminology as an entity defined in terms of itself (Tomalin 2006) such as nested possessives (NP occurs within an NP), or relative clauses (clause occurs within a clause).

Within linguistic theory recursion has proven to be a useful tool to account for certain phenomena in human language. Recently it has even been argued that recursion is possibly the only device which can explain the distinguishing characteristics of human language and that therefore it might be considered the only true language universal.

This idea has been put forward in an influential article by Hauser, Chomsky and Fitch (2002) (HCF from here). In their view, our ability for language is located in ‘the language faculty’, which itself consists of two parts: the faculty of language in the broad sense (FLB) including pragmatics, semantics, breathing system, sensory-motor system, and on the other hand the faculty of language in the narrow sense (FLN). This second part constitutes a subdivision of the first and contains only those properties that distinguish language from other human communication systems, as well as from all non-human communication systems. They hypothesize that FLN is equivalent to recursive syntax, since this mechanism would account for most of the unique properties of human language proposed by Hockett (1960). The above definition of recursion allows for the definition of all members of a set without listing them. That is why recursion could account for discrete infinity, creativity, and displacement because it is an unambiguous way to display understanding of theory of mind.

Since in their view recursion is located in the syntactic component of the language faculty, the implication is that recursion is not to be found in non-human communication systems (since they do not use syntax) and must be found in all human language systems (since there are no human languages without syntax).

### 2.2. *Pirahã*

These strong and challenging hypotheses have evoked a chain of reactions, including a reply by Everett (2005). He claims that Pirahã, a Muran language spoken in the northwest Amazon shows no evidence of recursive syntax at all. It

needs to be stressed, however, that like every human being they think and reason recursively: they are able to entertain a string of propositions in asymmetrical relationship (such as conditional or causal propositions) like everyone else. Because of this Everett argues that rather than being a syntactic device in an autonomous language faculty, recursion is part of human cognition and that this gives it its universal value. Syntactic recursion is just one possible way to deal with universal recursive thought, which could be reflected through other strategies as well, as is illustrated in (1).

- (1) (a) *The man who ate the farinha returned to the house.*  
 (b) *The man ate the farinha. The same returned to the house.*

In this set of sentences we have to do with two events in both cases. However the relation between the two events is expressed by different means. In (1a) subordinating syntax is used (an instance of recursion) whereas in (1b) we see a paratactic structure. However, (1b) could get the same message across as (1a) when presented in the appropriate context. In that case it would be our natural ability for linking propositions that yields recursive interpretation. This second strategy is employed in Pirahã, as can be seen in (2).<sup>2</sup>

- (2) *xigihí xágaisi kohoái-p-í sigíaihi xaboópápáhá kaioó-xio*  
 DEM farinha eat-IMPERF-PROX same return house-DIR  
 ‘The man ate the farinha. The same returned to the house.’

Both views recognize recursion as an important part of human communication and even of our humanness but their theoretical convictions lead them to analyse the data in very different ways: HCF analyse Pirahã as containing recursive syntax, whereas Everett doesn’t do so any longer (but see Everett 1986 for his earlier analysis).

In the debate particular attention is paid to the suffix *-sai* in Pirahã, which has been adduced as an indicator of embedding since it occurs in contexts of semantic embedding.

- (3) *Pii-boi-sai ti kahá-p-i-hiab-i-hai*  
 water-come-NOM 1 go-IMPF-EP-NEG-PROX-REL.CERT  
 ‘If it rains I will not go’  
 (Everett 1986: 38)

This sentence has been analysed in two ways. Everett (2005) argues that it consists of two separate clauses, *It rains* and *I don’t go*, and that the recursive, subordinate interpretation is derived on the basis of pragmatics and discourse.

<sup>2</sup> The abbreviations used in this paper are 1 = first person, 3 = third person, DEM = demonstrative, EMPH = emphatic marker, EP = epenthetic vowel, INTENS = intensifier, NEG = negation, NOM = nominalizer, OSERV = observative, PROX = proximal, REL.CERT. = relative certainty, DIR = direction.

There is no trace of recursion in the syntax. In their reply to Everett Nevins Pesetsky and Rodrigues (2007) have refuted this analysis arguing that this is an instance of recursive syntax, *-sai* being the indicator of embedding. They base their analysis on the same theoretical assumption as HCF that recursive syntax is universal and, ironically, on Everett's own older work on Pirahã in which he analyses *-sai* is an indicator of embedding on the same theoretical assumptions.

Now which analysis reflects the data in Pirahã best? It seems that scholars are arguing against each other on the basis of their theoretical conviction and forget that the answer is in the data themselves. Instead of supporting their analyses with new conclusive data, the debate has become purely theoretical leading to unsolvable circular argumentation: evidence for universality of recursion is seen in data, which are analyzed in these terms on the assumption that recursion is universal. And the same goes for the argument against recursive syntax. It is clear that new data is necessary, and a new analysis of the language sensitive to the categories relevant to that particular language system. The remainder of this paper sets out to do exactly that. On the basis of my own fieldwork data I will revisit the analysis of *-sai* and will investigate whether recursive syntax is necessary and desirable in an analysis of Pirahã.

### 3. THE SUFFIX *-SAI* REVISITED

In everyday spontaneous speech embedded clauses are relatively rare (Karlsson 2007) so in order to test the exact behaviour of the suffix *-sai* as an embedder it was necessary to set up an elicitation task in addition to spontaneous language data.<sup>3</sup>

#### 3.1. *Background*

In the debate about Pirahã two claims have been made about the suffix *-sai*. First it has been said to be a nominalizer, and second it has been analysed as an obligatory marker of embedding in conditional clauses (Everett, 1986). For the purpose of this paper I will focus on the second question only.

In the experiment we used sentences consisting of two clauses in an asymmetrical relation. Whatever this relationship may be (causal, conditional, temporal) their common feature is that they consist of two parts with distinctive functions. In a sentence like:

(4) [*If it rains*] [*I don't go to the forest*]

The first clause is called the ANTECEDENT and the second one the CONSEQUENCE (Haiman 1978) since the second is a consequence of the first. In English one can recognize the antecedent by its syntactic marker 'if' and this stays the same when the second clause is fronted. What changes, however, is its discourse function. This is summarized in (5):

---

<sup>3</sup> The experiments were all designed and carried out together with T. Gibson and M. Frank, MIT.

- (5) (a) *[If it rains]*                      *[I don't go to the forest]*  
 ANTECEDENT/TOPIC                      CONSEQUENCE/COMMENT
- (b) *[I don't go to the forest]*   *[if it rains]*  
 CONSEQUENCE/TOPIC                      ANTECEDENT/COMMENT

This means that if *-sai* is a syntactic marker of antecedent or consequence we expect it to always attach to the same clause irrespective of its position in the sentence. However, if *-sai* is a marker of discourse functions such as TOPIC and COMMENT we expect its occurrence to be sensitive to linear position rather than to the syntactic function of the clause it attaches to.

### 3.2. Description of the task

To test these hypotheses, a sentence repetition task was designed. First a trial set was run with clearly ungrammatical sentences; the informants were asked to repeat grammatically to see whether they were participating and had understood the task. Then they were asked to do the same for a sentence consisting of an antecedent and a consequence with a clear semantic connection (this connection can be temporal, conditional or causal, but in all cases the relation is subordinate), as in (6).

- (6) *Pi-boi-bai-(SAI)*,                      *ti kahá-p-i-hiabi-(SAI)*  
 Water-come-INTENS-(SAI)   1   go-IMPPF-EP-NEG-(SAI)  
 'If it rains I don't go (to the forest)'

*-sai* was attached to either the first or the second clause in the input and in both cases the informants were asked to repeat the sentence grammatically. Any systematic changes would give information on whether *-sai* is obligatory in this type of sentence and in which positions it is acceptable. This would give information on whether it has a syntactic or discourse function.

The informants consisted of 7 men and 2 women. Every subject was presented with two instances where *-sai* was attached to the first clause (*piboibai*) and two where it was attached to the second (*kahápihiaba*).

### 3.3. Results

The total number of relevant responses from the informants was 39. Out of all responses *-sai* was absent in 6 cases, it was attached to one of the two clauses (either on *piboibai* or on *kahápihiaba*) in 24 cases and in 9 cases it was attached to both clauses.

These results are summarized in table 1 below. In this table the input is set out against the output. P's represent the *piboibai* (it's raining) clause, K's represent the *kahápihiaba* (I don't go) clause. The input column specifies whether *-sai* was attached to the P-clause or to the K-clause. The output columns give information on which clauses the suffix was attached to. PK represents the P-clause being output before the K-clause; and KP represents the K-clause being

output before the P-clause. In those cases where *-sai* occurred on one clause only it shows which clause the suffix was attached to.

**Table 1**  
Occurrence of *-SAI*

Input	Output							
	None		One				Both	
	PK	KP	P-SAI, K	P, K-SAI	K-SAI, P	K, P-SAI	P-SAI, K-SAI	K-SAI, P-SAI
P-SAI, K	1	2	2	0	7	0	5	0
P, K-SAI	0	3	1	0	14	0	4	0

On the basis of this information we can conclude that *-sai* cannot be an obligatory marker of embedding since in 6 cases it is not present in the sentence at all even though in the input sentence it was always there.

Our second hypothesis was that it should systematically occur on one of the clauses if its function were to indicate which clause is the embedded antecedent and which is the consequence. Since the table demonstrates that *-sai* can occur on either clause or even on both this analysis seems unlikely. This also excludes the possibility of being a correlative marker. Although there are cases where the suffix occurs on both clauses, this is only 9/39 of the cases. So on the basis of these results it is unlikely that *-sai* is an obligatory marker of embedding and also that its function is to specify which part must be interpreted as the embedded or main clause.

Now what is the standard way to express a sentence like *if it rains I don't go to the forest* in Pirahã? It seems that the default way of organizing such sentences is in an order different from the given input. Even though the input is always in PK order, the table shows that in 2/3 of the cases the output was KP. In a structure like that *-sai* can be left out, or it can be attached to the K (i.e. the first clause). *-sai* does not occur on the P part (i.e. the second clause) or on both clauses. In most cases where *-sai* occurs on the K-part, additional evidential suffixes are added and far most common way to repeat the sentence is:

- (7) *Ti kahá-p-i-hiabi-SAI-xáagahá pi-boi-bai-koí*  
 1 go-IMPF-EP-NEG-SAI-OBSERV water-come-INTENS-EMPH  
 'I don't go (to the forest), it rains'

*-sai* occurs on the P-clause as well but in such case the order is reversed, P being the first clause. In our sample we have no occurrences of *-sai* attached to the P-clause in second position. This is what we see confirmed in those cases where *-sai* occurs on both clauses. In such case the attested order of clauses is only PK and no KP cases have been recorded.

Summarizing we can derive the following constraints: K comes before P and *-sai* must come on the first item. But other word orders are possible provided that *-sai* is attached to the first item again or to both. What can be the function of such morpheme?

One plausible way to account for these data is to treat *-sai* as a marker of INFORMATION STRUCTURE rather than being a purely syntactic indicator of embedding. On the basis of Haiman's definition we can exclude the possibility of a syntactic marker of antecedent or consequence in conditional sentences. Since in that case it would have to be bound to one particular clause. Instead, linearity seems to be the crucial factor. *-sai* is attached to the first clause in a semantic dependency for canonical (KP) or non-canonical order (PK). In canonical order we can have *-sai* attached to first clause or it can be absent. For non-canonical order we must have *-sai* on the first clause, and the second one is optional (resulting in both marked if it occurs). In one case it occurs without *-sai*. However, on the basis of these experimental results it is hard to tell in what way exactly it marks the first clause. It could mark topic, old information, new information, it could stand for 'main message' whereas the second clause would be interpreted as a comment or afterthought leaving the kind of relation between the two clauses open to discourse and common sense (*I don't go to the forest if/since/when it is raining*).

Everett has analyzed *-sai* as a marker of old information (Everett 2007 and personal communication) but for this to be confirmed we need to know more about the semantic differences between the clauses and see them in natural discourse instead of experimental elicitation. Furthermore we need to test whether the combinations that were not attested in this experiment are truly ungrammatical. It could be connected with the semantics of this particular sentence, or could just not have occurred by chance.

Even though it is not clear what the exact function of *-sai* is, we are able to conclude on the basis of its distribution that an analysis of this suffix as a syntactic marker of embedding is not justified.

#### 4. CONCLUSION

The above study has shown that *-sai* most probably is not a syntactic marker of embedding even though more experiments and spontaneous language data are necessary to give a positive answer to what its function is instead. In order to provide a full answer to the question whether Pirahã has recursive syntax or not, clearly more areas need to be investigated part of which is currently being done (Stapert, Gibson and Frank in preparation). In addition to testing the grammaticality of sentences, the ungrammaticality of those clauses that do not occur in the sample needs to be confirmed as well before making definite conclusions.

However, this account has shown that independent collection and analysis of new data is essential in the solution of highly theoretical questions such as the one addressed in this paper. Despite the informational gaps that need to be filled in the



future, I have shown that new data are the only way to break the vicious circle of theoretical argumentation. They provide the key to turn this circle into an upward spiral where data and theory constantly feed into each other.

## REFERENCES

- Everett, Dan. 1986. Pirahã. In Desmond Derbyshire and Geoffrey Pullum (eds.) *Handbook of Amazonian Languages I*, Berlin: Mouton de Gruyter, 220-326.
- Everett, Dan. 2005. Cultural Constraints on Grammar and Cognition in Pirahã: Another Look at the Design Features of Human Language. *Current Anthropology* 76 (4), 621-646.
- Everett, Dan. 2007. Cultural Constraints on Grammar in PIRAHÃ: A reply to Nevins, Pesetsky, and Rodrigues (2007). <http://ling.auf.net/lingBuzz/000427> (March 2007).
- Haiman, John. 1978. Conditionals are topics. *Language* 54 (3), 564-589.
- Hauser, Marc D. Noam Chomsky & William.T. Fitch. 2002. The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science* 298, 1569-1579.
- Karlsson, Fred. 2007. Constraints on multiple initial embedding of clauses. *International Journal of Corpus Linguistics* 12 (1). 107–118.
- Nevins, Andrew, David Pesetsky, Cilene Rodrigues. Reassessment of Pirahã exceptionality. <http://ling.auf.net/lingBuzz/000411> (March 2007).
- Tomalin, Marcus. 2006. *Linguistics and the formal sciences: the origins of generative grammar*, Cambridge: Cambridge University Press.