A corpus analysis of Pirahã grammar: An investigation of recursion

Steven T. Piantadosi (Rochester), Laura Stearns (Wellesley)
Daniel L. Everett (Bentley), Edward Gibson (MIT)

Background: Recursion is proposed to be a core property of human language / Universal Grammar

- Hauser, Chomsky & Fitch (2002) argued that the grammars for all human languages are _recursive_, such that

"*there is no longest sentence (any candidate sentence can be trumped by, for example, embedding it in 'Mary thinks that . . .'), and there is no nonarbitrary upper bound to sentence length.*" (Hauser et al., 2002, p. 1571).

Background: Everett (2005)

- Everett (2005) argued that the grammar of Pirahã, an isolate language spoken by a small hunter-gatherer community in the Amazonian rain forest, is *not recursive*. (cf. Hale, 1980)

- Pirahã is proposed to have a regular grammar, generating *a finite language* (with no recursive rules)

Background: Everett (2005)
The Immediacy of Experience hypothesis

Everett (2005) makes two kinds of claims:

1. *Syntactic descriptive*: Pirahã lacks recursive syntactic structures

2. *Syntactic explanatory*: **Immediacy of Experience Principle**: "Pirahã culture constrains communication to non-abstract subjects which fall within the immediate experience of interlocutors.This constraint explains [the] very surprising features of Pirahã grammar and culture." (see Everett, 2012, for a derivation of the lack of recursion from the IEP via evidentials)

We will only discuss the descriptive claim here: whether or not Pirahã has recursive syntactic structures

Background: The definition of recursion

- Syntactic recursion: a syntactic category embedded within another of the same category, e.g., an S within an S; an NP within an NP ("self-embedding" in Nevins et al., 2009)

  - Conjunction: S → S Conj S; NP → NP Conj NP; VP → VP Conj VP
    - [[John and Mary] and Bill]
  - Clausal complements: S → NP VP; VP → V CP; CP → Comp S
    - [John thinks that [Mary said that [the girl cried]]]
  - Possessives: NP → NP 's N; NP → N
    - [[[[John]'s mother]'s brother]'s house]
  - Possessives: NP → NP of NP
    - [the house of [the brother of [the mother of [John]]]]

Background: some terminological confusion

1. The definition of *recursion*:

   (a) self-embedding of a syntactic category, thus allowing for an infinite number of sentences (Everett, 2005).

   (b) "Merge": effectively compositionality of any two syntactic elements (Nevins, Pesetsky & Rodrigues, 2009):

   - *"In a model with category-neutral Merge, however, a language that lacks recursion would be considerably more exotic. No sentence in such a language could contain more than two words. Pirahã is manifestly not such a language." (p. 366)*

   - Definition (b) is absurd (as Nevins et al. observe)
   - Everett is assuming definition (a).

Background: some terminological confusion

2. The term *Universal Grammar (UG):*

(a) Chomsky (in more recent discussions of Everett's work) and Nevins et al. assume that UG is whatever is biologically necessary to learn human language.

(b) Everett uses the term to refer to a specific claim about the nature of human language from HCF: that it allows recursion (self-embedding) in the syntax

- It makes no sense to falsify UG in the sense of (a): this is just a descriptive term
- Everett intends (b)

Goal of the current work:
To evaluate whether Pirahã grammar is recursive (allows self-embedding)

- Weakness of previous work:
  - No quantitative evaluations
    - Everett: Some structure looks to be non-recursive (finite language)
    - Nevins et al.: No, this structure looks recursive

    How to decide between the two?

Goal of the current work: To evaluate whether Pirahã grammar is recursive

- Methods:
  - Experimental elicitation: Trip to Pirahã winter, 2007
  - Analysis of corpus of stories collected by Steve Sheldon & Dan Everett

- Look for structures that are hallmarks of recursion (cf., Everett, 2010):
  - Relative clauses
  - Embedded clauses of saying / thinking
  - Embedded possessives
  - Conjunction, disjunction

Goal of the current work: To evaluate whether Pirahã grammar is recursive

- General Caveat:
    - Everett's descriptive hypothesis is that Pirahã *lacks* recursion
    - In general, it is impossible to prove a negative claim: if recursive structures are rare, it may be hard to find evidence of them
    - See Piantadosi & Gibson (under review) for quantitative statistical approaches to similar questions within typology, exploring proposed linguistic universals

The Pirahã

- Indigenous people of the Amazon basin

- Hunter-gatherers, little agriculture

- Approximately 750 people in ~6 villages

- Minimal contact or trade with outsiders

- Generally uninterested in outside cultures

Working with Pirahã:

- Arlo Heinrichs: 1959-1966

- Steve Sheldon: 1967-1976

- Dan Everett: 1977-2007

Experimental elicitation, January 2007 Edward Gibson, Michael Frank, Ev Fedorenko, Dan Everett

Strategy: case studies of recursive constructions

1. Possessive elicitation

   – *Kohoi's spouse's parent's dog*

   (NP → NP's N = tail recursion)

2. Relative clause elicitation

   – *The man who killed the jaguar fell down.*

   (S → NP VP; NP → NP COMP S = full/nesting recursion)

Language elicitations: Preliminary conclusions

- ## No evidence of recursive structures
  - Possessives: we couldn't elicit recursive possessive structures (task demands)
  - Relative clauses: Most elicited sentences in relative clause contexts were not embedded (*paratactic*)

- ## **Failed to falsify "no recursion" claim**
  - ➢ **But no strong evidence for the lack of recursion either**

Corpus analyses

- Corpus was collected by Steve Sheldon & Dan Everett

- 15 stories (14 by Sheldon, 1 by Everett):
  - Approximately 1000 sentences
  - Stories were transcribed morpheme by morpheme by Sheldon, with overall glossed meaning
  - Words tagged for part of speech (nouns verbs, pronouns, adjectives, determiners, etc.)
  - Some glosses, and parses added by trained undergraduate (L.S.), using primarily Sheldon's glosses.

Examples

 # 7.17.1: Hoagaixoxai said she is not giving birth.

 # Hoagaixoxai spoke.

```
(S (NPsubj hi/3/PRP )
    (VP gA/speak/VB -sai/old_info )
    (NPsubj hoagaIxOxaI/Hoagaixoxai/NNP ))
```

 # Hoagaixoxai spoke.

```
(S (NPsubj hi/3/PRP )
    (VP gA/speak/VB -sai/old_info ))
```

 # She is not giving birth.

```
(S (NPsubj i/3/PRP )
    (VP op/give_birth/VB -i/state -hiaba/neg -xaI/REL_CERT )))
```

Corpus analyses: Initial attempt

- Initial attempted analysis:
  - o Formalize several grammars and compare their performance on predicting POS tag sequences (following Perfors, Regier, & Tenenbaum 2011)
  - o Compare grammars with/without recursive rule expansions (a nonterminal that can yield the same category)

- Weaknesses:
  - o Not clear how to choose grammars to compare uncontroversially.
  - o Potentially hard to interpret the results

Corpus analyses: Shallow parsing

- Alternative analysis: **Shallow parsing** by hand
  - ➢ Subject noun phrase; Object noun phrase; Verb phrase; Possessives; Embedded sentences (if any); etc.
  - o Initially performed by co-author Laura Stearns
  - o Collaboration among all authors to evaluate potential recursive sentences (recently including Steve Sheldon).

- Examine parses for the presence or absence of certain <u>hallmark structures of recursion</u> (cf. Everett, 2010):

  - ➢ Relative clauses
  - ➢ Embedded sentences: complement clauses
  - ➢ Possessive structures
  - ➢ Conjunctions
  - ➢ Disjunctions

Corpus analyses: Shallow parsing

- Limitations:
  - Depends primarily on D.E. for linguistic judgments
    - The "no-recursion" claim is Dan's hypothesis, and he may have unconscious cognitive biases in support of this hypothesis (cf. Gibson & Fedorenko, 2010, 2011)
  - Many key examples are ambiguous with respect to their syntactic analysis
  - Lack of examples may result from low frequency structures (rather than ungrammaticality)

Corpus analyses: Searching for structures that are hallmarks of recursion

- ➤ Relative clauses
- ➤ Embedded sentences: complement clauses
- ➤ Possessive structures
- ➤ Disjunctions
- ➤ Conjunctions

Corpus analyses: Searching for structures that are hallmarks of recursion

Embedded clauses: RCs, complement clauses

To find strong evidence for recursive syntax from embedded clauses it is best to find examples with two dependents of one clause on the outside of another clause:

E.g., [SUBJ [embedded clause] main-verb]

No instances like this in the corpus, so no strong evidence of recursion

Corpus analyses: Relative clauses

- **0 sentences transcribed by either Sheldon or Everett as relative clauses**

- Back of the envelope calculation:
  - Relative clauses occur in English about 6 per 1000 words (Reali, 2007).
  - With ~3600 words, we should expect to see around 20.

(Limitation of this comparison: Need comparisons with matched discourse topic, socio-economic status, education, etc.)

Corpus analyses:
Potential complement clauses

- ~160 instances of "NP said / speaks" followed by a clause: e.g.,

2.3.1: I spoke. He is moving on the ground. TixohOI is crying.
"I said that TixohOI is crying on the ground."

```
(S (NPsubj ti/1/PRP )
   (VP igA/speak/VB xai/do/VB -ai/INTENS ))
(S (NPsubj hi/3/PRP )
   (VP (NPobj big/ground/NN ) a/move_vertical/VB -I/proximate )))
(S (NPsubj * )
   (VP is/cry/VB -Aaga/be -haI/REL_CERT )
   (NPtopic-subj TixohOI/TixohOI/NNP )))
```

The analysis of these clauses as embedded or matrix clauses is underdetermined by the evidence: There is no convincing argument yet to analyze these clauses as embedded.

Note: Semantic dependency does not entail syntactic dependency: e.g.,

"You drink you drive you go to jail." (Everett, 2010)

Corpus analyses: Possessives

- **We searched for instances of potential embedded possessives.**
  - **68 non-embedded possessives**

- Back of the envelope:
  - Possessives occur about 60 per 1000 nouns.
  - So we would expect 0.06 * 60 = 3~4 if possessives were themselves just as likely to be modified as nouns.

Corpus analyses: Possessives

- Based on Sheldon's translations: three potential recursive possessives

```
# 7.30.1: So Hoagaixoxai grabbed his wife's arm there, like this.
(S (NPsubj hoagaIxOxaI/Hoagaixoxai/NNP )
   (NPsubj hi/3/PRP )
   aigIa/thus/RB
   (VP (NPobj apipA/arm/NN giaIi/there/RB
            (NP-POSS kagi/expected_associate/NN
                     (NP-POSS hi/3/PRP)))
       abAa/grab/VB -hoi/go -haI/REL_CERT )
   agIiso/like_this/ )
```

Corpus analyses: Possessives

- But in this analysis, an adverbial "there" occurs between the head noun and following possessive, which doesn't happen in simple possessives
- Alternative analysis: repeated argument of the verb "grabbed": "arm", and "his wife": (there are many instances of repeated arguments of the verb elsewhere)

```
# 7.30.1: So Hoagaixoxai grabbed his wife there on the arm, like
   this.
(S (NPsubj hoagaIxOxaI/Hoagaixoxai/NNP )
   (NPsubj hi/3/PRP )
   aigIa/thus/RB
   (VP (NPobj apipA/arm/NN)
       giaIi/there/RB
       (NPobj kagi/expected_associate/NN
            (NP-POSS hi/3/PRP))
       abAa/grab/VB -hoi/go -haI/REL_CERT )
   agIiso/like_this/ )
```

(Everett proposed this analysis, and Sheldon agrees that it is better.)

Corpus analyses: Possessives

- All three transcriptions from Sheldon's corpus are similar
  - No strong instances of use of possessives beyond one level deep

- Weak evidence in support of a non-recursive analysis
  - It could be that the contexts in the corpus didn't support the use of possessives. But our experimental elicitations also didn't find the use of any recursive possessives.

- *Preliminary conclusion*: recursive possessives are at least rare in Pirahã

Corpus analyses: Conjunction / Disjunction

- There is no word in Pirahã for "and" (conjunction) or "or" (disjunction) that Sheldon or Everett has encountered

- However Sheldon's original glosses of many sentences have conjoined NPs or conjoined VPs:
  - E.g., *"Opisi and the snake climbed the tree"*

But none of the potential conjoined NP examples have conjoined NPs, just a translation that might have a conjunction in English

Most of these arise because of repeated or elaborations of arguments, or topics.

No strong evidence for recursive structures

Corpus analyses: Searching for structures that are hallmarks of recursion

- ➤ Relative clauses
- ➤ Embedded sentences: complement clauses
- ➤ Possessive structures
- ➤ Disjunctions
- ➤ Conjunctions

Initial conclusion: *No strong evidence for recursion among these categories.*

But there are many instances of repeated topics and elaborated NPs: *recursive structure*?

Corpus analyses: Topics / elaborated NPs

- Our analyses of possessives and potential conjunctions made us aware that there are many instances of repeated topics and elaborated NPs (~75 per 1000 sentences, excluding repeated pronouns): *often 2 or more full NPs that refer to the same verbal argument*
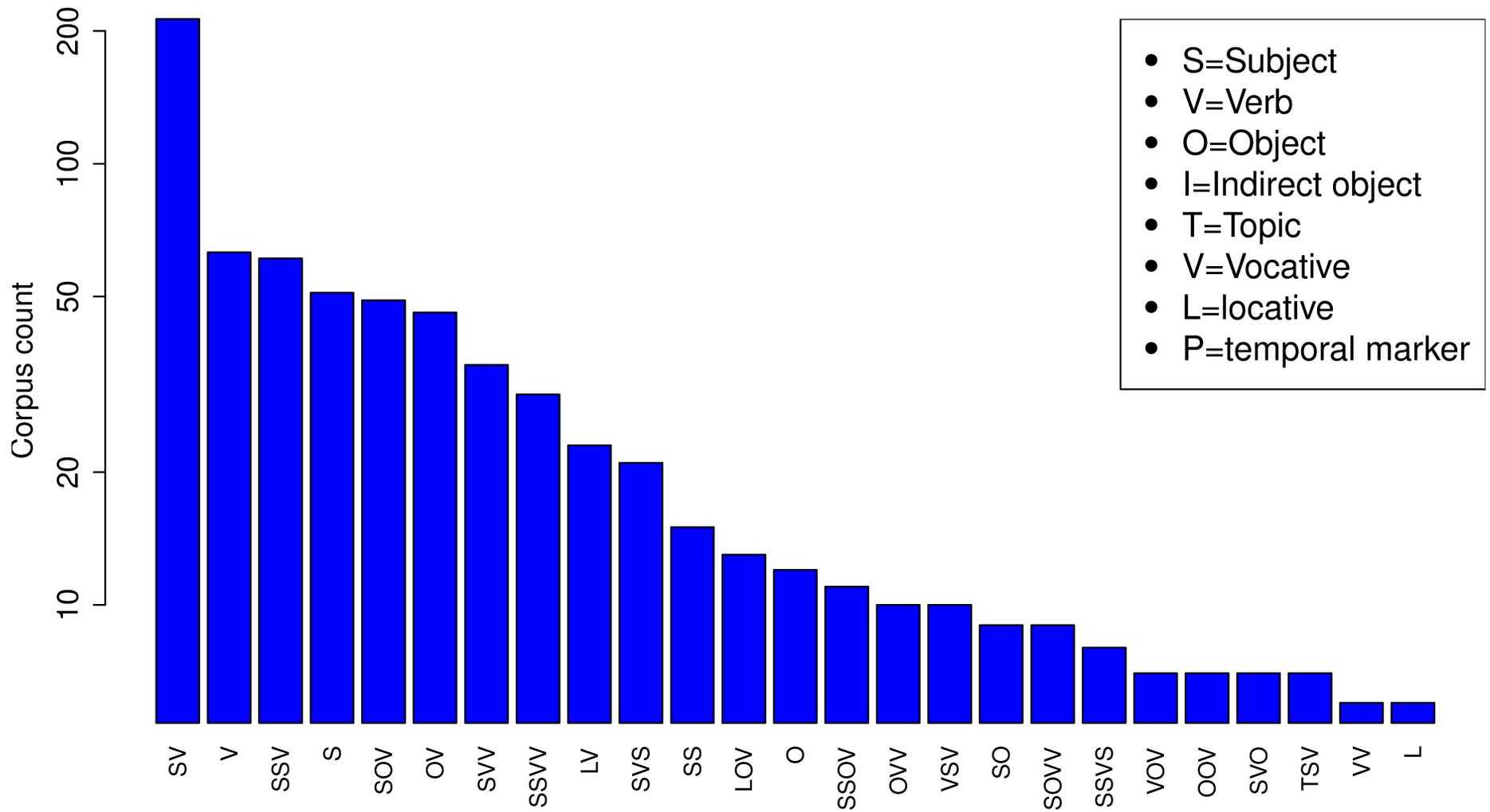
 # 12.17.1: The foreigner intends to sleep.

```
(S (NPsubj hi/3/PRP) aigIa/thus/RB

    (NPsubj aoI/foreigner/NN)

    (NPsubj batIo/Martins/NNP)

    (NPsubj ao/foreigner/NN) aigIa/thus/RB

    (VP ait/sleep/VB -aI/do -gA/cont -hoI/move -xiI/INTENT -haxA/certain))
```
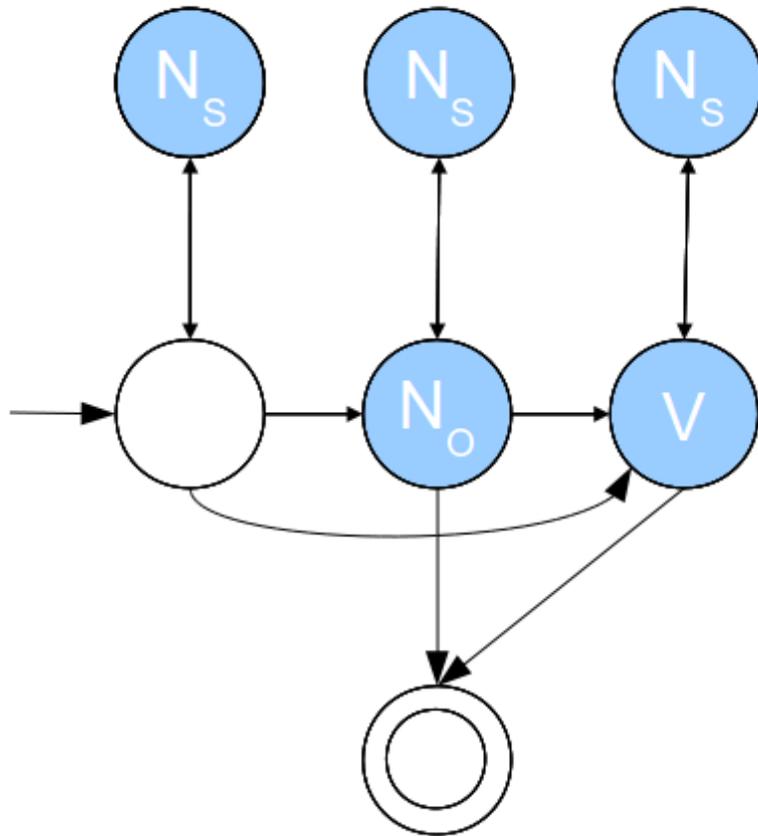
 # 10.2.1: His mother, Itaha, spoke.

```
(S (NPsubj hi/3/PRP)

    (VP gA/speak/VB xai/do/VB -sai/old_info)

    (NPsubj itahA/Itaha/NNP)

    (NPsubj ibigaIso/parent/NN oOxiai/there/RB))
```

Most frequent sentence structures

- S=Subject
- V=Verb
- O=Object
- I=Indirect object
- T=Topic
- V=Vocative
- L=locative
- P=temporal marker

Corpus analyses: Topics / elaborated NPs

S ➔ A B C
S ➔ B C
S ➔ C

A ➔ A N_subj
A ➔ N_subj

B ➔ N_obj
B ➔ A B

C ➔ Verb
C ➔ A C

Sample Finite State Transition Network & corresponding Regular Grammar that generate an arbitrary number of NP subjects or NP objects, either before or after the verb

Corpus analyses: Topics / elaborated NPs

- This is a recursive structure: it includes self-embedding of non-terminals below the sentence in order to generate sentences of potentially arbitrary length

- Caveat: these examples may be false starts followed by repairs
  - *Perhaps they shouldn't be represented in the core grammar of syntax*

- Future work should target these structures in elicitation experiments

Corpus analyses: Summary
The search for recursive structures

➢ Relative clauses: No examples
➢ Embedded sentences: complement clauses: No conclusive examples
➢ Possessive structures: No conclusive examples
➢ Conjunctions / disjunctions: No conclusive examples

Initial conclusion: *No strong evidence for recursion among these categories.*

➢ Topics / repeated arguments
  • *Possible evidence of recursive structure*

Potential explanations for differences in proportions of sentences with recursion across languages

- The Immediacy of Experience Principle (Everett, 2005)
  - Currently an *ad hoc* proposal: To be convincing, one needs to show correlations between the relevant linguistic features and the relevant cultural features across languages

- Perhaps there are cultural pressures for longer / shorter sentences
  - The existence of a writing system
  - Education / contact with other cultures
  - Communicative pressures: e.g., the need to communicate over distance in the forest
- If so, one of these could explain fewer recursive structures: structures with recursive rules tend to be longer, so there will be fewer of them

Concluding remarks

- We have provided suggestive evidence that Pirahã may have sentences with recursive structures

  - Evidence for the existence of recursion in Pirahã: *Topics / repeated arguments*

- Open question: Potential explanations for the rarity of many kinds of recursive structures

- Caveat: initial stages of project

Acknowledgements

- Thanks to:

  Steve Sheldon, for providing a corpus; and for providing judgments between glosses over email

  Mike Frank, Ev Fedorenko for collaboration on elicitation experiments in 2007