Experimental Evidence for Complex Syntax in Pirahã

September 2010

Abstract

The existence of complex clauses in the Amazonian language Pirahã has been controversially debated. We present a novel analysis of field data demonstrating the existence of complex clauses in Pirahã. The data concern the tone of the morpheme 'sai' and stem from a field experiment where a second language speaker of Pirahã presented sentences and Pirahã speakers were asked to correct them saying the correct sentence alound. The experimental items contained the morpheme 'sai' in two different clausal environments: a nominalizer and a conditional environment according to Everett's 1986 description. Our phonetic analysis shows an effect clausal clausal environment on the pitch of 'sai'. The native Pirahã speakers pronounced conditional 'sai' with lower pitch than nominalizer 'sai'. We show furthermore that the experimenters pitch on 'sai' shows the opposite pattern from that of the native Pirahã speakers and hence the Pirahã's pitch could not just have been copied. The effect of the clausal environment on the tone of 'sai' can be explained by a complex clause analysis of Pirahã, while existing alternative proposals do not explain the difference.

1 Introduction

Complex syntax is present in all well-studied languages spoken today for example in the form of complement clauses (Cristofaro 2003 and others). At the same time, human languages (or animal predecessors of them) must have existed that didn't exhibit any complex syntax. Hence at some point of human (pre-)history, complex syntax must have arisen. To find out when complex syntax arose would not only offer us a glimpse into prehistoric language, but would also allow us to understand better the biological foundation of complex syntax: If complex syntax was found to have arisen at a markedly later time from the biological prerequisites for it, the biological foundation of complex syntax is likely a by-product of other evolutionary developments (i.e. what Gould & Lewontin (1979) call a *spandrel*). Only if complex syntax had been present from essentially the time when the biological prerequisites for it were in place, could complex syntax be an adaptive trait of modern human. At present evidence on the historial genesis of complex syntax is for the most part circumstantial (see e.g. Givón 2009 for a review). Recently though Everett (2005, 2009) has claimed that a language still spoken today, namely Pirahã, lacks complex syntax entirely. Since the Pirahã tribe lives for the most part a hunter-gather lifestyle and the Pirahã have been shown to not have knowledge of numbers (Gordon, 2004; Frank *et al.*, 2008), it is initially plausible that their language may have not developed from an initial stage of language where no complex syntax is present.[1] Hence, it is interesting to evaluate Everett's claims carefully. Up to this point, however, the discussion has been limited by a lack of concrete relevant Pirahã data. In this paper, experimental evidence is presented that argues that complex syntax is in fact present in Pirahã contrary to Everett (2005, 2009).

Pirahã is spoken by an estimated 360 speakers in 2006 according to the Brazilian Government Health Organization Funasa 2006 as given on the website of the Instituto Socioambiental, www.socioambiental.org. The Pirahã tribe lives in the Brazilian state Amazonas in a remote location and are not connected to modern transport and communication networks. This has made it difficult to conclusively assess Everett's claims about the language and a number of recent papers debate the existence of complex syntax in Pirahã (Everett, 2005; Nevins *et al.*, 2009a; Everett, 2009; Nevins *et al.*, 2009b). One specific issue of debate is the question whether

---

[1]XXX(personal communication) points out that another possible scenario is that Pirahã culture at some point had numbers and other developments, but subsequently lost these (cf. Beller & Bender 2008). Specifically, there is historical evidence that the Pirahã have invaded a large area in the Amazon by force. According to XXX, this history suggests that the Pirahã had an advanced military organization that would have required the use of numbers. However even if the Pirahã indeed lost cultural developments including complex syntax, an absence of complex syntax at the present would still indicate that complex syntax conveys no direct evolutionary benefits to humans.

the suffixes *sai/saí* serve as syntactic markers of clausal embedding: In his first sketch of a grammar of Pirahã, Everett (1986) proposed that there are two morphemes *sai* and *saí* distinguished by a high tone on the final vowel. The morpheme *sai*, Everett analyzed *sai* as a nominalizer, while he analyzed *saí* as a conditional. This analysis evidently entails that complex clauses consisting of two nuclear clauses exist in Pirahã. More recently Everett (2005, 2009), however, suggests that Pirahã lacks embedding of clauses and propose two different analyses of *sai/saí*. Most recently, Everett (2009) claims that there is no underlying tonal distinction, but only one morpheme *sai* that is "occassionally" pronounced with high pitch. He proposes to analyze *sai* as a marker of old information, rather than a marker of sentential embedding.

In this paper, we present an analysis of field data collected by Eugenie Stapert, Daniel Everett, Michael Frank, and Ted Gibson in January 2007 concerning *sai* (Stapert, 2007). We report on an aspect of these data that Stapert doesn't consider in her 2007 work: Namely, we investigate whether the pitch of clause-final *sai* is determined by the environment the clause it is attached to occurs in. We show that if the *sai*-clause could plausibly serve as a semantic argument to the immediately preceding clause, *sai* is pronounced with lower pitch than if this isn't the case. These data are straightforwardly explained if the clause with *sai*-marking is subordinated to the other clause that determines the tone on *sai*, while it is not predicted by the analysis of *sai* by Everett (2009). Before we present the new data in detail, we summarize the three different analysis of *sai/saí* Everett has proposed in 1986, 2005, and 2009. We then describe the experimental data and show that the data argues in fact against all three of Everett's proposal. Finally, we show that the experimental data can be accounted for if Pirahã indeed possess complex syntax, though not in exactly the way Everett (1986) proposed.

1.1 Previous Analyses of '*sai*'

In three papers from 1986, 2005, and 2009, Everett proposes three different analysis of *sai/saí*. As mentioned above, Everett (1986) claims that there are two different morphemes *sai* distinguished by the lexical tone of the final phoneme: nominalizer and conditional.[2] The morpheme *sai* with a low tone on /i/, Everett analyzes as a nominalizer. Two example of its use are the following:[3]

---

[2](Everett, 1986, 264) raises the possibility that this may be the same morpheme with the tone distinction derived from other factors, but does not spell out an analysis.

[3]Cited examples follow the original sources with respect to transcription, glosses, and translation unless otherwise noted. Specifically, we represent the glottal stop as 'x' as in most of Everett's writings.

(1) a. hi ob-áaxáí   kahaí kai-**sai**   (Everett 1986, (232))
         3 see/know-INTNS arrow make-NOMLZR
         'He really knows how to make arrows.'

     b. tiobáhai hóoí ai-**sai**   xabahíoxoi (Everett 1986, (262))
         children bow make-NOMLZR incorrect
         'The children's bow making is incorrect.'

The morpheme *saí* with a high tone on the final segment, Everett (1986) analyzes as a conditional conjunction. Two examples of *saí* are quoted in (2).

(2) a. Pi-boi-hiab-i-**saí**   ti ahá-p-i-í
         water-come-NEG-EP-COND 1 go-IMPERF-PROX-COMPLETE-CERT(?)
         (Everett 1986, (239))

         'If it doesn't rain, I'll go.'

     b. gí-hi aho-a-áti   pii ap-ái   p-i-**saí**   baósaí
         2-3 speak-REMOTE-UNCERT water enter-ATELIC IMPERF-PROX-COND cloth
         ib-ái   t-op-í   (Everett 1986, (240))
         hit-ATELIC ITER-go-PROX
         'Tell him that if he goes to take a bath, to wash the clothes.'

In his second discussion of *sai*, Everett claims that Pirahã lacks embedded clauses. Everett (2005) claims there to be no tone distinction on *sai/saí* and analyzes it uniformly as a nominalizer as in the two following examples:[4]

(3) hi ob-áaxáí   kahai kai-**sai**   (Everett, 2005, 25a)
     he see-attractive arrow make-nominalizer

     'He knows how to make arrows well.'

(4) pii-boi-**sai**   ti kahapi-hiab-a   (Everett, 2005, 32)
     water-vertically move-nominalizer I go-negative-declarative

     'If it rains, I will not go.'

Everett claims that the nominalized constituent and the finite verb in both (3) and (4) are not part of one complex sentence, but are a form of *paratactic conjoining*.

   Everett (2009) proposes a third analysis of *sai*. He maintains that there is only one morpheme *sai*, but now proposes that it marks old information. One important prediction of this

---

[4]In some examples in Everett (2005), *sai* is glossed as *nominative*, where *nominalizer* is clearly intended. This is corrected in (3).

analysis that Everett notes is that in all examples where *sai* occurs the corresponding example without *sai* should also be grammatical. In support of this prediction, Everett gives the following example:[5]

(5) Pii-boi-baaí-hai.   Ti kahápi-hiaba. (Everett, 2009, (14))
     water-MOVE:DOWN-INTNS-INTENT 1 go.away-NEG
     'It is raining a lot. I will not go.'

Addressing his 1986 report of a tone difference between conditional and nominalizer *sai*, Everett claims that:

> The conditional sentence in [(5)] is not marked by *sai*, but by context and (usually) rising intonation. Rising intonation is commonly used whether or not *sai* is present. The latter fact means that occasionally a *sai* 'conditional' will have higher pitch, though a *sai* 'nominalizer' will not. But the pitch difference is a function of conditional intonation. *sai* itself is marked (underlyingly) by low tone, so there is not evidence for two *sai*s, [. . .] (Everett 2009, p. 412)

In the following section, I report on elicited data that argues against all three of Everett's proposals.

2 Experimental Data

2.1 Method and Subjects

The data we report on were not gathered by us, but during a field expedition in January 2007 led by Daniel Everett and comprising a team of five linguists and psychologists, Daniel Everett, Edward Gibson, Jeanette Saekel, Eugenie Stapert, and Michael Frank. A subset of them conducted an experiment in the field using a sentence correction task. Stapert (2007) reports her analysis of the results of this study. In individual sessions, nine Pirahã speakers were instructed to repeat utterances that were presented to them by a Pirahã-English interpreter. Some of the utterances the interpreter presented were ungrammatical; specifically, the word order was incorrect. The Pirahã had been instructed to repeat the interpreter's utterances aloud but as they did so, make any corrections of the Pirahã that they deemed necessary. Since the interpreter was not a native speaker of Pirahã, this task was natural for the Pirahã speakers – it

---

[5]Female speakers of Pirahã generally don't distinguish between the phonemes /s/ and /h/ as Everett (1986) reports. Since (5) contains *-hai*, it only makes Everett's point if the sentence is uttered by a male.

was essentially the same task, as they would perform when helping a second language learner correct her or his errors. Nine individual sessions with nine different Pirahã speakers were conducted. Each session was audio-recorded in its entirety. Eugenie Stapert transcribed all nine recorded sessions and we rely on her transcription in the following.

In this note, we do not report on the aspects of these data that the experiments designers had in mind (see Stapert 2007). Instead we set out to compare the production of 'sai' in two conditions. The first of these, we call the nominalizer condition. In this subjects were asked to repeat the Pirahã sentence in (6) or a word order variant of it. We give the two translations predicted for the sentence by Everett's analyses from 1986 and 2009 respectively.

(6)   Hiaitíihi xobáaxái kaíi   kai-sai
       Pirahã   good at   house make-SAI

       'The Pirahã are good at making houses.'

       'The Pirahã are good. They make houses.'

The second condition, we refer to as the conditional condition and the sentence subjects were asked to repeat is shown in (7). Again, the first translation is that predicted by Everett's 1986 analysis, while the second is the one predicted by the 2009 analysis.[6]

(7)   Pi-boi-bai-sai   ti kahápihiaba
       rain-MOVE:DOWN-INTNS-SAI 1 go-NEG

       'If it rains, I'm not going.'

       'It is raining. I'm not going.'

In our study, we extracted all and only productions of (6) and (7) in the grammatical word order shown here. We looked at both the productions by the Pirahã and those of the interpreter. To test for the tone of *sai*, we performed pitch extraction with the PRAAT software and then computed the maximum pitch on the syllable *sai*.

The three analysis predict different tone markings for the occurrences *sai* in (6) and (7). According to the 1986 analysis, both versions of 'sai' are sentence connectives and the tone of 'sai' depends on the role the 'sai'-marked clause has in the embedding sentence. In (6), the 'sai'-sentence could be a nominal argument or adjunct of the preceding clause assuming that the verb *xobáaxái* ('good at') can optionally combine with an extra nominal. In (7), however,

---

[6]The 1986 analysis would, in fact, require a high tone on *sai* for the sentence to have the conditional interpretation, though it is shown in (i) without tone on *sai*. As the results below show, the sentences were in fact presented with significantly higher tone on *sai* in the conditional condition than the nominalizer condition.

there is the 'sai'-clause precedes a second clause which has the verb *kahápihiaba* ('going'). Because of the difference in position and verb, an analysis as conditional could be pragmatically more plausible in (7). Hence, we may expect that the nominalizer condition carry a low tone in contrast to a high tone for the conditional condition on Everett's 1986 analysis. In contrast, the 2005 analysis of Everett doesn't lead us to expect a tone difference between the two conditions. Finally, the 2009 analysis makes a weaker prediction than the other two analysis: it predicts that there may be a tone difference between the two conditions with only the conditional condition exhibiting high tone, but this should not be obligatory.

2.2   Results

Since the recordings contained environmental sounds, pitch extraction was not possible from all relevant extracted recordings. Furthermore, Pirahã speakers in six cases omitted 'sai' from their production (Stapert, 2007). For the interpreter (I) and the Pirahã (P1–9) for which there was sufficient data, the number of pitch measurements obtained from the recordings is shown in the table below. For two of the Pirahã participants (P3 and P6) in the experiment, no pitch measurement could be made in at least one of the two conditions. Therefore, their data is excluded from the analysis in the following. Of the remaining subjects, two (P8 and P9) are female.

speaker	Cond	Nom
I	26	11
P1	6	2
P2	3	7
P4	4	2
P5	6	2
P7	3	5
P8	3	1
P9	3	4

The graph in figure 1 shows the distribution and means of pitch maxima on 'sai' in the two conditions for the seven Pirahã participants and also the interpreter that conducted the experiment.

insert figure 1 about here

7

As can be seen the mean pitch maximum in the nominalizer condition is greater than that in the conditional condition for six out of seven Pirahã participants. This difference is significant at the .05 level for participants P2 and P5 (p-value = 0.0016 for P2 and p-value = 0.036 for P5 according to Welch two-sample t-test).

The average pitch maximum on 'sai' across Piraha speakers was 190.6 Hz. (SD = 47.0) in the conditional condition and 211.6 Hz. (SD=43.1) in the nominalizer condition. A two-factor ANOVA was computed to evaluate the effect of the conditional and nominalizer condition on the maximum pitch of *sai* as well as that of subjects. The effect of subjects was significant at the maximum $p < .00001$ level ($F(6,1) = 50.4, p < 2.5e - 16$), which is unsurprising given that subjects differed by gender. In addition, we find a significant effect of condition on the maximum pitch at the $p < .0005$ level ($F(1,6) = 17.4$, $p = 0.00014$). The interaction of subject and condition was not significant at the $p < .05$ level ($F(6,1) = 1.8$, $p = 0.128$).

The comparison with the interpreters pitch maxima shows that the difference between conditions in the Pirahã data could not be due to copying from the interpreter: Namely, we observe a difference in the opposite direction for the interpreter: the average maximum pitch in the conditional condition is 11 Hz. higher than that in the nominalizer condition. A Welch two Sample t-test was conducted to test the effect of condition on maximum pitch of 'sai' in the interpreters speech, and showed the condition to be significant at the $p < .05$ level (t(29.7) = -2.36, p = 0.025). Therefore, if the Pirahã had copied the tones of the interpreter, they should have shown the reverse pattern of pitch maxima from the one we actually observed. Taken together, these results suggest that all the Pirahã participants pronounce 'sai' with different tone depending on whether a conditional or a nominalizer interpretation of 'sai' is pragmatically plausible.

3 Discussion and Conclusions

None of the three analyses by Everett (1986, 2005, 2009) predicts the tone difference between the conditional and nominalizer conditions we observe: The 1986 analysis predicts an obligatory tone difference, but in the opposite direction. The 2005 and 2009 analyses predict either no tone difference or an optional tone difference in the opposite direction from the one observed.

We take our evidence to indicate that the tone difference is obligatory with low tone *sai* occurring in the conditional condition and high tone *saí* occurring in the nominalizer condition. This assumption explains that the Pirahã speakers introduced the observed tone distinction

even though they received input with a tone difference in the opposite direction.

What can account for the distribution of *sai* and *saí* in Pirahã? There are numerous differences between the two conditions, and it is difficult to rule all of them out as relevant factors without further research. Note, however, that it seems unlikely that the factors internal to the nuclear clause that *sai/saí* is suffixed to account for the difference: In both cases, the coda of the preceding syllable is *ai* with low tone, hence, morphophonological factors are unlikely to account for the difference between *sai* and *saí*. And though the conditions differ in many grammatical properties of the clause *sai/saí* attaches to (animacy of the subject, transitivity, ...), Everett (1986: 288–9) describes Pirahã as a language where postverbal suffixes exclusively relate to the entire event described by the clause they attach to, not its internal make up. Hence, sentence internal grammatical differences are not expected to determine the distribution of *sai* and *saí*.

From this reasoning, it follows that the relationship of the *sai/saí*-clause to the second nuclear clause most likely determines the choice between *sai* and *saí*. What the relevant factor is is impossible to tell on the basis of just two conditions. Two possible hypotheses that seem worth pursuing are the following: 1) A lexical explanation: There are two morphemes, conditional *sai* and nominalizer *saí*, distinguished by lexical tone. 2) A prosodic explanation: Sentence prosody could explain the tone distinction we observed since in the data we considered *sai* always occurred utterance-medially, while *saí* occurred utterance-finally. Specifically, such an account would postulate that there is only one morpheme *sai/saí* but it is pronounced with low tone if it occurs in the middle of a complex clause and with high tone if it occurs on the right edge of a complex clause. In the absence of further evidence we cannot determine which of these two hypotheses is correct. Since there are clearly systematic discrepancies between the tones as transcribed by Everett in published work and the evidence discussed here, we must leave the matter open at this point.[7]

Independently of the specific factor determining the choice between *sai* and *saí*, the result provides evidence for the existence of complex clauses in Pirahã. Namely, if there was no grammatical relationship between the two nuclear clauses, it would not be expected that the relationship of the two nuclear clauses should determine the tone of *sai/saí*.

We conclude therefore that the new experimental evidence argue for the existence of complex clauses in Pirahã. Previously, Nevins *et al.* (2009a) have argued towards this conclusion

---

[7]Note also that though the tone difference is statistically significant overall, most Pirahã speakers do not consistently produce a measurable difference in maximum pitch of *sai/saí*. Hence, it is not possible to reliably classify individual occurrences of *sai/saí* in other recordings as carrying high or low tone.

with several very interesting arguments. However, the evidence Nevins *et al.* (2009a) have used was not obtained directly from recordings of Pirahã, but from descriptions of Pirahã by Everett (1986) that Everett (2009) now claims to have been mistaken in crucial respects. Since our argument relies on recorded data obtained in a controlled setting, we hope that disputes about the validity of the data will not arise to the same extent. If so, our findings also provide an argument for the application of more rigorous methodology in field research. Specifically, such methodology should be applied where the findings are particularly surprising or important.

Bibliography

Beller, Sieghard, & Andrea Bender. 2008. The limits of counting: Numerical cognition between evolution and culture. *Science* 319.213–215.

Cristofaro, Sonia. 2003. *Subordination*. Oxford, UK: Oxford University Press.

Everett, Daniel. 1986. Pirahã. In *Handbook of Amazonian Languages*, ed. by Desmond C. Derbyshire & Geoffrey K. Pullum, volume I, 200–325. Berlin, Germany: de Gruyter.

——. 2005. Cultural constraints on grammar and cognition in Piraha: Another look at the *Design Features* of human language. *Current Anthropology* 46.621–646.

——. 2009. Pirahã culture and grammar: A response to some criticisms. *Language* 85.405–442.

Frank, Michael C., Daniel L. Everett, Evelina Fedorenko, & Edward Gibson. 2008. Number as a cognitive technology: Evidence from piraha language and cognition. *Cognition* 108.819–824.

Givón, T. 2009. *The Genesis of Syntactic Complexity*. Amsterdam, Netherlands: John Benjamins.

Gordon, Peter. 2004. Numerical cognition without words: Evidence from Amazonia. *Science* 306.496–499.

Gould, Stephen J., & Richard C. Lewontin. 1979. The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proceedings of the Royal Society of London. Series B, Biological Sciences* 205.581–598.

Nevins, Andrew, David Pesetsky, & Cilene Rodrigues. 2009a. Pirahã exceptionality: A reassessment. *Language* 85.355–404.

——, ——, & ——. 2009b. Evidence and argumentation: A reply to Everett (2009). *Language* 85.671–681.

Stapert, Eugénie. 2007. Linguistic theory and fieldwork in interaction: the case of Pirahã. In *Proceedings of Conference on Language Documentation and Linguistic Theory*, ed. by Peter K. Austin, Oliver Bond, & David Nathan, 235–242. London: SOAS.
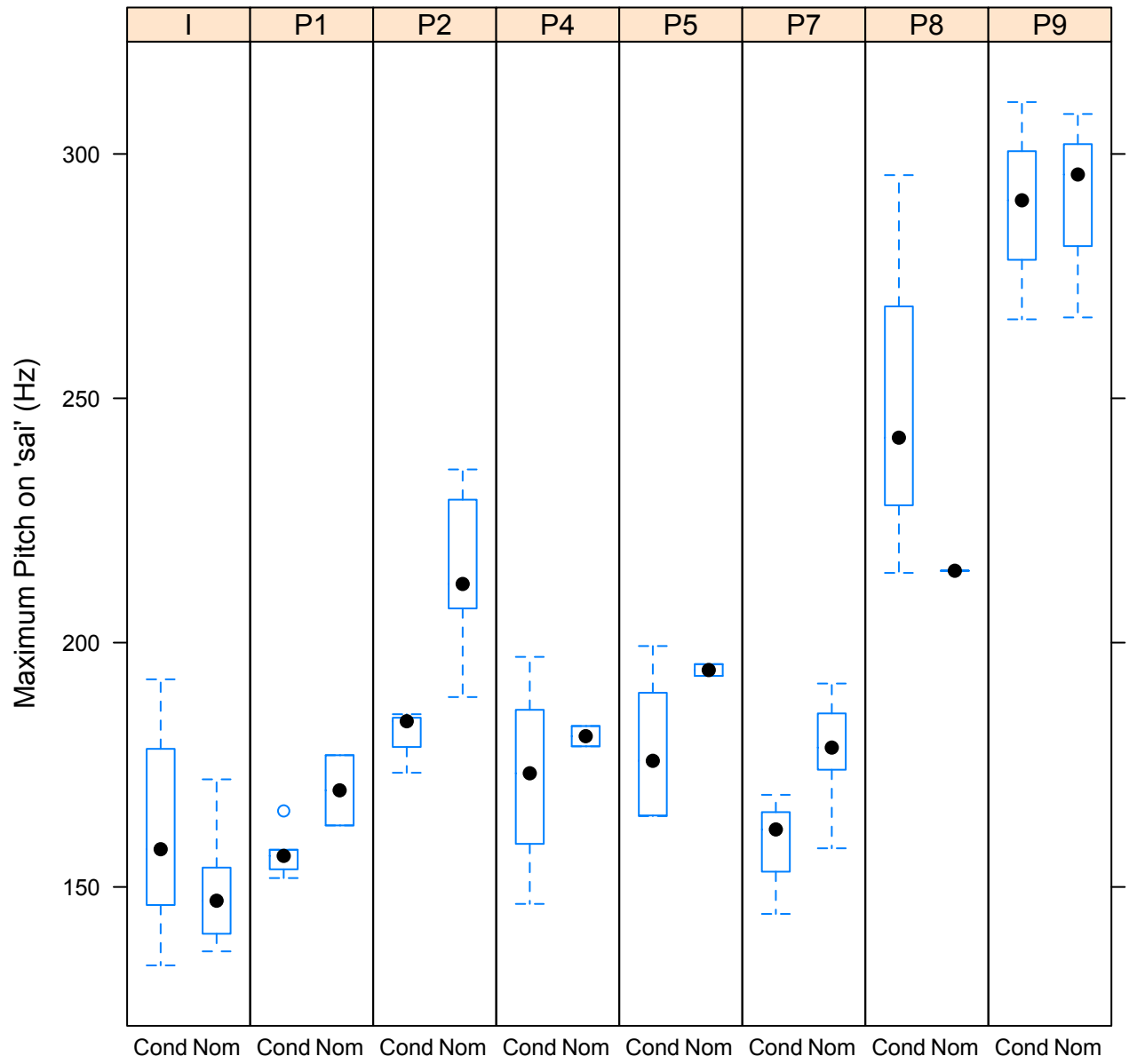
Figure 1: Maximum pitch on 'sai'