Brazilian Portuguese and the Null Subject Parameter

I. THEORETICAL FRAMEWORK

In this paper I will examine the syntactic properties of Brazilian Portuguese with respect to the Null Subject parameter, focusing in particular on 'long' wh-movement of the subject. The theory of grammar within which this research has been carried out is that of Government-Binding (henceforth, GB) as developed by Chomsky (1980, 1981) and others. The core grammar is assumed to be organized as in (1):

\[
\begin{align*}
\text{D-structures} & \quad \text{S-structures} \\
\phantom{\text{D-structures}} & \downarrow \\
\text{Phonological component (PF)} & \quad \text{Logical form (LF)}
\end{align*}
\]

At D-structure all argument positions (A-positions) are filled; these representations are then mapped onto S-structures via the syntactic transformation rule \text{move} \ a. At LF the scope of quantifiers (including negation and wh-operators) is established via the rule of quantifier raising (QR) and perhaps also wh-raising (Aoun, Hornstein and Sportiche, 1981), while in PF are found rules such as stylistic movement and phonological rules proper.

Within the GB framework, much investigation has focused on the null subject languages (henceforth, NSLs), that is, languages which allow subject position to be filled by a phonetically null element, as in the Spanish example in (2):

(2) Salió.
'S/he left'

GB analyses of NSLs have identified the following as the cluster of properties associated with the null subject parameter:

This paper is a revised and condensed version of a paper written in Spring 1982 with the title "Brazilian Portuguese and the Pro-Drop Parameter." I am grateful to Carlos Quicoli for having spent many hours as both informant and linguist in discussing the original paper with me. As the former he was very patient in answering innumerable questions about Portuguese sentences and their interpretations (grammaticality judgements in this paper except where otherwise noted are his) and as the latter he was an excellent sounding board for ideas. Thanks are also due to Tim Stowell for his observations and corrections on the earlier draft, some of which have worked their way into the present version. Estou muito obrigada.
These properties are illustrated in Spanish in (4):

(4)  a. [e] Protestó contra las armas nucleares.  
     ‘S/he protested against nuclear arms’

        b. Lo hizo Ana. 
           ‘Ana did it’

        c. ¿Quién dijiste que llegó ayer?
           ‘Who did you say that arrived yesterday?’

        d. María, dijo que [e] iba a protestar la decisión.  
           ‘María said that (she) was going to protest the decision’

        e. (*El) Parece que los contras tienen el apoyo de la CIA. 
           ‘It seems that the contras have the support of the CIA’

The property of greatest interest has been (3(iii)). The *[that-trace] filter
was originally proposed by Chomsky and Lasnik (1977) to account for
the ungrammaticality of sentences such as (5a) in English, whose S-structure
is assumed to be as in (5b):

(5)  a. *Who did you say that arrived yesterday?

        b. [§ who[i [§ did you say [§ t[i that [§ t[i arrived yesterday]]]]]

Within the GB framework, the ungrammaticality of (5a) is attributed to
the Empty Category Principle (ECP), given in (6):

(6)  [NP e] must be properly governed. (where [e] = NP-trace or wh-
     trace)

Government and proper government are defined in (7) and (8):

(7)  In [γ...β...α...β...] α governs β iff
     (i) α = X° or α is co-indexed with β
     (ii) where φ is a maximal projection, if φ dominates β then φ
domina tes α
     (iii) α c-commands β. 
         (Chomsky, 1981)

(8)  α properly governs β iff
     (i) α governs β and 
     (ii) α is lexical.

Finally, c-command is defined in (9):

(9)  α c-commands β iff the first branching node dominating α also
dominates β.
Now consider the S-structure (5b) as given in the tree diagram (10):

\[
\begin{array}{c}
\text{(10)} \\
\text{\[ S \\
\text{COMP} \]}
\end{array}
\]

The presence of the overt complementizer *that* creates a branching COMP, and thus the first branching node dominating \([t']\), the trace left in the lower COMP by successive cyclic wh-movement, does not also dominate the trace in subject position. Hence the c-command requirement for government is not met and the subject trace is not governed by \([t']\), which is the only potential proper governor. Note that the INFL node is assumed to govern subject position, assigning nominative Case to the subject. However, INFL is not lexical and so cannot properly govern.\(^4\) Now suppose there is no overt complementizer present. Then the first branching node dominating the trace in COMP is \(\bar{S}\), which also dominates subject position. Then the trace in COMP may properly govern the trace in subject position and the ECP is satisfied, accounting for the grammaticality of (11):

(11)  Who did you say arrived yesterday?\(^4\)

Suppose now that the S-structure of the Spanish example in (4c) is identical to that of the English example in (10). Then the trace in COMP should fail to c-command the trace in subject position and the latter should violate the ECP. However, the sentence is grammatical, as it also is in Italian, another NSL. Further, the optionality of the overt complementizer which provided the 'escape hatch' for the English example is unavailable in these languages:

(12) a. *¿Quién dijiste llegó ayer?  
b. *Chi hai detto è arrivato ieri?

Initial analyses of sentences such as (4c) within the GB framework attempted to attribute the apparent ability of NSLs to escape the *[that-trace] filter (more properly, to escape ECP violations) to the ability of INFL to act as a proper governor in these languages (cf. Taraldsen, 1978). The intuitive notion was that those languages which allow a null subject have a characteristically rich verbal morphology marking inflectional endings, and so INFL in these languages was in some sense 'lexical'. Thus INFL in Spanish but not in English is a potential proper
governor for a trace in subject position, making the presence or absence of an overt complementizer irrelevant.

Rizzi (1980) showed, however, that this account of NSLs was incorrect. Specifically, he demonstrated that in Italian interpretation of the scope of negative elements is constrained by the ECP in the same manner as in non NSLs. In Italian, as in Spanish, postverbal negative elements require the negative particle _non_ before the verb, while preverbal negative elements appear without _non:_

(13) a. Mario *(non) ha visto _nessuno._
   ‘Mario neg has seen no one’

b. _Nessuno_ *(non) ha visto Mario.
   ‘No one has seen Mario’

c. *(Non) ha telefonato _nessuno._
   ‘No one has telephoned’

Following work by Kayne (1979) on scope properties of negation in French, Rizzi proposes that elements such as _nessuno_ are negative polarity elements which must appear in the c-command domain of a negative operator and which undergo QR at LF, thus being linked to this negative operator. In addition, he proposes a mechanism by which such elements incorporate a negative operator when they appear in preverbal subject position:

(14) _nessuno_ → [+neg] when c-commanded by VP

At LF, then, (13a,b) appear as in (15):

(15) a. [non + nessuno] [Mario ha visto _x_]
   ‘For no _x_, Mario has seen _x_’

b. [neg + nessuno] [x ha visto Mario]
   ‘For no _x_, _x_ has seen Mario’

Now consider the following sentences:

(16) a. Non pretendo che tu arresti _nessuno._
   ‘I do not require that you arrest no one’

b. Non pretendo che _nessuno_ ti arresti.
   ‘I do not require that no one arrest you’

The LF of (16a) is as in (17):

(17) [non + nessuno] [pretendo [che [tu arresti _x_]]]
   ‘For no _x_, I require that you arrest _x_’

_A priori_, it would seem that (16b) has two possible LF representations, one in which _nessuno_ is interpreted with the _non_ of the main S, and one in which _nessuno_ undergoes (14). However, only the latter is a possible LF representation:
(18) a. *[non + nessuno [pretendo [che [x_i ti arresti]]]]
   'For no x, I require that x arrest you'

   b. [non [pretendo [che [neg + nessuno [x_i ti arresti]]]]]
   'I do not require that for no x, x arrest you'

Note that x, the variable left by the application of QR to nessuno, is subject to the ECP. If INFL is a proper governor in Italian, then (18a) should be a possible LF representation. However, the representation is excluded, indicating that the variable in subject position has no proper governor and that INFL in fact cannot properly govern.

The negative interpretation facts led Rizzi to a reexamination of the apparent ECP violations in Italian. He concluded that this property was in fact directly related to another property, namely, free inversion of the subject. Postverbal subjects are assumed to be adjoined to VP (I will here ignore the postverbal subjects of ergative verbs, cf. Burzio (1981)). Following Chomsky (1982), preverbal subject position is assumed to be occupied by expletive [pro], where [pro] is the empty category containing the features [+pronominal — anaphor]. [pro] may have a definite reference, as in (19), or it may be expletive, as in (20), corresponding to English expletive it:

(19) [pro] estamos cansadas
   'are tired'
   +1 pers
   +fem

(20) [pro] resulta que decidieron a favor nuestro.
   'It turns out that they decided in our favor'

Thus the S-structure of the Italian example (21) is (22):

(21) Ha telefonato Mario.
   'Mario has telephoned'

(22) 

Rizzi proposed that in the case of long wh-movement of the subject, the subject is extracted from postverbal rather than preverbal position. Assuming an extended definition of c-command by which an element may c-command, and hence govern, up through its maximal projection, including a maximal projection created by adjunction, the verb may then
properly govern the trace of the subject. The trace in COMP left by successive cyclic movement is thus irrelevant. Rizzi in fact gives evidence related to ne-cliticization in Italian which shows that even short wh-movement of the subject is from postverbal position.

Let us now return to the properties of NSLs as given in (3). Property (iii) has been linked to property (ii). Properties (i) and (v) are different manifestations of the empty category [pro]. I have said nothing as to what licenses [pro] in Spanish and Italian but not in English and French; I will address this question below. Property (iv) may be considered a manifestation of the 'Avoid Pronoun' principle, informally stated in (23) (cf. Chomsky, 1981):

(23) Avoid phonetically realized pronouns when [PRO] or [pro] is available.

At this point we are ready to examine the facts of Brazilian Portuguese.

II. BRAZILIAN PORTUGUESE

Brazilian Portuguese is commonly assumed to be an NSL like Spanish and Italian. As such it appears to exhibit the cluster of properties given in (3), as shown by the examples in (24):

(24) a. Saiu.
'S/he left'

b. Correm as crianças rapidamente.
'The children run fast'

c. Quem você disse que chegou ontem?
'Who did you say that arrived yesterday?'

d. João disse que [e] tinha de comprar um casaco.
'João said that he had to buy an overcoat'

e. (*Ele) Parece que João admira Gabriela.
'lt seems that João likes Gabriela.'

((24a,d) are from Chao (1980); (26b) is from Zubizarreta (1982).)

However, sentences such as (24b) do not have the same interpretation as sentences with inverted subjects in Spanish and Italian. An inverted subject in Portuguese is only interpretable as a focused subject, and even then is only marginally acceptable with transitive verbs:


b. Não leem estes livros AS CRIANÇAS.
'THE CHILDREN don't read these books'

Only the subjects of passives and of ergative verbs in the sense of Burzio (1981) can freely appear postverbally without stress:
(26) a. Foram omitidos pormenores importantes pelo autor deste relatorio.
   'Important details were omitted by the author of this report'
   b. Rebentou a guerra em janeiro.
   'The war broke out in January'

((25) and (26) are from Zubizarreta (1982) with Brazilian Portuguese judgements.)

Subject inversion in embedded clauses appears to be even more limited than in matrix clauses:

(27) a. "Corria Iara pela praia.
   'Iara ran along the beach'

b. *Mário disse que corria Iara pela praia.
   'Mario said that Iara ran along the beach'

Recall that in Rizzi's analysis free subject inversion is crucial to explaining the apparent ECP violations in NSLs. Further, while Portuguese lacks free subject inversion, particularly in embedded clauses, it does appear to permit long wh-movement of the subject from embedded clauses, as in (24c). We could retreat to the position of the early GB analyses and claim that INFL is a proper governor in Portuguese. But there are examples of ECP effects which falsify this stand.

It appears that in Portuguese S is a (weak) bounding node for subjacency, and thus wh-island constraints are generally respected:

(28) *A quem você disse o que José deu?
   'To whom did you say what José gave?'

However, violations of subjacency involving extraction of relative pronouns from embedded questions are acceptable to varying degrees:

(29) a. ?O aluno a quem não sei que coisas êles tinham ensinado estava totalmente confundido.
   'The student to whom I don't know what things they had taught was totally confused'

b. O seu irmão, a quem não sei que estorias êles contaram, estava muito preocupado.
   'Your brother, to whom I don't know what stories they told, was very worried'

Violations involving subject extraction, nevertheless, are ungrammatical:

(30) *O diretor que você me perguntou que filme rodou em Portugal tem muito talento.
   'The director that you asked me which movie filmed in Portugal is very talented'
The relevant S-structure is (31):

(31) O diretor \([\text{s} \text{ que}_i \text{ [s você me perguntou [s que filme] [s t_i rodou t_j em Portugal]]}]\) tem muito talento

\([t_i]\) is separated from its coindexed antecedent by \(\text{s}\) and so is not properly governed by it. On the assumption that successive cyclic movement cannot pass through a COMP already filled by a wh-element, there is no coindexed trace for \([t_i]\) in the lower COMP. Thus INFL is the only potential proper governor. The ungrammaticality of (30) indicates that INFL fails to properly govern the subject trace. Thus in Portuguese, as in Spanish and Italian, INFL is not a proper governor.

There are two possible analyses which suggest themselves to explain the contradictory Portuguese facts. It may be the case that free inversion and the possibility of null subjects are in fact two different parameters. This is the approach taken by Safir (1982). The second possibility is that Portuguese is in the process of becoming a non NSL. This then necessitates a different explanation for the long wh-movement cases.

Let us explore the second possibility. Although INFL in NSLs is no longer assumed to be a proper governor, there is still a (rather unarticulated) requirement that INFL be ‘rich’ enough to identify the features of the missing subject, thus allowing the language the option of a non-phonetically realized pronoun, or \([\text{pro}]\). This of course is the intuition of traditional grammars. Thus we find the following in the Esbozo de una nueva gramática de la lengua Española (Real Academia Española):

Las desinencias personales de la conjugación española son tan claras y vivaces que casi siempre hacen innecesario y redundante el empleo del pronombre sujeto . . . (p. 421)

Note also that French, as a descendant of Latin, was once a null subject language and ceased to be so when phonological changes neutralized the inflectional endings in the spoken language.

In light of the above, let us examine the actual situation in Portuguese. Brazilian Portuguese originally had a complete verbal paradigm of six distinct forms, three persons each in the singular and the plural. Thus the present tense conjugation of falar ‘to speak’ was as in (32):

(32) eu falo ‘I speak’ nós falamos ‘we speak’
    tu falas ‘you (sg) speak’ vós falais ‘you (pl) speak’
    \{êle\} fala ‘s/he speaks’ \{êles\} falam ‘they speak’
    \{ela\} fala ‘s/he speaks’ \{elas\} falam ‘they speak’

Both the singular and plural second person forms have dropped out of use in speech and, for all practical purposes, in writing. A distinction is still maintained between ‘familiar’ and ‘formal’ you via distinct pronominal forms: você (pl. vocês) is the familiar form of address and o senhor or a senhora (pl. os senhores, as senhoras) are the formal forms. All of these pronouns use the third person verbal ending. Thus the verb form
fala, for example, corresponds to five pronominal subjects: êle, ela, você, o senhor and a senhora. The same is true of the plural falam.

Thus the verbal paradigm is now reduced to four inflectional endings, two of which are five-ways ambiguous. In the imperfect tenses and the present subjunctive the number is further reduced to three, since the first and third person singular are non-distinct. The result is that in many cases INFL is not ‘rich’ enough to unambiguously identify the null subject.

The above facts are interesting in the light of some informal observations on Brazilian speech habits. There appears to be resistance to using third-person verb forms without an accompanying overt pronominal subject. Further, the use of such overt subjects is losing its emphatic force. Recall that in NSLs embedded subjects which are coreferential to the higher subject (or to any NP in the higher clause) are preferably null, with the overt pronoun receiving a highly preferred disjoint reference reading. However, in (33) there is no preference for the disjoint reference reading vs. the coreferential reading:

(33)  Ele diz que êle está doente.
       ‘He says that he is sick’

Recall also the Avoid Pronoun principle as stated in (23), which was offered as an explanation for the preferred disjoint reference reading for the Spanish equivalent to (33). Note that in Spanish, which is unambiguously an NSL, overt subject pronouns are restricted to [+HUMAN]. The English pronoun it (not expletive it) has no phonetic corresponding form in Spanish:

(34)  a.  {[e]  Da al parque.
       {[*Ella]}
       ‘It (e.g. the window) looks out over the park’

       b.  Sabes como êle es: sonríen los padres y lloran las madres.
           {[*é]}
       ‘You know how it is: the fathers smile and the mothers cry’

Overt pronouns in the corresponding Portuguese sentences are, however, perfectly grammatical:

(35)  a.  Ela dá ao parque.

       b.  Você sabe como êle é: os pais sorriem e as mães choram.

On the other hand, expletive elements in Portuguese are still realized as null elements, as in (24e). Rizzi (1980) discusses a similar case with respect to certain northern Italian dialects. These dialects do allow free subject inversion and have empty expletive elements; however, a definite subject cannot be null:

(36)  a.  El vien.
       ‘He is coming’
b. *Vien.
c. Vien Giorgio.
d. (*El) Piove.

'It is raining'

Safir (1982) discusses similar facts in German and Dutch, which only allow null expletive subjects. The existence of sentences such as (24e), then, poses no problem for the hypothesis that Portuguese is becoming a non NSL.

If Portuguese is in fact a non NSL, then a new analysis of cases of long wh-movement of the subject is needed. The answer suggests itself in the analysis of wh-movement of the subject in French. In French, as in English, a trace in subject position preceded by a COMP containing the overt complementizer que and the coindexed trace of successive cyclic movement is not properly governed since the c-command requirement is not met:

(37) *Qui as-tu dit que viendra?
    'Who did you say that will come?'

There is a grammatical variant of (37):

(38) Qui as-tu dit qui viendra?

Pesetsky (1978) proposed that qui is a pronominal form of the complementizer que and represents a merger of que with the trace in COMP (referred to as the que → qui rule). Qui thus inherits the index of the trace and properly governs the subject trace, since with the merger there is only one element in COMP and the c-command requirement is met. This analysis is supported by the fact that qui also shows up in relative clauses in which the relativized constituent is the subject of the clause, regardless of whether the antecedent is human or not:

(39) {la fille} {que je crois qui est arrivé(e)} {la première}
    {le cheval} {le premier}
    'the girl/the horse which I believe arrived first'

At first glance, there does not appear to be a corresponding rule in Portuguese. The complementizer to embedded sentences is invariably que, regardless of whether wh-movement has affected the subject or the object:

(40) a. Quem você disse que [t] beijou Maria?
    'Who did you say that kissed Maria?'
b. Quem você disse que Maria beijou [t]?
    'Who did you say that Maria kissed?'

But it is instructive to look at relative clauses. In restrictive relative clauses, only que appears in COMP for both relativized subjects and objects:
There is no distinction corresponding to English who/which/that.

Traditional grammars of Portuguese insist that in non-restrictive relative clauses the relative pronoun for a relativized human subject can be quem or o/a qual, but not que:

In fact, in spoken Portuguese que is used almost exclusively. O/a qual may replace que, particularly in written usage, while quem is highly unacceptable if not ungrammatical. Thus in spoken Portuguese the relative pronoun and the complementizer are being merged in the form que, which on the surface is ambiguous.

Given this background, we can assume that que in (43) is equivalent to French qui:

(43) Quem você disse que chegou ontem?

The S-structure is as in (44):

(44) \[ [\text{S quem} [\text{S você disse [S que [S que [t_1 chegou ontem ]]]}}] \]

Quem in the embedded COMP is pronominal que, merging the complementizer with the coindexed trace in COMP and thus c-commanding and properly governing the subject trace.

Note that this analysis also accounts for the ECP effect in (30). Since the only potential proper governor for a trace in subject position is a coindexed trace merged with the complementizer in an immediately adjacent COMP, a subject can never be extracted from a wh-island. Recall the S-structure of (30), repeated below as (45):

(45) O diretor [\text{S que [S você me perguntou [S que filme [S, t_1 rodou t_1 em Portugal ]]]}] . . .

Even if we assume that the trace in the COMP of \( \bar{S} \) has merged with the complementizer, it cannot properly govern the coindexed trace in subject position of \( S \) because there is an intervening \( \bar{S} \).

This analysis of wh-movement of the subject in Portuguese may seem
less plausible than the corresponding French analysis, given the phonetic change effected by the *que—qui* rule, which gives some overt proof for the analysis. Nevertheless, the relative clause facts show that Portuguese *que* does in some cases have a pronominal character, which I am claiming is the result of a rule merging the complementizer with the trace in COMP.7 By this means, then, Portuguese can on the surface show *[that-trace] violations despite the fact that as a (partially) non NSL it does not have free subject inversion.8

**CONCLUSIONS**

This paper has two main conclusions:

1. Brazilian Portuguese is in the process of becoming a non NSL and thus lacks certain properties associated with the null subject parameter.

2. Wh-movement of embedded subjects in Brazilian Portuguese satisfies the ECP by means of a *que—qui* type rule. Thus a wh-moved subject must always pass through the immediately adjacent COMP and there are no true cases of long subject extraction.

Paula Kempchinsky
University of California, Los Angeles

**NOTES**

1. In NSLs an overt pronominal embedded subject is usually interpreted as disjoint in reference with the matrix subject. When the embedded subject is stressed, however, coreference is much more likely:

   (i) María; dijo que ella *iba a protestar la decisión*.

   (ii) María; dijo que ELLA; iba a protestar la decisión (y no Susana).

2. The parentheses notation is to be interpreted as follows: (*x) indicates that the sentence is ungrammatical if *x* is included, while *(x) indicates that the sentence is ungrammatical if *x* is excluded. Thus (4e) is ungrammatical if *él* appears in the string.

3. *Lexical categories* are those which can be defined with the categorial features [±N], [±V], i.e. [+]N (nouns), [+]V (adjectives), [−N] (verbs) and [−V] (prepositions) (but cf. Kayne (1981) on the status of prepositions as proper governors).

4. An alternative account of the that-trace facts is given by Aoun, Hornstein and Sportiche (1981), using the definition of government as in (i):

   (i) α governs β if for every maximal projection γ, γ dominates α iff γ also dominates β

   (Aoun and Sportiche, 1983)

   They propose that the potential proper governor for a trace in subject position is not the trace in COMP, but rather COMP itself. They further propose the COMP-indexing convention given in (ii), which applies at S-structure:

   (ii) \([\text{COMP} \overline{x}_i \ldots ] \neq [\text{COMP}_i \overline{x}_i \ldots ]\) iff COMP dominates only

   *i*-indexed elements

14
Thus in a structure such as (10) COMP-indexing cannot apply, since COMP does not dominate only i-indexed elements. Hence COMP is not coindexed with subject position and the trace has no proper governor.

5. There seems to be a third LF representation available for (16b) in which the negative element nessuno undergoes QR and is adjoined to its own S, becoming construed as an existential operator:

\[
(i) \quad \text{[non \pretendover{\chi} \{x_{\chi}, \{x_{\chi} \text{ ti arresto}\}}]} \\
\text{‘I do not require that for some } x_\chi \text{ you arrest you’}
\]

This interpretation is possible only when the negative element is in the scope of a negative operator in a higher clause and is itself in the subjunctive complement to certain matrix verbs, cf. Campos (1981), Kempchinsky (in preparation).

6. Portuguese as used throughout this paper will consistently mean Brazilian Portuguese.

7. A different approach to cases of long wh-movement of the subject is taken by Chao (1980). She argues for a resumptive pronoun strategy, where the empty category in the embedded subject position is a resumptive pronoun with no phonetic content coindexed with the wh-element in the higher COMP by a rule of resumptive pronoun interpretation. As support for this, she gives evidence of overt resumptive pronouns in relative clauses which allow Complex NP violations. While it is true that some dialects of Brazilian Portuguese have resumptive object pronouns, it is shown that even those dialects which do not do have apparent *[that-trace] violations, so that this is not a totally satisfactory approach. The question is further complicated by the null object pronoun phenomenon in Brazilian Portuguese, which has not been treated here.

8. The que—qui solution given here to account for the wh-movement facts was proposed in the original version of this paper. Since that time, Zubizarreta (1982) has appeared, in which a similar solution is proposed to account for wh-movement of the subject in continental Portuguese, which, like Brazilian Portuguese, lacks free inversion of the subject although it does allow null subjects with more freedom than Brazilian Portuguese. Note that regardless of the correctness of my proposal that Brazilian Portuguese is becoming a non NSL, the facts of continental Portuguese support Safir’s (1982) analysis, alluded to above, which separates free subject inversion from the NSL parameter. Both varieties of Portuguese demonstrate that free subject inversion cannot be the only mechanism by which languages show apparent ECP violations.

REFERENCES


Kempchinsky, P. In preparation. Scope of negation in Spanish subjunctive clauses. UCLA.