J. Linguistics **49** (2013), 33–53. doi:10.1017/S0022226712000217 © Cambridge University Press 2012 First published online 11 September 2012

Embedded null subjects in Capeverdean[I]

JOÃO COSTA
FERNANDA PRATAS

CLUNL – Faculdade de Ciências Sociais e Humanas, Universidade Nova de Lisboa

(Received 4 March 2010; revised 20 June 2012)

The status of Capeverdean as a *pro*-drop language is controversial. Baptista (2002) contends that this Portuguese-based creole has null referential subjects with some types of predicates, while Pratas (2002, 2007) proposes that it has only expletive null subjects. She argues that the rare cases of root null subjects can be analyzed as instances of null expletives. The aim of this paper is to show that in Capeverdean there is an asymmetry in the distribution of null referential subjects. These are ruled out in root contexts, but allowed in some embedded contexts; this is the case when the null subject is bound by a *wh*-operator or a quantifier. Following Holmberg's (2005) and Holmberg, Nayudu & Sheehan's (2009) analysis of null subjects, we offer an analysis of Capeverdean null subjects exploring the properties of T in the language (in particular, the lack of a rich inflectional system), the syntax of subjects, and the type of null category available. We claim that Capeverdean embedded null subjects are variables, licensed by an operator in the matrix clause. We show that these specific properties explain minimal differences between null subjects in Capeverdean and Brazilian Portuguese.

1. Introduction

Recent descriptions and debates on the status of null subject languages reveal that classic views of the null subject parameter, such as Rizzi's (1982), are too simplistic; languages do not split into just two options, allowing or disallowing null subjects. Comparing languages like Italian and English, Rizzi concludes that the licensing of null subjects is a parametric option. Summarizing Rizzi's observations, the following set of properties derives from the language-particular value of the null subject parameter:

(1) (a) Licensing of referential null subjects.
    (b) Licensing of expletive null subjects.

[1] We are very grateful to our Capeverdean consultants from Santiago Island, especially Ana Josefa Cardoso and José António Brito. We also want to thank three anonymous *Journal of Linguistics* referees and the editor, for their enriching comments and suggestions; to Nina Hyams, for her careful English editing and perceptive questions, and to our colleagues Alexandra Fiéis, Inês Duarte, Charlotte Galves and Maria Lobo, for relevant insights. Research for this paper was partly funded by FCT, through the project Events and Subevents in Capeverdean (PTDC/CLE-LIN/103334/2008).

https://doi.org/10.1017/S0022226712000217 Published online by Cambridge University Press

(c) Pronominal verbal inflection.
(d) Availability of subject-verb inversion.
(e) Lack of *that*-t effects.

Under the earlier formulations, the properties listed in (1) are expected to cluster together if a language has a positive setting for the null subject parameter. However, crosslinguistic research reveals that the correlations are not so robust. In fact, after surveying around 100 languages, Gilligan (1987) concludes that only four of the correlations hold, and all of them are unidirectional:

(2) (a) If referential null subjects are allowed, then null expletive subjects are allowed.
(b) If there is subject–verb inversion, there are null expletives.
(c) If there is subject–verb inversion, there are no *that*-t effects.
(d) If there are no *that*-t effects, there are null expletives.

Recent studies confirm Gilligan's (1987) view that the null subject parameter cannot be viewed simply as a split between languages with and languages without null subjects, in which all the properties in (1) cluster together. Figueiredo Silva (1996), Holmberg (2005), Neeleman & Szendrői (2007) and Holmberg et al. (2009), among others, show that there are several types of null subject languages.[2] Holmberg et al. (2009) propose a split between 'consistent' null subject languages and 'partial' null subject languages; in partial null subject languages null subjects are allowed in a much more restricted set of environments than in consistent null subject languages. Brazilian Portuguese and Finnish are examples of partial null subject languages. For instance, Brazilian Portuguese allows indefinite null subjects and embedded referential null subjects controlled by a c-commanding antecedent, but disallows referential third person null subjects in root contexts. Holmberg (2005, 2010) explains these differences in terms of the properties of the head of the inflectional domain – T.

Another productive line of inquiry addresses the issue of how different types of null subjects are licensed, particularly in partial null subject languages. Modesto (2007, 2008) and Miyagawa (2011) claim that null subjects can be licensed by a c-commanding operator when the referential properties of the inflectional domain are not available.

In this paper, we discuss null subjects in Capeverdean, a Portuguese-based creole. We argue that this language is a clear case of a partial null subject

[2] For a thorough review of the evolution of views on the null subject parameter, we refer the reader to Holmberg & Roberts (2010).

https://doi.org/10.1017/S0022226712000217 Published online by Cambridge University Press

language, and describe an asymmetry between root and embedded null subjects. To anticipate that discussion, we will show the following:

(3) (a) Capeverdean lacks root referential null subjects.
(b) Capeverdean has expletive null subjects.
(c) Capeverdean has embedded null subjects bound by a matrix operator antecedent.

We provide an analysis of embedded null subjects in Capeverdean showing that both the properties of the inflectional head (along the lines of Holmberg 2005) and the licensing mechanisms for the null subject (along the lines of Modesto 2007, 2008) need to be considered.

The paper is structured as follows. Section 2 spells out our assumptions regarding the syntax of null subjects, along the lines of Holmberg's work, namely, that the properties of the inflectional head T are relevant, and sets out the research questions:

(i) Does Capeverdean allow null subjects?
(ii) Is it a partial or a consistent null subject language (in the senses defined in Holmberg et al. 2009)?
(iii) If it is partial, what is the distribution of the null subjects allowed?
(iv) Furthermore, what is the licensing mechanism for these null subjects?

Section 3 discusses the properties of null subjects in root questions, showing that the language does not allow null referential subjects, but has null expletive subjects; these facts motivate the proposal that Capeverdean is a non-consistent null subject language. Section 4 focuses on embedded null subjects. Section 5 describes the difference between Capeverdean and Brazilian Portuguese in terms of the syntax of subjects and topics; we claim that the Capeverdean embedded null subject is a *pro*, licensed as a bound variable, as proposed in Modesto (2000) for Brazilian Portuguese, and not a deleted topic, licensed by a moved topic, as proposed, also for Brazilian Portuguese, in Modesto (2008). Finally, Section 6 presents the further predictions of our proposal and offers some concluding remarks.

2. Null subjects: Background assumptions

As mentioned in the introduction, the null subject literature of the past two decades recognizes that, whereas certain languages have null subjects across the board (e.g. Italian, Spanish or European Portuguese), others only allow null subjects in certain syntactic environments. One of the earliest findings regarding asymmetries in the availability of null subjects was the difference between expletive and referential null subjects. German, for instance, has expletive null subjects in root contexts, but does not allow referential null subjects (Gilligan 1987).

Holmberg (2005, 2010) and Holmberg et al. (2009) suggest that the split between 'consistent' and 'non-consistent' null subject languages depends on

https://doi.org/10.1017/S0022226712000217 Published online by Cambridge University Press

the properties of T. Following in the spirit of early proposals that null subjects are available when the inflectional head (T) is pronominal (e.g. Rizzi 1982, Barbosa 1995, Alexiadou & Anagnostopoulou 1998), Holmberg (2005, 2010) and Holmberg et al. (2009) suggest that a definiteness feature in the head of the inflectional domain (T) is relevant for characterizing a language as null subject. In languages like English or Swedish, this feature is absent (see Holmberg & Roberts 2010 for independent evidence). In consistent null subject languages, T bears a D(efiniteness)-feature, which is licensed under agreement with the referential subject or under incorporation of a related pronominal category. On the other hand, in partial or non-consistent null subject languages T is not specified for D(efiniteness) and, consequently, null subjects are restricted to indefinite environments. Since *pro* does not bear a D-feature in partial null subject languages, it may have a c-commanding antecedent. This is schematized in (4):

(4) Subj T[+Df] – T and Subj enter agree relation
Ø T[+Df] – T assigns interpretation to the subject under agree
Subj T[αDf] – Value of T determined by the subject
Ø T[αDf] – Only indefinite reading is possible

Another relevant factor for determining the syntax of null subjects is how null categories are licensed. As clearly stated in Neeleman & Szendrői (2007), not all null forms are licensed in the same way and this derives a wide typology of constructions involving these categories.

Brazilian Portuguese provides a clear example of the relevance of the licensing mechanism, as discussed in Modesto (2000).

Brazilian Portuguese is undergoing a change with respect to the setting of the null subject parameter (Duarte 1995, among others). In a diachronic study of popular theater plays written in the 19th and 20th centuries, Duarte (1993a) provides clear evidence that Brazilian Portuguese is losing referential null subjects. She observed an increasing rate of overt referential subjects in the corpora she examined, as shown in (5).

(5) *Rate of overt referential subjects in Brazilian Portuguese*
1845 – 20% 1882 – 23% 1937 – 46%
1975 – 67% 1992 – 74%

Duarte (1993b) also observed another piece of evidence indicating the ongoing change in Brazilian Portuguese, which is that the language does not display the typical effects of the Avoid Pronoun Principle (Chomsky 1981), in that it allows embedded overt subjects co-referential with a matrix subject:

(6) O povo brasileiro$_i$ acha que ele$_i$ tem uma grave doença.
the people Brazilian thinks that he has a serious illness
'Brazilian people think they have a serious illness.'

https://doi.org/10.1017/S0022226712000217 Published online by Cambridge University Press

The intended reading for a sentence like (6) would be ruled out in a consistent null subject language. For instance, in European Portuguese (6) is ungrammatical with the co-referential interpretation for the two subjects.

However, it is not the case that Brazilian Portuguese has lost null referential subjects altogether. Figueiredo Silva (1996), among other authors, notes that referential null subjects can still be found in some specific contexts. One such context is question–answer pairs with 1st or 2nd person verbal morphology, as in (7):

(7) A: O que você comprou?
the what you bought
'What did you buy?'
B: Ø comprei um livro.
bought a book
'I bought a book.'

Referential null subjects are also found in embedded complement clauses, in spite of the loss of the Avoid Pronoun Principle. As argued in Modesto (2000), in a sentence like (8) the embedded pronoun is optional:

(8) O Pedro disse que (ele) ganhou na loto. (Modesto 2000: 44)
the Pedro said that he won in.the lotto
'Pedro said he won the lotto.'

The example in (8) is particularly interesting because the null subject is possible but its interpretation is limited and different from what happens in typical *pro*-drop languages such as European Portuguese (Modesto 2000):

(9) (a) *Brazilian Portuguese*
O Pedro_i disse que $Ø_{i/*j}$ ganhou na loto.
the Pedro said that won in.the lotto
'Pedro said that he won the lotto.'
(b) *European Portuguese*
O Pedro_i disse que $Ø_{i/j}$ ganhou na loto.
the Pedro said that won in.the lotto
'Pedro said that he won the lotto.'

As shown in (9), the embedded null subject is obligatorily co-referential with the matrix subject in Brazilian Portuguese, which is not the case in European Portuguese. Modesto (2000) contends that in Brazilian Portuguese *pro* is available but cannot be identified by the inflectional head. According to his proposal, *pro* in Brazilian Portuguese is instead licensed and identified as a null bound variable (along the lines of Cinque 1988). Crucially, for a variable

https://doi.org/10.1017/S0022226712000217 Published online by Cambridge University Press

interpretation to obtain, it must be c-commanded by an antecedent. This proposal explains why the embedded subject cannot have a disjoint reading. Such a reading is only possible when *pro* is licensed as a pronoun with free reference and not as a bound variable.

Modesto (2008) further develops this analysis in order to account for some of the properties of embedded null subjects in Brazilian Portuguese. In particular, he aims to explain why they do not behave like PRO in control structures, and also why they do not display features of A-movement, contra Nunes (2008). Modesto (2008) proposes that null subjects in Brazilian Portuguese are instances of topic deletion. This analysis explains, among other things, the fact that the relation between a topic subject and a null subject cannot cross islands, and also the existence of intervention effects, when a *wh*-phrase is moved across a null subject to the left of a preverbal subject.

We will return to Modesto's two analyses below. What is important at this stage is the observation that two factors are relevant in the different null subject constructions, both the properties of the head of the inflectional domain (T) and the licensing mechanisms for null categories. Thus, several possibilities emerge in evaluating a given language:

(10) (a) Does the language allow null subjects or not?
(b) If yes, does the language allow null subjects consistently or partially?
(c) If partially, what is the distribution of null subjects?
(d) What is the licensing mechanism for the null subjects allowed?

These are the questions we intend to address for Capeverdean.

3. Null subjects in Capeverdean

The status of Capeverdean with respect to the null subject parameter has been the subject of some controversy in the literature. Baptista (2002) contends that Capeverdean is a null subject language, whereas Pratas (2002, 2007) claims that the language is a non-*pro*-drop language, specifically, a language with obligatory referential lexical subjects.

In part the debate was obscured by the fact that the authors used a uniform set of tests to determine the language's *pro*-drop status. However, we know that languages may be consistent or non-consistent regarding the null subject. As Gilligan (1987) has observed, languages do not behave uniformly with respect to the null subject criteria listed in Rizzi (1982). In fact, as shown in Pratas (2002), Capeverdean displays mixed properties. We describe these properties in the next subsections.

https://doi.org/10.1017/S0022226712000217 Published online by Cambridge University Press

3.1 *No root referential null subjects*

As shown in (11), root referential null subjects are ruled out in Capeverdean for all persons:

(11) (a) *(N) papia txeu.
1SG speak much[3]
(b) *(Bu) papia txeu.
2SG speak much
(c) *(E) papia txeu.
3SG speak much

Baptista (2002) proposes that subject clitic pronouns are agreement markers lexicalizing the head position of agreement projections, and not the subject of the root clause. Under this analysis, the sentences in (11) are actually null subject sentences because Spec,TP is empty.

To test these competing hypotheses, we draw on De Cat (2005). This paper provides several arguments to decide between two competing analyses of French subject clitics: one in which the clitic is an agreement marker (as in Rizzi 1986), and one in which the clitic is an argument (as in De Cat 2005). De Cat's argumentation is the following: if the clitic is the argument, it should only be doubled in contexts in which a pronoun and a DP are independently found to be doubled, as in contexts of contrastive topic or left dislocation, for instance:

(12) Jean, je l'ai vu.
Jean I him.have seen
'Jean, I saw him.'

If the clitic is an agreement marker, its role in the sentence is to mark agreement and not argumenthood. In that case, it is expected that clitics appear independently of the discourse function of the argument. As De Cat (2005) notes, this is not the case, since subject clitics only double a DP in topic contexts:

(13) A: Qui est arrivé?
who BE arrived
'Who arrived?'
B: (a) #Jean, il est arrivé.
Jean 3SG BE arrived
'Jean, he arrived.'

[3] Abbreviations used in example glosses: Ø = null subject; ADV = adverbial; BE = copula; COMP = complementizer; LOC = locative; NEG = negation; PASS = passive; PROG = progressive; PRON = pronoun; PST = past tense; SG = singular; TMA = tense/mood/aspect marker.

https://doi.org/10.1017/S0022226712000217 Published online by Cambridge University Press

(b) Jean est arrivé.
Jean BE arrived
'Jean arrived.'

(14) A: Qu'est-ce que Jean a fait?
'What has Jean done?'
B: Jean, il a chanté.
Jean 3SG has sung
'Jean sang.'

On the basis of this type of data, De Cat (2005) concludes that subject clitics in French are true arguments and not just argument markers.

Returning to Capeverdean, the language has a full array of pronominal forms. As shown in the table in (15), the language has strong, weak and clitic pronominal forms:

(15) *Capeverdean pronominal forms*

	Strong pronoun	Weak pronoun	Subject clitics	Object clitics
1sg	*ami*	*mi*	*N*	*-m*
2sg (informal)	*abo*	*bo*	*bu*	*-bu/-u*
2sg (formal, masc)	*anho*	*nho*	*nhu*	*nhu*
2sg (formal, fem)	*anha*	*nha*	*nha*	*nha*
3sg	*ael*	*el*	*e*	*-l*
1pl	*anos*	*nos*	*nu*	*-nu*
2pl	*anhos*	*nhos*	*nhos*	*nhos*
3pl	*aes*	*es*	*es*	*-s*

As we can see in the following examples, clitic doubling by a strong pronoun is possible, but limited to certain discourse settings. If the clitic is an argument, it should only be doubled in contexts in which a pronoun and a DP are independently found to be doubled, as in contexts of contrastive topic or left dislocation.

As De Cat finds from French, clitic doubling in Capeverdean is restricted to topic doubling environments:[4]

(16) A: Kenha ki txiga?
who that arrive
B: (a) #Djon, e txiga.
Djon 3SG arrive
(b) Djon./Djon ki txiga.
Djon Djon that arrive

[4] As in French, the SV order 'Djon txiga' is a legitimate order in wide focus contexts, and in answers to questions like 'What happened?'

https://doi.org/10.1017/S0022226712000217 Published online by Cambridge University Press

(17) A: Kusé ki Djon faze?
what that Djon do
'What has Djon done?'
B: Djon, e kanta./E kanta.
Djon 3SG sing 3SG sing
'Djon, he sang.'/'He sang.'

Given the logic of De Cat's statement that an argument clitic will only be doubled where a pronoun and a DP are independently found to be doubled, as in contrastive topic or left dislocation, we have evidence that the clitic in Capeverdean is the subject of the clause, and that root null subjects are ruled out.[5]

3.2 *Expletive null subjects*

Unlike languages with obligatory overt subjects, Capeverdean has expletive null subjects, as illustrated in (18):

(18) Sata txobe na Lisboa.
PROG rain in Lisbon
'It's raining in Lisbon.'

The contrast between referential and expletive subjects is highly relevant. A consistent null subject language will allow for both kinds of null subjects. On the other hand, expletive null subjects are found in some languages that do not have referential null subjects (Safir 1982, Vikner 1995, Holmberg 2005), which indicates that the licensing of null expletives is independent of the licensing of referential null subjects. As such, one can set the expletive data aside (see Pratas 2007 for an analysis of null expletives in Capeverdean), and focus on the distribution of null and overt referential pronouns.

[5] Baptista (2002) mentions two other cases in which root null subjects would be possible in Capeverdean: with stage-level predicates, and with copular predicates, such as in (i) and (ii), respectively:

(i) (Bu) sta livri.
you are free

(ii) (El) e nha pai
he is my father

Pratas (2002, 2007) shows that (i) is not productive at all in the language, based on corpus studies and judgments of native speakers who strongly reject null subjects in these contexts. As for (ii), it is ambiguous with a presentational sentence, involving an expletive subject, as in the French counterpart *C'est mon père* 'This is my father'. In that context, it is expected that the subject is null, since, as we will see, expletives are null in the language.

https://doi.org/10.1017/S0022226712000217 Published online by Cambridge University Press

3.3 *Indefinite null subjects*

In contrast to referential null subjects, indefinite subjects in Capeverdean can be null, as illustrated in the following example:

(19) Li pode fumadu.
LOC may smoke.PASS
'People may smoke here.'

3.4 *Inversion restricted to unaccusative contexts*

Another property that is relevant for the current discussion concerns the availability of VS word order. Some languages with referential *pro* may have a VS word order with all verb classes, under specific discourse conditions (see Rizzi 1982, and Gilligan 1987 for a refined version of the correlation). As shown in (20), subject–verb inversion is disallowed in Capeverdean both with transitive and with unergative verbs:

(20) (a) *Le Djon livru.
read Djon book
(b) *Le livru Djon
read book Djon
'Djon read the book'
(c) *Ta trabadja Djon.
TMA work Djon
'Djon works.'

This holds irrespective of the discourse settings and is independent of the definiteness of the subject. As shown in (21), there is no improvement in acceptability when the inversion occurs with indefinite subjects.

(21) (a) *Le tres omi livru.
read three man book
'Three men read the book.'
(b) *Ta trabadja tres omi.
TMA work three man
'Three men work.'

The only verbs allowing subject–verb inversion are the unaccusatives. As shown in the examples in (22), inversion is limited to indefinite subjects:

(22) (a) (dja) Txiga tres omi.[6]
TMA arrive three man
'There arrived three men.'

[6] Some informants report that the TMA marker *dja* improves the grammaticality of the sentence. However, there is no consensus on this point. In Pratas (2007), several hypotheses regarding the nature of this TMA marker in this context are raised. One interesting view is that it acts as a lexicalization of a Davidsonian argument, required in presentational contexts with unaccusative verbs. If that is right, it can be assumed that it sits in Spec,TP, and if so these sentences should also be considered instances of null expletive contexts.

https://doi.org/10.1017/S0022226712000217 Published online by Cambridge University Press

(b) *(dja) Txiga Djon/*(dja) More Djon.
TMA arrive Djon TMA die Djon
(c) Djon (dja) txiga./Djon (dja) more.
Djon TMA arrive Djon TMA die
'Djon arrived.'/'Djon died.'

Definiteness effects in inversion contexts show up in many languages that do not have consistent referential null subjects (see Pinto 1997, among others). Moreover, given that Capeverdean has null expletives (see (18) above), inversion with unaccusatives may be the counterpart of English presentational sentences like *There comes a man*, in which the preverbal position is occupied by a null expletive and the argument is necessarily indefinite (assuming Belletti's 1988 analysis for expletive-associate constructions in unaccusative contexts).

3.5 *No* that-*t effects*

As mentioned above, one of the properties typically associated with languages that have obligatorily overt referential subjects is *that*-t effects, whereby extraction is barred from a subject position in the presence of a complementizer. As shown in (23), and argued at length in Nicolis (2005), in Capeverdean extraction out of an embedded subject position after an overt complementizer is fully grammatical:

(23) Kenha ki bu ta pensa ma kunpra livru?
who COMP 2SG TMA think COMP buy book
'Who do you think bought the book?'

This fact would be compatible with the view advocated by Baptista (2002) that Capeverdean is a null subject language. However, as argued in Gilligan (1987), lack of *that*-t effects is not a sufficient condition for the identification of a language as *pro*-drop, since it implies only the availability of null expletives. These, as we have seen, are also possible in languages like German, which require overt referential subjects.

In short, we have identified the following properties of null subjects in root contexts in Capeverdean:

(24) (a) Null subjects cannot be referential in root contexts.
(b) Null subjects can be expletive in root contexts, in expletive-associate environments and with weather-predicates.
(c) Null subjects can be indefinite in root contexts.

On the basis of these properties, it is legitimate to postulate that Capeverdean is a NON-CONSISTENT null subject language. Let us see, then, if Capeverdean null subjects behave as predicted under Holmberg's (2005) analysis.

As mentioned above, Capeverdean lacks a specification for person inflection. This may be taken as a sign that there is no incorporation of any kind in

https://doi.org/10.1017/S0022226712000217 Published online by Cambridge University Press

T, making this functional category an inherent bearer of [+Definiteness] feature. There are now two options: either there is no [D] feature in T, and no kind of null subject is licensed, or the feature is underspecified. If the latter option is right, two types of null subjects are possible: expletive and indefinite. This is indeed confirmed by the facts, as we saw above. However, matters turn out to be more complicated. In the next section, we show that referential null subjects are also possible in the language, but only in embedded contexts.

4. Null subjects in embedded complement clauses

Thus far we have only looked at matrix contexts, concluding that Capeverdean is a non-consistent null subject language, which disallows referential null subjects. One might expect that referential null subjects are ruled out in both root and in embedded contexts, just as in English or French (Chomsky 1981, Montalbetti 1984). On the other hand, if we consider other non-consistent null subject languages, such as Brazilian Portuguese, Capeverdean null subjects in embedded contexts might be expected.

Prima facie, the distribution of pronominal subjects in embedded contexts does not challenge the conclusion that referential null subjects are ungrammatical. Referential null subjects are also ruled out in the subject position of complement clauses, even when their reference can be easily retrieved from context:

(25) (a) Djon fla m-**e** ta bai ku nos.
Djon say COMP-3SG TMA go with us
'Djon says that he goes with us.'
(b) *Djon fla ma Ø ta bai ku nos.
Djon say COMP TMA go with us

However, this is not the whole picture. Certain embedded contexts permit referential null subjects. Two such cases, an embedded complement clause within a *wh*-question and a relative clause, are illustrated in (26) and (27), respectively:

(26) (a) Kenha$_i$ ki ta atxa ma Ø$_{i/*j}$ kore faxi labanta mo.
who COMP TMA think COMP run fast raise hand
'Whoever thinks he ran fast raise (their) hand.'
(b) Kenha$_i$ ki ta atxa ma **el**$_{*i/j}$ kore faxi labanta mo.
who COMP TMA think COMP PRON run fast raise hand
'Whoever thinks he ran fast raise (their) hand.'

(27) (a) Ningen$_i$ ka atxa livru ki Ø$_{i/*j}$ perdeba.
no.one NEG find book that lose.PST
'No one has found the book that (they) lost.'

https://doi.org/10.1017/S0022226712000217 Published online by Cambridge University Press

(b) Ningen$_i$ ka atxa livru ki **el**$*_{i/j}$ perdeba.
no.one NEG find book that 3SG lose.PST
No one has found the book that he lost.'

As shown in the (a) examples, the subject of an embedded complement or relative clause can be null.[7] There is, however, an important difference between the (a) and (b) sentences in (26) and (27). In the (a) sentences, the embedded null subject is necessarily coindexed with the matrix subject, whereas in the (b) examples, the overt subject is obligatorily disjoint from the matrix subject. Thus, the subject is not really optional. Whether the subject is null or overt crucially affects its interpretation.

This behavior is consistent with Chomsky's (1981) Avoid Pronoun Principle further qualified in Montalbetti (1984), according to which overt pronouns in null subject languages are to be avoided whenever co-reference is intended. What we observe in the data just presented is that Capeverdean displays a mixed behavior. It has obligatory overt subjects in embedded contexts when the matrix subject is a DP (see 25a, b), but the subject of the embedded clause must be null when the matrix subject is a *wh*-operator, as in (26), or a quantified DP, as in (27), and co-indexation is intended.

Holmberg's (2005) analysis is compatible with the facts just presented, but it is insufficient. As in root contexts, the embedded T in Capeverdean is underspecified for the [D] feature, and is thus unable to license a null subject (unless it gets an indefinite reading). If, on the other hand, the null subject is a variable licensed by a matrix operator, then there is no clash. We expect, then, that embedded null subjects in Capeverdean are restricted to operator-variable contexts.

However, as shown for Brazilian Portuguese, non-consistent null subject languages are not uniform. A full understanding of this pattern requires a closer comparison between Capeverdean and Brazilian Portuguese.

5. Types of null subjects and embedded null subjects

As shown in Section 2 above, Brazilian Portuguese also has embedded null subjects which are obligatorily co-referential with the matrix subject. Modesto (2000) argues that the embedded *pro* in Brazilian Portuguese is a variable, which explains its obligatory co-reference. In a modified version of this analysis, Modesto (2008) suggests instead that the embedded null subject is a deleted topic, licensed by a moved topicalized item.

[7] This is similar to other partial null subject languages. Whenever possible, languages select a null anaphoric form for an antecedent like 'nobody' (Menuzzi 1999). This is not the case for languages in which null subjects are completely ruled out ('Nobody thinks that he is crazy').

https://doi.org/10.1017/S0022226712000217 Published online by Cambridge University Press

Let us test these two proposals, spelled out in A and B, against the Capeverdean data:

A. The embedded null subject is a *pro*, licensed as a bound variable, as proposed in Modesto (2000) for Brazilian Portuguese.
B. The embedded null subject is a deleted topic, licensed by a moved topic, as proposed in Modesto (2008) for Brazilian Portuguese.

The two hypotheses make different predictions. If the null subject is a moved (and deleted) topic, it creates a chain with a topic licensing it. This type of chain is subject to locality and subjacency relations. As such, and as demonstrated by Modesto (2008), null subjects in Brazilian Portuguese cannot occur in islands.[8] If the null subject is a *pro* licensed as a variable, it is predicted that it can occur in islands. The latter prediction is borne out in Capeverdean, as shown by a sentence such as (28), in which the null subject is inside a relative clause:

(28) Ningen$_i$ ka atxa livru ki Ø$_i$ perdeba.
no.one NEG find book that lose.PST
'No one has found the book that he lost.'

We will therefore adopt for Capeverdean a version Modesto's (2000) analysis of embedded null subjects in Brazilian Portuguese. We propose that *pro* is available in Capeverdean, but is restricted to contexts in which it establishes a relation with an operator. It follows that the subject must be c-commanded by its antecedent, as in the examples presented above (see (26a) and (27a)). We predict, however, that *pro* cannot be bound by a non-c-commanding 'no one', as in (29).

(29) [Amigu di [ningen]$_i$]$_j$ ka atxa livru ki Ø$_{*i/j}$ perdeba.
friend of no.one NEG find book that lose.PST
'No one's friend has found the book that he lost.'

Other properties of the embedded null subject in Capeverdean confirm its status as a bound variable.

(30) *The antecedent of the null subject is not necessarily animate*
Mas di ses avion$_i$ ka ta lebaba ningen kantu Ø$_i$ kai.
more than six plane NEG TMA take.PST no.one when fall
'More than six planes didn't take anyone when they fell.'

[8] As noted by an anonymous *JL* referee, this observation has been challenged for Brazilian Portuguese. See Rodrigues (2004) for discussion. We will not go into the discussion of the Brazilian Portuguese data here. For the purposes of this paper, it is sufficient to note that null subjects of the relevant type are legitimate in islands in Capeverdean.

https://doi.org/10.1017/S0022226712000217 Published online by Cambridge University Press

As discussed in Holmberg & Sheehan (2010), the fact that this relation is not restricted to animate antecedents shows that this is not a logophoric relation.

(31) *The binding relation is established with the subject in the first clause up*
Ningen$_i$ ka fla ma tudu algen$_j$ ta atxa ma Ø$_{*i/j}$
no.one NEG say COMP every person TMA think COMP
sta duente.
BE sick
'No one said that everybody thinks (they) are sick.'

(32) *Sloppy readings*[9]
Ningen$_i$ ka fla ma Ø$_i$ kore faxi i tudu algen$_j$
no.one NEG say COMP run fast and every person
fla ma Ø$_{*i/j}$ laba mo.
say COMP wash hand
'No one said (they) ran fast and everybody said (they) washed their hands.'

(33) *Split antecedents allowed*[10]
Tudu alunus$_i$ fla tres omi$_j$ pa Ø$_{i+j}$ laba losa.
every student tell three men for wash dishes
'Every student told three men for (them all) to wash the dishes.'

As a whole, these arguments confirm the hypothesis that the relation established between the antecedent and the embedded null subject is not one of control, replicating findings by Holmberg & Sheehan (2010).

Note that the analysis proposed by Modesto (2008) for Brazilian Portuguese cannot be fully extended to Capeverdean, since there is no perfect overlap between the two languages. In Capeverdean, the referential null subject is only available with *wh*-antecedents or when the antecedent is a quantified DP. In sentences in which the subject of the matrix clause is a non-quantified DP, *pro* is ruled out. In Brazilian Portuguese, the antecedent may be a non-quantified DP, as in (9a) above, repeated here as (34).

(34) *Brazilian Portuguese*
O Pedro$_1$ disse que Ø$_{1/*2}$ ganhou na loto.
the Pedro said that won in.the lotto
'Pedro said that he won the lotto.'

This difference calls for an explanation.

[9] We are using the type of examples Miyagawa (2009) uses to test strict and sloppy readings in pronominal forms in Spanish and Japanese. They do not involve VP-ellipsis, but rather the recoverability of antecedents from a first conjunct.

[10] The embedded sentence in this example is non-finite despite the fact that it may have a lexical subject (Pratas 2007).

https://doi.org/10.1017/S0022226712000217 Published online by Cambridge University Press

Both Brazilian Portuguese and Capeverdean allow embedded null subjects, and Brazilian Portuguese permits embedded null subjects if they are coreferent with a matrix referential subject. Both Brazilian Portuguese and Capeverdean permit embedded null subjects when the subject is coindexed with a quantified or interrogative matrix subject. We suggest that in Capeverdean, *pro* may be a bound variable. The crucial difference between the two languages therefore lies in the nature of the subject antecedents.

Modesto (2008) argues that all subjects in Brazilian Portuguese occupy an A-bar position from which they are able to bind a variable. This position is a topic position in the left periphery of the clause, which attracts DPs, in particular the subject DP. According to Modesto (2008), this is a consequence of the fact that Brazilian Portuguese is a topic-prominent language.[11] Modesto (2000), Costa & Galves (2002) and Costa, Duarte & Silva (2006) argue that the A-bar status of subjects does not follow from the fact that subjects occupy a peripheral position, but from the specific properties of the extended IP in Brazilian Portuguese. Independent evidence for the mixed A- and A-bar–status of the extended IP comes from the fact that topics can trigger verbal agreement in the language, as shown in (35):[12]

(35) Essas florestas chovem muito. (Galves 2001; Duarte 2008: 9, ex. (4))
those forests rain.3PL a.lot
'In those forests, it rains a lot.'

For Capeverdean, we assume a much simpler analysis. Following Pratas (2002, 2004, 2007), we assume that the matrix subject is in Spec,TP, which has A-properties. The fact that these subjects can serve as operators will then depend on its inherent status. If they are quantified, they have an inherent operator status and, as such, they can bind a variable. In Capeverdean, in contrast to Brazilian Portuguese, only *wh*-antecedents (which occupy an A-bar–position) or expressions that have an inherent operator status can serve as binders for the null variable in embedded subject position.

Independent evidence for the claim that preverbal subjects in Capeverdean remain in Spec,TP, in contrast to what Modesto (2008) argues for Brazilian Portuguese, comes from the fact that subjects are adjacent to the verb in Capeverdean but not in Brazilian Portuguese:

(36) (Onti) Djon (*onti) bai sinema.
ADV Djon ADV go movie.theater
'Yesterday Djon went to the movies.'

This gives independent support to the different nature of embedded null subjects in Capeverdean and in Brazilian Portuguese. In the latter, the null

[11] See Costa (2011) for discussion. Topic here is meant as the current topic of discourse, syntactically present in the clause.

[12] For a discussion on the productivity of these structures, see Costa (2011).

https://doi.org/10.1017/S0022226712000217 Published online by Cambridge University Press

subject is a deleted topic, which is confirmed by the fact that it establishes a relation with a dislocated topic, as suggested in Modesto (2008). In Capeverdean, it is a pronominal category, which remains in Spec IP, as shown by the adjacency requirement. As such, the two languages have in common the superficial fact that both have embedded null subjects. Their different properties, however, suggest that they are not of the same type.

We are now able to provide an analysis of the full array of subjects in Capeverdean:

(37) (a) *Root referential lexical subject*
T bears an underspecified [D] feature. Its value is fixed by the lexical subject in Spec,TP, under Agree.

(b) *Root indefinite null subject*
T bears an underspecified [D] feature. No subject is present to assign its reference. An indefinite reading obtains.

(c) *Root expletive null subject*
T bears an underspecified [D] feature. No subject is present to assign its reference. An indefinite reading obtains.

(d) *Root referential null subject*
T bears an underspecified [D] feature. There is no referent to fix the value of T. Indefinite reading clashes with the referential value. Ungrammaticality obtains.

(e) *Embedded referential null subject bound by operator*
T bears an underspecified [D] feature. *pro* is a variable. *pro* is bound by overt operator.

(f) *Embedded referential null subject bound by referential DP*
T bears an underspecified [D] feature. *pro* is a variable. Antecedent DP is not a legitimate binder. Ungrammaticality obtains.

Returning to the four questions raised in (10) at the end of Section 2 above, we are now able to provide complete answers:

(38) (a) Does the language allow null subjects or not?
Yes.

(b) Does the language allow null subjects consistently or non-consistently?
Non-consistently.

(c) If non-consistently, what is the distribution of null subjects?
Null subjects are allowed in root expletive and indefinite contexts, and in embedded contexts when bound by an operator.

(d) What is the licensing mechanism for the null subjects allowed in the language?
Since T bears an underspecified definiteness feature, null subjects are licensed when their value is compatible with an indefinite

https://doi.org/10.1017/S0022226712000217 Published online by Cambridge University Press

reading (in indefinite and bound variable contexts), cases where no clash obtains between the value of *pro* and the value of T.

6. Properties and predictions

The analysis of embedded null subjects as *pro* variables bound by A-bar–elements makes two further predictions. First, it is expected that quantifiers other than negative words and *wh*-expressions license embedded null subjects, which is confirmed by the following set of data:

(39) Tudu/Mas di seis/Metade di alunos ka atxa livru
all more than six half of the.students NEG find book
ki Ø perdeba.
that lose.PST
'Every/more than six/one half of the students haven't found the book that they had lost.'

Secondly, the analysis proposed here has consequences for current debates on the nature of the null subject parameter. As has been observed by other authors (Rizzi 1986, Sigurðsson 1993, Coelho et al. 2001, Holmberg 2005, among others), the null subject parameter is not a uniform phenomenon. Given this view, the availability of *pro* is not the crucial criterion to determine the status of a language as *pro*-drop, but rather the combination of its availability and its licensing and identification mechanisms. This predicts a much richer array of typological possibilities for null subject languages, as has been independently argued in Holmberg (2005). Note that, because we crucially involve T in the analysis of the null subject types, this proposal has no bearings on the typology of null objects available for the languages under discussion in this paper.

For Capeverdean, our analysis suggests that *pro* is available in the language as a bound variable. This is crucially different from what happens in consistent *pro*-drop languages, like Spanish or Italian, in which *pro* is licensed by a specific featural make-up of T.

If our analysis is broadly correct, Capeverdean enriches the typology of possible null subject patterns. It is a non-consistent null subject language, in which *pro* is licensed in a different way. It shares with Brazilian Portuguese the feature specification of T, which explains the lack of referential null subjects in root contexts, but it diverges from Brazilian Portuguese in the mechanism for licensing embedded *pro*, because of the different syntax of subjects and topics. This is in line with Holmberg's (2010) view on the null subject parameter, in which the parameter derives from properties of T and from the availability of licensing mechanisms for *pro*. In fact, Capeverdean provides direct evidence for Holmberg's (2010) claim that non-consistent null subject languages may constitute a case in which *pro* is available

https://doi.org/10.1017/S0022226712000217 Published online by Cambridge University Press

only if bound. The specific type of binding will depend on independent properties of each particular language. In the case of Capeverdean, we argue that the specific type of binding depends on the availability of *pro* functioning as a variable, and on the syntax of subjects.

Our analysis indirectly contributes to another debate on the syntax of embedded null subjects: the movement theory advocated in Nunes (2008) and disputed in Modesto (2008). The availability of embedded null subjects in islands supports the claim put forward in Modesto (2008) that embedded null subjects are not derived by A-movement.

To conclude, in this paper, we summarize it as follows:

(i) We have provided evidence against the view that Capeverdean is a *pro*-drop language.
(ii) We have shown that in spite of its non-*pro*-drop status, Capeverdean allows embedded null subjects, obligatorily identified by a matrix quantified antecedent or a *wh*- antecedent.
(iii) We have analyzed the embedded null subject available in Capeverdean as a variable only partly similar to the null subject available in Brazilian Portuguese.
(iv) We have proposed that the difference between the licensing mechanisms for variable embedded subjects in Brazilian Portuguese and Capeverdean stems from properties of topics and subjects in the two languages.
(v) We have argued that the behavior of subjects in Capeverdean contributes to a better understanding of the family of null subject constructions in that it provides evidence for further distinguishing the availability of null forms from their identification and licensing mechanisms.

REFERENCES

Alexiadou, Artemis & Elena Anagnostopoulou. 1998. Parametrizing Agr: Word-order, V-movement and EPP-checking. *Natural Language & Linguistic Theory* 16, 491–539.
Baptista, Marlyse. 2002. *The syntax of Cape Verdean creole, the Sotavento varieties*. Amsterdam & Philadelphia, PA: John Benjamins.
Barbosa, Pilar. 1995. *Null subjects*. Ph.D. dissertation, MIT.
Belletti, Adriana. 1988. The case of unaccusatives. *Linguistic Inquiry* 19.1, 1–34
Biberauer, Theresa, Anders Holmberg, Ian Roberts & Michelle Sheehan (eds.). 2010. *Parametric variation: Null subjects in Minimalist theory*. Cambridge: Cambridge University Press.
Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.
Cinque, Guglielmo. 1988. On *si* constructions and the theory of ARB. *Linguistic Inquiry* 19, 521–581.
Coelho, Izete, João Costa, Maria Cristina Figueiredo Silva, Fátima Oliveira & Sérgio Menuzzi. 2001. Ordem VS e sujeito nulo em PE e PB. Presented at the II Colóquio do Projecto PE/PB, Universidade do Ceará.
Costa, João. 2011. Topic prominence is not a factor of variation between Brazilian and European Portuguese. In Janine Berns, Haike Jacobs & Tobias Scheer (eds.), *Romance languages and linguistic theory 2009*, 71–88. Amsterdam & Philadelphia, PA: John Benjamins.

https://doi.org/10.1017/S0022226712000217 Published online by Cambridge University Press

Costa, João, Inês Duarte & Cláudia Silva. 2006. Construções de redobro em português brasileiro: sujeitos tópicos vs. soletração do traço de pessoa. Leitura. *Revista do Programa de Pós-Graduação em Letras* (UFAL) 33, 135–145.
Costa, João & Charlotte Galves. 2002. External subjects in two varieties of Portuguese: Evidence for a non-unified analysis. In Claire Beyssade, Reineke Bok-Bennema, Frank Drijkoningen & Paola Monachesi (eds.), *Romance languages and linguistic theory 2000*, 109–125. Amsterdam & Philadelphia, PA: John Benjamins.
De Cat, Cécile. 2005. French subject clitics are not agreement markers. *Lingua* 115, 1195–1219.
Duarte, Maria Eugênia Lamoglia. 1993a. Do pronome nulo ao pronome pleno: a trajetória do sujeito no português bra-sileiro. In Roberts & Kato (eds.), 107–128.
Duarte, Maria Eugênia Lamoglia. 1993b. O enfraquecimento da concordância no Português Brasileiro. In Roberts & Kato (eds.), 387–408.
Duarte, Maria Eugênia Lamoglia. 2008. O sujeito de referência indeterminada em sentenças infinitivas. *Revista do GEL* (*Araraquara*) 5, 9–30.
Figueiredo Silva, Cristina. 1996. *A posição sujeito no Português Brasileiro: frases finitas e infinitivas*. Campinas: Editora da Unicamp.
Galves, Charlotte. 2001. *Ensaios sobre as gramáticas do português*. Campinas: Editora da Unicamp.
Gilligan, Gary Martin. 1987. *A cross-linguistic approach to the* pro-*drop parameter*. Ph.D. dissertation, University of Southern California.
Holmberg, Anders. 2005. Is there a little Pro? Evidence from Finnish. *Linguistic Inquiry* 36, 533–564.
Holmberg Anders. 2010. Null subject parameters. In Biberauer et al. (eds.), 88–124.
Holmberg, Anders, Aarti Nayudu & Michelle Sheehan. 2009. Three partial null-subject languages: A comparison of Brazilian Portuguese, Finnish and Marathi. *Studia Linguistica* 63.1, 59–97.
Holmberg, Anders & Ian Roberts. 2010. Introduction. In Biberauer et al. (eds.), 1–57.
Holmberg, Anders & Michelle Sheehan. 2010. In Biberauer et al. (eds.), 125–152.
Menuzzi, Sérgio. 1999. *Binding theory and pronominal anaphora in Brazilian Portuguese*. The Hague: Thesus.
Miyagawa, Shigeru. 2011. Minimal parametric variation. Presented at the 21st Colloquium on Generative Grammar, Sevilla.
Modesto, Marcelo. 2000. *On the identification of null arguments*. Ph.D. dissertation, University of Southern California.
Modesto, Marcelo. 2007. Null subjects in Brazilian Portuguese and Finnish: They are not derived by movement. In Stanley Dubinsky & William D. Davies (eds.), *New horizons in the analysis of control and raising*, 231–248. Dordrecht: Springer.
Modesto, Marcelo. 2008. Topic prominence and null subjects. In Theresa Biberauer (ed.), *The limits of syntactic variation*, 375–410. Amsterdam & Philadelphia, PA: John Benjamins.
Montalbetti, Mario. 1984. *After binding: On the interpretation of pronouns*. Ph.D. dissertation, MIT.
Neeleman Ad & Kriszta Szendrői. 2007. Radical *pro* drop and the morphology of pronouns. *Linguistic Inquiry* 38, 671–714.
Nicolis, Marco. 2005. *On* pro *drop*. Ph.D. dissertation, University of Siena.
Nunes, Jairo. 2008. Inherent case as a licensing condition for A-movement: The case of hyper-raising constructions in Brazilian Portuguese. *Journal of Portuguese Linguistics* 7, 83–108.
Pinto, Manuela. 1997. *Licensing and interpretation of inverted subjects in Italian*. Ph.D. dissertation, University of Utrecht.
Pratas, Fernanda. 2002. O sistema pronominal do Caboverdiano. MA dissertation, Universidade Nova de Lisboa.
Pratas, Fernanda. 2004. TP in Capeverdean: An almighty functional projection? Presented at the Lisbon Workshop on Alternative Views on the Functional Domain, Universidade Nova de Lisboa.
Pratas, Fernanda. 2007. *Tense features and argument structure in Capeverdean predicates*. Ph.D. dissertation, Universidade Nova de Lisboa.
Rizzi, Luigi. 1982. *Issues in Italian syntax*. Dordrecht: Foris.

https://doi.org/10.1017/S0022226712000217 Published online by Cambridge University Press

Rizzi, Luigi. 1986. On the status of subject clitics in Romance. In Osvaldo Jaeggli & Carmen Silva-Corvalán (eds.), *Studies in Romance linguistics*, 391–419. Dordrecht: Foris.
Roberts, Ian & Mary Kato (eds.). 1995. *Português Brasileiro: uma viagem diacrônica*. Campinas: Editora da Unicamp.
Rodrigues, Cilene. 2004. *Effects of loss of morphology in partial* pro-*drop languages*. Ph.D. dissertation, University of Maryland.
Safir, Ken. 1982. *Syntactic chains and definiteness effect*. Ph.D. dissertation, MIT.
Sigurðsson, Halldór Ármann. 1993. Argument-drop in Old Icelandic. *Lingua* 89.2–3, 247–280.
Vikner, Sten. 1995. V°-to-I° movement and inflection for person in all tenses. *Working Papers in Scandinavian Syntax* 55, 1–27.

Authors' addresses: *Centro de Linguística da Universidade Nova de Lisboa,*
Faculdade de Ciências Sociais e Humanas,
Avenida de Berna, 26-C, 1069-061 Lisbon, Portugal
jcosta@fcsh.unl.pt
fpratas@fcsh.unl.pt

https://doi.org/10.1017/S0022226712000217 Published online by Cambridge University Press