

What Early Clitics Can Tell Us about Early Subjects

Larisa Avram* and Martine Coene
University of Bucharest and UIA-Antwerp

1. Introduction

It is a well-known fact that children drop subjects during the early stages of linguistic development, irrespective of the value of the subject drop parameter in the target language. It has also been argued that children acquiring null subject languages (NSL) drop the subject in an adult-like fashion at a very early stage, during which children acquiring a non-null subject language (NNSL) may still occasionally allow non-adult-like dropped subjects (Rizzi 1998). Few studies, however, focused on the status of early null subjects in NSLs or on the link between early subjects and other facets of the *pro*-drop parameter, such as the availability of overt postverbal subjects, of clitic doubling constructions, or of a morphologically rich agreement system.

In this article we investigate the setting of the *pro*-drop parameter in Romanian, a NSL. On the basis of data coming from two longitudinal corpora of child Romanian we argue that early null subjects in NSLs are null constants (as defined in Rizzi 1998). Moreover, we show that there is a link between the emergence of object clitics (OC) and the adult-like use of (null) subjects. In particular, adult-like null subjects are claimed to emerge during the same developmental stage as OCs. The use of adult-like subjects and OCs is related to the valuation of Agreement features.

2. *Pro* as a clitic: predictions for acquisition

The analysis builds on the theoretical assumption that there is a link between the availability of null subjects, overt postverbal subjects and clitic doubling constructions in NSLs: (i) clitic_i...doubling DP_i; (ii) *pro*_i...doubling DP_i.

Recent theoretical research has defined the null subject *pro* available in NSLs as an agreement pronominal affix, on a par with weak pronouns and/or clitics (Cardinaletti and Starke 1994, Kato 1999). On such a view, both [+pronominal] Agreement and weak pronouns are D elements with a theta-role, case and phi-features, and they can be 'doubled' by a strong pronoun or an overt DP.

Not only has *pro* been shown to exhibit the morphological deficiency and semantics of weak pronouns, but it has also been argued to enter doubling constructions, just like clitics (Torrego 1998). A parallel can be assumed between clitic doubling and *pro*. *pro* is a copy of the phi-features on T when Agreement is

* This paper was partly developed during the first author's stay at NIAS, The Netherlands, as a Mellon NIAS Fellow. Their support is gratefully acknowledged.

[+pronominal], while OCs are a copy of the phi-features of a null object (Avram and Coene 2002). In NSLs the EPP is satisfied by *pro* which occupies [Spec IP]. *pro* is doubled by a DP when the subject is overt in the same way in which the clitic is doubled by the overt direct object in clitic doubling constructions:

- (1) a. pro_i a căzut Ion_i
 pro has fallen Ion
 b. I_i - am văzut *pe* Ion_i
 him-have-1st sg. seen *pe* Acc. marker Ion

Pro-drop and clitic doubling are taken to reflect the same phenomenon. *Pro*-drop languages will allow postverbal subjects because they allow null clitics to satisfy the EPP. The prediction for language acquisition is straightforward: the presence of adult-like postverbal subjects in child language can provide evidence in favour of the acquisition of the null subject parameter. The analysis also predicts that OCs and *pro* emerge during the same stage. Since both OCs and *pro* involve Copy and Match/Agree with the phi-features of a (discourse) antecedent, their emergence is argued to be directly linked to the valuation of agreement features.

If this analysis is on the right track, the *pro*-drop parameter should be set when the value of Agreement [+/- pronominal] is set and the acquisition of the *pro*-drop parameter in NSLs and OCs might be related to the valuation of the same (set of) feature(s). Since *pro* is phonologically null, the analysis of early clitic constructions can prove instrumental in the analysis of early subjects. The absence of OCs could correlate to a yet unset *pro*-parameter. Previous studies of early Romanian showed that object clitics are absent during the early stages and, after emergence, they continue to be randomly omitted until rather late (Avram 2000, Avram and Coene 2003a). Assuming that there is a correlation between OC constructions and *pro*-drop, the immediate implication is that early (null) subjects and clitics in child Romanian begin to be used in an adult-like fashion during the same stage. The presence and the behaviour of clitics can be taken as an indicator of the setting of the *pro*-drop parameter.

3. Data and methodology

The data on which our analysis relies are provided by two longitudinal corpora of child Romanian, consisting of weekly 60 minute recordings of natural unstructured conversations of monolingual Romanian children (B.1;03-3;02 and A.1;09-3;06) with their mother and an investigator, some of which have already been transcribed in Childes format (MacWhinney and Snow 1990). For the present analysis, we only examined one (or two) recorded 60' file(s) per month for the period 1; 09 – 2; 02 for both A. and B., as summarized in Table 1:

Child	Nr. of files	Age
Bianca	7	1;09 – 2; 02
Antonio	7	1;09 – 2; 02

Table 1: Data used in the analysis

We first analysed all the utterances in the natural speech samples with an explicit or unambiguously identifiable implicit verb or a null auxiliary or copula. Imperatives and surrogate imperatives (such as subjunctives used with imperative force) were excluded.

These utterances were divided into three categories: (i) with an overt preverbal subject; (ii) with an overt postverbal subject; (iii) with a null subject. This was done in order to evaluate the relative productivity of null subjects overall. These data are presented in the Appendix. However, as will be seen, the focus was on the qualitative analysis of the data. The properties of early postverbal subjects were compared to postverbal subjects in an early NNSL (French) as well as to postverbal subjects in the target language. We then compared the developmental route of null and postverbal subjects to that of OCs (for OCs we used the data in Avram and Coene 2003a).

The development of the following elements was investigated: (i) the type of DP subjects (overt or null); (ii) the distribution of overt DP subjects (pre- and postverbal); (iii) the emergence of OCs; (iv) the inflectional paradigm of the verb, with a focus on agreement (person and number) markers.

If there is a link between OCs and the *pro*-drop parameter, we expect to detect a change in the distribution of null subjects and in the number and/or type of postverbal subjects at approximately the same stage when OCs emerge.

4. Early null subjects

4.1 Early null subjects in Romanian

Null subjects have been attested in both early NNSLs and early NSLs. Two facts meet general agreement. Firstly, for NNSLs, the rate of null subjects is higher in infinitive clauses, which led to the postulation of a developmental link between null subjects and root infinitives. Secondly, the *pro*-drop parameter seems to be correctly set at a very early stage in NSLs, since the number and the distribution of early null subjects in these languages parallels those of null subjects in adult grammar. In particular, early null subjects have been attested in *wh*-questions and in subordinate clauses, which has been taken as evidence in favour of their adult-like status. However, the null subjects in the first files of the examined corpus of child Romanian are subjects of finite declaratives during the early stages (before 2;01), when no embedded clauses or *wh*-questions are attested. In the absence of these distributional contexts one has no solid evidence in favour of their adult-like use. Measuring children's competence to use null subjects cannot be done on the basis of a comparison of rates of use of null subjects in early vs. adult grammars because the available longitudinal data are not constrained enough.

In NNSLs it has been noticed that in finite clauses an overt subject is usually present; null subjects are often associated with root infinitives. But there is no 'root infinitive' stage in NSLs, where one can only talk about a non-finite stage, i.e. a stage during which Tense and/or Agreement may be omitted. In child Romanian, the past participle (with omitted auxiliary) is used to indicate past tense but null subjects have been attested both with inflected verbs, as well as in clauses with omitted auxiliary:

- (2) a. a scăpat
has dropped (A. 1; 09)
b. (am) spart
(have) broken -Past.Part. (A. 1; 09)

Lexical verbs, copulas, and auxiliaries, when present, are always fully inflected, but the data show that the child uses these early inflected forms as ‘impostors’ (Avram and Coene 2003b). The number and person distinctions in the verbal paradigm are not mastered yet. As can be seen in Table 2, plural markings are absent:

	Singular	Inflection omitted	Plural required	
			and used	sg. used instead
A. 1;09 [99 utterances]	73	26	0	1
B. 1;10 [12 utterances]	3	9	0	0

Table 2: Singular/plural distinction in the verb paradigm in early files

The first person is practically absent in the early files. The children refer to themselves mainly using the second or the third person, sometimes during the same conversational turn. The data are summarized in Table 3 (ambiguous cases were not counted):

	1 st person	2 nd person		3 rd person	
		for 2 nd person	for oneself	for 3 rd person	for oneself
A. 1;09[99 utterances]	1	0	22	12	29
B. 1;10[12 utterances]	0	0	0	1	2

Table 3: Person distinctions in the verb paradigm in early files

A person feature clash is attested on three occasions when a 3rd person subject combines with a 2nd person lexical verb in the A. corpus:

- (3) *Antonio n- ai.
Antonio not have-2nd pers. sg.

This suggests that the verbal paradigm is significantly impoverished¹. Agreement is underspecified during this very early stage. The early null subjects

¹ The relatively delayed acquisition of person and number agreement has been attested for child German (Clahsen 1991), child French (where Ferdinand 1996 notices an overgeneralization of the 3rd person singular), child Spanish and Catalan (Grinstead 1994) or early Dutch (van Kampen 2003).

cannot be identified by syntactic agreement which has not been valued yet. In the absence of unambiguous agreement markers (number and, more importantly, person) on the verb the system will have to rely on a different identification mechanism. Traditional grammar labels non -finite constructions as non-personal. The lack of unambiguous person marking in child language suggests that all these early clauses are actually ‘non-personal’. In the absence of Agreement, the null subject cannot be syntactically identified, as is the case in the adult system. Our claim is that early null subjects in Romanian are null constants (Rizzi 1998), located in the [Spec IP] of a clause whose Agreement features are underspecified. The identification of null subjects during this early stage is discourse bound.

4.2 Early postverbal subjects and the *pro*-drop parameter

The fact that the value of the *pro*-drop parameter is not set yet during the early stages (before 2;01) is also supported by the status of early overt postverbal subjects. *Pro*-drop has often been associated with the availability of postverbal subjects. The prediction for acquisition would be that if the *pro*-drop is positively set, we expect to encounter postverbal subjects in the corpus. But the mere presence of overt postverbal subjects cannot be taken as evidence in favour of a positively set parameter. Firstly, early postverbal subjects have been attested in NNSLs (such as French) as well (e.g. Friedemann 2000). Secondly, the identified postverbal subjects, relatively rare in the examined files (see Appendix) are not adult-like from the beginning. Firstly, until 2;01, they are exclusively ‘no choice’ DPs, i.e. DPs whose restrictive condition narrows down the choice of a legitimate value for the variable which they introduce to one single entity:

- (4) strică Antonio
breaks Antonio (A. 1;09)

No indefinite or quantificational postverbal subjects have been identified in the early files, in spite of the fact that indefinite DPs are attested as early as 1;10 in the B. corpus and 1;09 in the A. corpus (Avram and Coene 2003b). No such restriction is at work in the adult grammar of the target language, where indefinite subjects are allowed (and in some cases preferred) in postverbal position. Actually, early preverbal subjects are also exclusively inherently no choice DPs (they are, like postverbal subjects, either demonstratives or proper names)². Whereas adult grammar distinguishes between pre- and postverbal subjects with respect to definiteness, disallowing non-specific indefinites from occurring as subjects in preverbal position, no such distinction can be detected in the early files where all the overt subjects are high on the definiteness scale: they are exclusively demonstratives or proper names during the early stage. Theoretically, this should come as no surprise. Null epithets (the ancestors of Rizzi’s null constant) are null R-expressions, the counterpart of overt proper names and definite DPs with linguistic antecedents. In semantic terms, Rizzi’s [anaphoric, -pronominal, -variable] null constant behaves

² The Romanian data confirm previous findings for child Dutch and child French (van Kampen 2002), where early subjects are also based on nouns with inherent reference.

like a no choice DP, which introduces a variable externally anchored to an individual at the top of the saliency stack of the discourse.

Further evidence that early postverbal subjects in child Romanian differ from postverbal subjects in the adult language is provided by prosody. In several cases, there is a break between the verb and the postverbal subject, as if the overt DP represented “added” information. Such prosody factors show that the postverbal subject is not (always) in its *in situ* position.

The data lead to the conclusion that early postverbal subjects in child Romanian are not adult-like. Moreover, they evince properties that have been detected with early postverbal subjects in NNSLs. This shows that the *pro*-drop parameter has not been set yet during this early stage in NSLs either.

For child French, a NNSL, it has been noticed that null and postverbal subjects disappear around the same age (Friedemann 2000). This may indicate that the *pro*-drop parameter has been negatively set. However, in a NSL, null and postverbal subjects are available in the target grammar. The data examined so far show that early null and postverbal subjects in a NSL are, just like in NNSLs, non-adult like. The obvious question which arises at this point is what exactly can serve as an indicator that the *pro*-drop parameter has been positively set in a NSL.

4.3 Null subjects and object clitics

If clitic doubling and null subject doubling are instances of the same phenomenon, that of clitic chains, as assumed in section 2, the immediate prediction is that OCs, which are overt, might provide a clue with respect to the setting of the *pro*-drop parameter in a NSL. Previous analyses of OCs in child Romanian show that they are totally absent during the early stages. The first clitics are ‘impostors’, which merely fill in a syntactic position (1;10 - 2;01 in the B. corpus, 1;09-2;00.11 in the A. corpus). After their emergence, they continue to be randomly omitted or, in some cases, their phi-features fail to match the phi-features of their ‘double’. By analogy, one may assume that *pro* subjects are absent during the stage when OCs are absent and that, after their emergence, some of the numerous null subjects may still be null constants. At 2;01, after the emergence of genuine OCs, B. still uses the 2nd person when talking about herself, and at 2;02 she still occasionally uses the singular form instead of the plural:

- (5) *desene(le) face nani.
drawings-the sleeps

Obviously, such situations do not seem relevant if one looks only for quantitative relevance. But we believe that they indicate that the child still vacillates with respect to Agreement and that the verbal paradigm, though ‘richer’ (there is an increase in the use of 1st and 2nd person marking on the verb), is still deficient (see Tables 4 and 5).

	Singular	Infl. omitted	Plural required	
			and used	Sg. used instead
A. 2; 02 [90 utterances]	46	42	2	0
B. 2; 02 [105 utt.]	91	6	4	4

Table 4: Number distinction in the verb paradigm at 2; 02

	1 st person	2 nd person		3 rd person	
		for 2 nd person	for oneself	for 3 rd pers.	for oneself
A. 2;01[79 utterances]	4	0	6	34	12
A. 2;02[90 utterances]	0	0	7	53	6
B.2;2 [105 utterances]	32	1	11	50	3

Table 5: Person marking in the verb paradigm at 2;01-2;02

The ‘errors’³ in the domain of verbal inflection match the agreement errors attested in the domain of object clitics (Avram and Coene 2003a).

Overt subjects are occasionally superfluously used in contexts which indicate that the child is still valuing the +/- pronominal value of Agreement and uses an overt DP in subject position in order to satisfy the EPP in a non-adult like fashion. The child knows that the [Spec IP] has to be filled, but he/she has not valued Agreement as + or – pronominal yet:

- (6) Bianca vrea leu(l). (about herself)
Bianca wants lion-the (B. 2;01.18)

It is also after the emergence of the first OCs that postverbal subjects are no longer restricted to no choice DPs. Mass terms begin to be used as postverbal subjects (though they are attested in the corpus earlier):

- (7) nu mai e ciocolată
no more is chocolate
“There is no more chocolate left.” (B. 2;01.18)

After the emergence of the first OCs the type of overt subjects gets diversified: they are no longer exclusively demonstratives and proper names.

³ The child probably chooses a default form.

It is shortly after the emergence of OCs that the first complementizers are attested. Embedded clauses introduced by complementizers emerge and they have null subjects:

- (8) aicea, că pus
 here, because put-Past Part. (B. 2; 01.18)

The first *wh*-questions are also attested, with null and postverbal subjects, as in the target grammar.

However, no significant increase in the rate of postverbal subjects is attested before 2;02 in any of the two examined corpora. Postverbal subjects are relatively frequent in the adult grammar but they are still relatively rare in the early files.

In spite of the fact that DPs other than no choice ones begin to be used as preverbal subjects, the corpus still contains a significant percentage of demonstratives and proper names, as summarized in Table 6:

	Preverbal subjects		Postverbal subjects	
	Dem.	Proper n.	Dem.	Proper n.
B. 2;02	5 (35.7%)	3 (21.4%)	1 (7.1%)	6 (42.8%)
A. 2;02	17 (68%)	6 (24%)	1 (12.5%)	4 (50%)

Table 6: Use of demonstratives and proper names at 2; 02

The data seem to indicate that at 2;02 the use of subjects is not always adult-like. By analogy with OCs, for which it has been argued that there is a significant time lapse between emergence and adult-like use, one could advance the hypothesis that after the first adult-like subjects the child still continues to use non-adult like ones. Further research is, obviously, necessary, in order to examine whether the acquisition of *pro* and OCs continues to proceed in a parallel fashion after this age.

5. The hypothesis

We assume a continuity approach to syntactic development, according to which the child has full access to all the components of UG from the onset of acquisition. What the child has to learn, on the basis of the input, are language-specific features.

The proposed explanation for the identified developmental pattern of OCs and subjects is that the valuation of agreement features is delayed. The child ‘selects’ the agreement feature (or all the features comprised by agreement) at a very early stage. But the valuation of this feature is delayed. The result will be the early underspecification of Agreement. We define underspecification of Agreement as deriving from an early restricted derivational space. The valuation of agreement features requires a multiple phase or a multiple Spell-Out derivation (Chomsky 1999, Uriagereka 2002), i.e. it requires a large derivational space, which is not available from the outset of acquisition (Avram and Coene 2003a).

The valuation of one feature in one domain may reflect on other domains. Once the child has acquired the complete set of agreement features in the nominal domain, s/he will start using determiners in an adult-like fashion and OCs emerge (Avram and Coene 2003a). Initially the child acquires agreement in D. Since *pro* is

identified by co-indexation with AGR (which is D), its identification is possible only after the valuation of Agreement features. The child can define the target language as a null or non null subject language only when Agreement has been valued on the basis of the linguistic input (for similar conclusions reached from a different perspective, see Grinstead 1994 or Hoekstra and Hyams 1996).

This analysis also provides evidence that early null subjects in NSLS and in NNSLs are uniformly non adult-like and that there might be no asymmetry between the acquisition of subject clitics and object clitics. The stage when they begin to be used in an adult-like way is probably language specific, deriving from language specific properties, among which the availability of robust unambiguous clues in the input or morphological properties of the target language.

Appendix

Age	Total utt.	PreV subject	PostV subjects	Null subjects	First object clitics
1;09.03	1	0	0	1	0
1;10.01	12	5	1	6	0
1;10.29	23	9	4	10	filler
1;11.26	35	2	5	28	0
2;00.11	17	3	6	8	0
2;01.18	72	10	9	53	3 (100%)
2;02.13	105	14	14	77	2 (50%)

Distribution of subjects in all utterances examined and the emergence of object clitics: Bianca

Age	Total utt.	Pre V subject	PostV subjects	Null subjects	First object clitics
1; 09	99	39	10	50	filler
1;10.13	117	48	2	67	filler
1;11	71	23	2	46	0
2;00.11	28	6	0	22	3 (13.6%)
2;01.08	107	28	9	70	3 (100%)
2;01.15	79	21	4	54	-
2;02	90	25	8	57	2 (50%)

Distribution of subjects in all utterances examined and the emergence of object clitics: Antonio

References

- Avram, L. (2000) 'The unbearable heaviness of antecedents'. Paper presented at *Going Romance* 2000, Utrecht.
- Avram, L. and M. Coene (2002) 'Romanian Genitive/Dative Clitics as Last Resort', *Balkanistica* 15, 1-34.
- Avram, L. and M. Coene (2003a) 'From feature selection to feature valuation: the case of early clitics and determiners'. Paper presented at Glow 2003, Workshop II: Generative Approaches to Language Acquisition, Lund, April 12.
- Avram, L. and M. Coene (2003b) 'Why is it difficult to reach Agreement?', in: A.T. Pérez-Leroux and Y. Roberge (eds.) *Romance Linguistics. Theory and Acquisition*, Amsterdam/Philadelphia: John Benjamins, 247-262.
- Cardinaletti, A. and M. Starke (1994) 'The typology of structural deficiency: on the three grammatical classes'. *University of Venice Working Papers in Linguistics* 4, 41-109.
- Clahsen, H. (1991) 'Constraints on parameter setting'. *Language Acquisition* 1 (4), 361-391.
- Chomsky, N. (1999) *Derivation by Phase*. MIT Occasional Papers in Linguistics 18.
- Ferdinand, A. (1996) *The Development of Functional Categories*, doctoral dissertation, University of Leiden.
- Friedemann, M.-A. (2000) 'Early French postverbal subjects', in: M.-A. Friedemann and L. Rizzi (eds.) *The Acquisition of Syntax: Studies in Comparative Developmental Linguistics*. London: Longman, 63-83.
- Grinstead, J. (1994) *Consequences of the Maturation of Number Morphology in Spanish and Catalan*, MA Thesis, UCLA.
- Hoekstra, T. and N. Hyams (1995) 'The Syntax and interpretation of dropped categories in child language', in: J. Camacho, L. Choueiri and M. Watanabe (eds.) *Proceedings of the 16th West Coast Conference on Formal Linguistics*. Stanford University: CSLI, 123-136.
- Kampen, J. van (2002) 'Learnability order in the French pronominal system'. Paper presented at *Going Romance*, Groningen, November.
- Kampen, J. Van (2003) 'The rise of the standard EPP in the acquisition of non-*pro*-drop languages'. Poster presented at Gala 2003, Utrecht.
- Kato, M.A. (1999) 'Strong and weak pronominals in the null subject parameter'. *Probus* 11, 1-37.
- MacWhinney, B. and C. Snow (1990) 'The Child Language Exchange System: an update'. *Journal of Child Language* 17, 457-472.
- Rizzi, L. (1998) 'Remarks on early null subjects', ms. University of Geneva.
- Torrego, E. (1998) 'Nominative subjects and pro-drop Infl'. *Syntax* 1(2), 206-219.
- Uriagereka, J. (2002) 'Multiple Spell-Out', in: J. Uriagereka (ed.) *Derivations. Exploring the Dynamics of Syntax*. London/New York: Routledge, 45-65.