

The Acquisition of Romance Languages

Selected papers from
The Romance Turn II

The Acquisition of Romance Languages

Selected papers from
The Romance Turn II

held at Utrecht University
7-9 September 2006

Sergio Baauw, Jacqueline van Kampen
& Manuela Pinto (eds.)

Published by
LOT
Janskerkhof 13
3512 BL Utrecht
The Netherlands

phone: +31 30 253 6006
fax: +31 30 253 6406
e-mail: lot@let.uu.nl
<http://www.lotschool.nl>

Cover illustration: 'Child of Wisdom', Acrylic on canvas
© Marja van Kampen www.irish-art.com/marja/

ISBN-13: 978-90-78328-46-9
NUR 616

Copyright © 2007 by the individual authors

Contents

Introduction

Towards a Theory of Language Acquisition <i>Kampen, Jacqueline van</i>	1-4
---	-----

Papers

Object Clitics as Last Resort. Implications for language acquisition <i>Avram, Larisa and Martine Coene</i>	7-26
--	------

The Role of Syllable Structure in the Acquisition of Brazilian Portuguese <i>Bonilha, Giovana, Carolina Lisbôa Mezzomo and Regina Ritter Lamprecht</i>	27-43
---	-------

The L2 Acquisition of Spanish Focus. A case of incomplete and divergent grammars <i>Domínguez, Laura</i>	45-57
--	-------

On the Acquisition of Clitic <i>Se</i> and Aspect in Spanish <i>Escobar, Linda and Vicenç Torrens</i>	59-71
--	-------

Germanic and Romance Discourse Devices in Acquisition <i>Kampen, Jacqueline van and Manuela Pinto</i>	73-96
--	-------

Modality, Non-Finite Forms and the Manifestation of the RI Stage in Null and Non-Null Subject Languages <i>Liceras, Juana M., Susana Perales and Aurora Bel</i>	97-111
---	--------

Acquiring a Null Object Language. When children sound target-like, but really aren't <i>Lopes, Ruth E. Vasconcellos</i>	113-126
---	---------

Beyond Deficiency. Reconsidering the null subject phenomenon <i>Palasis-Jourdan, Katérina and Michèle Oliviéri</i>	127-143
---	---------

The Acquisition of the Morphosyntax-Pragmatic Interface in French L1. Evidence from reference with articles and pronouns <i>Rozendaal, Margot</i>	145-164
---	---------

The Relationship Between Word Prosodic Structure and Sentence Prosody. (Non)evidence from Brazilian Portuguese <i>Santos, Raquel S. and Paula Fikkert</i>	165-179
Null and Overt Subjects at the Syntax-Discourse Interface. Evidence from monolingual and bilingual acquisition <i>Serratrice, Ludovica</i>	181-200
Referring Expressions in Early Italian. A study on the use of lexical objects, pronouns and null objects in Italian pre-school children <i>Tedeschi, Roberta</i>	201-216

E-mail addresses of contributors

Avram, Larisa	larisa.avram@gmail.com
Baauw, Sergio	Sergio.Baauw@let.uu.nl
Bel, Aurora	aurora.bel@upf.edu
Bonilha, Giovana	gfgb@terra.com.br
Coene, Martine	martine.coene@ua.ac.be
Dominguez, Laura	L.Dominguez@soton.ac.uk
Escobar, Linda	maescobar@flog.uned.es
Fikkert, Paula	p.fikkert@let.ru.nl
Kampen, Jacqueline van	Jacqueline.vanKampen@let.uu.nl
Lamprecht, Regina	lamprecht@terra.com.br
Liceras, Juana M.	juana.liceras@sympatico.ca
Lopes, Ruth	ruth@floripa.com.br
Olivieri, Michèle	olivieri@unice.fr
Palasis-Jourdan, Katérina	katerina.jourdan@wanadoo.fr
Perales Haya, Susana	susana.perales@unican.es
Pinto, Manuela	Manuela.Pinto@let.uu.nl
Rozendaal, Margot	M.I.Rozendaal@uva.nl
Santos, Raquel S.	raquelss@usp.br
Serratrice, Ludovica	Serratrice@manchester.ac.uk
Tedeschi, Roberta	Roberta.Tedeschi@let.uu.nl
Torrens, Vicenç	vtorrens@psi.uned.es

Dear reader,

We are pleased to present you *The Acquisition of Romance Languages. Selected papers from The Romance Turn II 2006*. The workshop took place on September 7-9 2006 in the old inner city of Utrecht.

The workshop welcomed 3 guest speakers. We would like to thank them for presenting their views:

- Larisa Avram (English language and Literature, University of Bucharest)
“Romanian direct object clitics as Last Resort: implications for language acquisition”.
- Anne Christophe (Laboratoire de Sciences Cognitives et Psycholinguistique, CNRS Paris)
“Bootstrapping early language acquisition”.
- Ludovica Serratrice (School of Psychological Sciences, University of Manchester)
“Pronominal subjects at the syntax-discourse interface: Evidence from monolingual and bilingual acquisition”.

We were able to host this *Romance Turn II* workshop due to financial support from the Royal Netherlands Academy of Arts and Sciences (KNAW) and the Utrecht Institute of Linguistics OTS. The Netherlands Graduate School of Linguistics (LOT) allowed us to publish the proceedings in their series. Organizing the event itself required the efforts and energy of many colleagues and students. As editors of these proceedings we speak for all participants when we specifically acknowledge the support of the three PhD students who helped to make a success of the conference: Ivana Brasileiro, Antje van Oosten and Roberta Tedeschi.

The present volume includes 2 of the invited papers and 10 of the 19 papers that were presented at the conference. We hope you will enjoy reading them.

Sergio Baauw
Jacqueline van Kampen
Manuela Pinto

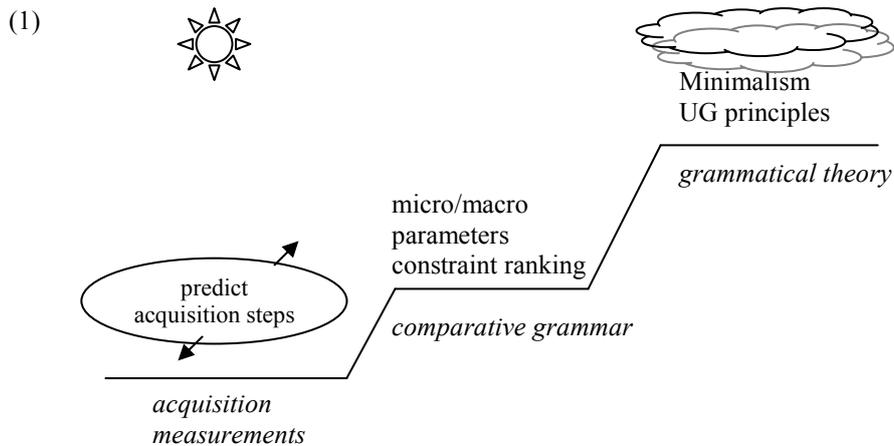
Towards a Theory of Language Acquisition

Jacqueline van Kampen

1. Acquisition measurements and goal

We have seen at the workshop a diversity of measurements and a no less diverse amount of different phenomena. Shared areas were acquisition phenomena in Romance languages and more general: the beginnings for a true theory of language acquisition. I will not attempt to evaluate the various contributions here. All of this must get its time to sink in. Yet, the general point of our endeavors, - the theory of language acquisition -, may be underlined to gain a further outlook.

The “hands-on” characterization of the workshop topic should mention the improved measurements techniques (elicitation tasks, judgment tasks, gap-filling tasks, story-retelling tasks, error quantifications, longitudinal acquisition graphs). We may now wonder where we expect to arrive when we hold on along those lines for the next five or ten years. Let me offer the following cartoon.



On the left, we have a plain, not to say a swamp, full of eco-diversity representing the diversity of acquisition measurements. On the right we see the high mountain of deductive grammatical theory with Baker’s typological macro-parameters and universal principles. It consists of deductions under the heavy sky of Minimalism.

In between, I imagine the sub-plateau of comparative grammar. Here we have the micro-parameters of the kind considered by Kayne (2000) to handle the North-Italian dialect variations. The flexibility of the micro-parameters may show how input controls micro-acquisitions. The sub-plateau also contains the ranking of Optimality constraints (Prince and Smolensky 1993), offering a potential space to handle the dynamics of acquisition. The underground of the sub-plateau could be the

more old-fashioned conception of markedness that goes back to the visions of Jakobson (1942) and Trubetzkoy (1939), two generations ago.

I think we will see how our work (experimental and file frequency measurements) will continue to find points of contact with comparative grammar. Besides that, I expect that a more constructive approach will develop, one that will predict beforehand the order and relative speed of the steps in language acquisition, given the type of grammar.

2. Is language acquisition a performance in slow motion?

The abstract theory of grammar in its general exposition will seldom fail to mention language acquisition as a crucial point, but only in a philosophical spirit. Such references to language acquisition are hardly ever followed by references to the actual study of language acquisition. The abstract theory of grammar is not comfortable with what we do. They may label that as surveys of performance data. I do not mean to complain about the theoretical grammarians. Grammatical theory on its own is bound to remain locked up in its own methods and devices. That is even a very good thing. The analysis of competence should precede the analysis of performance. Relevant performance data are hard to come by, but language acquisition might be seen as performance in slow motion. If there is a need for a general theory of grammatical acquisition as a performance activity, it is up to us to develop such a vehicle. I think we will do it. The questions to be considered are all about the hierarchies in the grammatical system and how broad decisions about the system support the subsequent more narrow decisions, cf. (2).

- (2) a. How does acquiring one set of features influence the other ones?
- b. Is there a predictable order of acquisition steps?
- c. Can the relative period of an acquisition phenomenon be predicted?
- d. Why exactly are interface conditions such a barrier for acquisition?
- e. What helps transfer in dialectical change and what blocks it in second language acquisition?

The ranking of parameters/features in an acquisition hierarchy is a major issue. The acquisition steps that set the parameters are taken in a linear order that impresses me as a causal chain. Each successive acquisition step shifts the focus of the learner to a new parameter. A direct example I know of is the acquisition of finiteness marking on the verb (I^0). The acquisition of finite verbs precedes the acquisition of determiners (D^0) in both Romance and Germanic languages (Van Kampen 2004). These steps offer the underpinning for all further acquisition steps. Agreement between subject DP and finite verb comes in later, and may appear as a final touch rather than as a structural underpinning.

The general issue and our chances at continued progress and relevance could be phrased in the following way. Our acquisition data are highly predictable. The acquisition order in first language acquisition and the relative period it takes to acquire a construction are parallel for all children acquiring the same language.

Order of acquisition and relative speed are not contingent properties. Hence, they have to be predictable from the type of grammar.

As far as second language learning is concerned, an experienced teacher can estimate someone's level from a short conversation. So, in second language acquisition we should be able to figure out how such estimates derive from the theory of grammar. We should get to understand why some properties of grammar are quite well learnable in second language acquisition, whereas other ones are hardly learnable at all. The answers to such questions will be given by our type of investigations, and they are bound to be relevant to assumptions about the design of grammar. I feel we should have high expectations about acquisition research.

3. The task at hand

I was inspired about this by an article in Scientific American (Ross 2006) that dealt with the memory organization of chess masters. These super-performers have developed an immense high-speed memory for configurations of chessmen in real games. Such a memory must have cost them some ten years of intensive daily training. Yet, there is no transfer from this memory to other games or activities. The memory is highly task-specific and flexible in quite specific ways only.

This reminds of language. We are all grandmasters in our native language, and the amount of training needed for the various parts of grammar and lexicon has been extensive and can roughly be predicted. Yet, there is only a limited amount of transfer when learning other languages. Second language acquisition is difficult and asks for a considerable amount of additional training. How does that tally with the undeniable evidence for a UG frame? A first approach may be that we see how the acquisition procedure in child language is structured and builds up an effective memory.

- (3) Predict the order of acquisition steps and the relative time, the training, needed for each acquisition step given a grammar

There is a highly practical side to this program as well. A theoretical insight in the training needed for grammatical and lexical habits may help us organize the massive amount of second language learning needed for the present century.

So, it is defensible that we work on these issues. The apparent diversity of our work is directed at a common point, the learnability of natural language as we see and hear it happen in actual practice.

References

- Jakobson, R. (1942) *Kindersprache, Aphasie und Allgemeine Lautgesetze*. Upsala: Uppsala Universitets Årsskrift.
- Kampen, J. van (2004) 'Learnability order in the French pronominal system', in: R. Bok-Bennema, B. Hollebrandse, B. Kampers-Manhe and P. Sleeman (eds.) *Selected Papers from Going Romance 2002*. Amsterdam/Philadelphia: John Benjamins, 163-183.

- Kayne, R.S. (2000) *Parameters and Universals Oxford Studies in Comparative Syntax*. New York: Oxford University Press.
- Prince, A. and P. Smolensky (1993) 'Optimality Theory: Constraint interaction in generative grammar', Technical report. Rutgers University Center for Cognitive Science.
- Ross, Ph.E. (2006) 'The expert mind', *Scientific American* August 2006.
- Trubetzkoy, N.S. (1939) *Grundzüge der Phonologie*, Prague: Travaux du Cercle Linguistique de Prague 7.

Section Papers

Object Clitics as Last Resort Implications for language acquisition*

Larisa Avram and Martine Coene
University of Bucharest and Leiden University, University of Antwerp

1. Aim

Various studies dealing with the emergence of Accusative direct object clitics (ADOCs) have revealed that they are not used in an adult-like fashion from the first observable stages. The data show either a relative late emergence and gradual target-like use or early emergence followed by a period of non-target-like use. For example, ADOCs have been argued to emerge later than determiners or other object pronominal clitics (Dominguez 2003) and also later than subject clitics (Hamann, Rizzi and Frauenfelder 1996; Hamann 2002). Clitics seem to be randomly omitted during the early stages (Jakubowicz et al. 1998; Haegeman 1996; Schaeffer 2000, Avram and Coene 2003; Wexler et al. 2004; Costa and Labo 2005; Perez-Leroux et al. 2006; Pîrvulescu 2006). In addition, *phi*-features and/or case errors have also been reported for French (Jakubowicz et al. 1998), Romanian (Avram 2001), and Spanish (Dominguez 2003). Such results clearly indicate that ADOCs are problematic across languages.

The present paper investigates the acquisition of ADOCs in Romanian. The novelty of the analysis derives both from the type of investigated data and from the research method. This is the first study which investigates the acquisition of ADOCs in Romanian on the basis of longitudinal data starting from the identification of the obligatory ADOC contexts in the adult system. A similar method was used for the study of early Accusative clitics in French (Pîrvulescu 2006). The present study, though, starts from the identification of the *obligatory* ADOC contexts, rather than from the *permissible* ones. This restriction is required by the properties of clitic doubling constructions in Romanian.

2. Assumptions with respect to language acquisition

Our analysis builds on the assumption that language is an optimal solution to legibility conditions (Chomsky 2000), i.e. only legible elements should be present in the expressions spelled out by the computational system. We adopt the view that language acquisition is a form of theory construction which implies selection and valuation of (a small inventory of) features. Parameter setting may reduce to a very small number of valued features, since valuation of one feature may lead to a cascade-like setting of parameter values (Avram and Coene 2005). A key feature is

* This work received financial support from the University of Antwerp and the Fund for Scientific Research-Flanders. The authors wish to thank Jacqueline van Kampen for her comments and suggestions on an earlier version of this paper.

Person (Schaeffer and Ben Shalom 2004, Coene and Avram 2004). In accordance with Longobardi's (2004) *Revised Denotation Hypothesis* we assume that "individuals are denoted by association with a (specified or default) person feature", the feature responsible for the projection of D. A positive setting for Person entails denotational uniqueness, i.e. a unique referent that has been previously introduced into the discourse or that is part of the (common) knowledge of the speaker/hearer. A similar proposal was advanced in Chomsky (1999) where the elimination of categorial features leads to interpreting Person as playing the role formerly assigned to [D] or [N] features. In earlier studies, D was associated with referentiality, therefore we infer that the referential property is taken over by the Person feature. The motivation for such a proposal goes back to Postal (1966), where it is argued that determiners and pronouns are expressions of the same Person feature, i.e. personal pronouns and determiners are in fact variants of the same category. Since clitics are D-elements, we expect the valuation of the Person feature to be an important ingredient in the acquisition of clitics.

Person is also relevant for the acquisition of Inflection. A positive setting, for example, will indicate the availability of Inflection-licensed null subjects (Coene and Avram 2004). The value of the Person feature in a particular language affects both the D-domain and the I-domain. The general idea is that a single feature can play an important part in a variety of constructions. Identifying correlations among the emergence and acquisition of all these functional domains may offer the starting point for a unified account of what might seem, at first sight, a bundle of unrelated phenomena.

On this approach, it is possible to account for the acquisition of clitics in terms of the valuation of the Person feature in the target language. One can also explain the obvious asymmetry between the complexity of clitic constructions and the speed with which they are acquired. There is one single feature which requires valuation and evidence in favour of a positive (or negative) setting can come from a variety of constructions. This means that there are numerous cues in the input with respect to the value of the Person feature.

Summing up, our analysis starts from the assumption that the Person feature is a key ingredient in the acquisition of pronominal clitics. The present study investigates its relevance for the emergence and development of ADOCs in Romanian.

3. ADOCs in Romanian

3.1 Obligatory ADOC contexts: the rationale

In this section we focus on the identification of obligatory clitic contexts in Romanian in an attempt at finding out what they tell the child about the properties of ADOCs in the target language, i.e. we analyze ADOCs in Romanian building on what is available in the input which the child receives.

3.2 ADOCs and overt antecedents

In Romanian, the presence of ADOCs is required whenever there is a putatively dislocated direct object DP, surfacing at the left periphery of the sentence, while the post-verbal complement position is empty. This is the case of (i) left dislocation structures with dislocated D-linked¹ direct objects (1); (ii) relative clauses (introduced by *care* ‘which, who’) (2); (iii) D-linked *wh*-questions (with *care* ‘which, who’) (3); (iv) right-dislocation structures (4):

- (1) Cartea, am dat *(-o).
book.the have given - Acc clitic 3rd sg fem
“I have given the book away.”
- (2) Mărul pe care *(l-) am mâncat.
apple.the *pe* which Acc clitic 3rd sg masc have eaten
“The apple which I have eaten.”
- (3) Pe care *(l-) ai ales?
pe which Acc clitic 3rd sg masc have chosen
“Which one have you chosen?”
- (4) *(L-) am mâncat # mărul.
Acc clitic 3rd sg masc have eaten # apple.the
“I have eaten the apple.”

In all these cases the clitic and its (overt) antecedent occur in the same clause. Importantly, following Cinque (1990), we adopt the view that in these structures the antecedent is base-generated in a left peripheral position². They contain a D-linked topic base-generated in a position at the left-periphery of the clause. One could analyze these constructions as predication structures, as suggested in Chomsky (1977), where topicalized arguments are base-generated in a left-dislocated A-bar position and licensed by rules of predication³.

To sum up, the data presented in this subsection show that the ADOC is obligatory when the left peripheral position (presumably SpecTopP) is occupied by a D-linked overt constituent co-indexed with the direct object and the post-verbal complement position is phonetically null. The antecedent is base-generated in this peripheral position.

¹ D-linking is used in the sense of Pesetsky (1987).

² Arguments in favour of a non-movement analysis include scope reading, sensitivity to strong islands, parasitic gaps, intonation. They all point to a systematic contrast between non-D-linked *wh*-phrases and D-linked *wh*-phrases. It is for the latter that one can argue in favour of a non-movement analysis.

³ We thank Jacqueline van Kampen for pointing this out to us.

3.3 ADOCs and null antecedents

ADOCs are also obligatory when the post-verbal complement position is phonetically empty but the null direct object has a (salient) antecedent in the preceding discourse (5):

- (5) A: Ce-ai făcut cu mărul?
“What have you done to the apple?”
B: *(L-) am mâncat.
Acc clitic 3rd sg masc have eaten
“I have eaten it.”

The ADOC is referentially anchored to an antecedent; it has *no choice* reference, i.e. the choice of a legitimate value for the variable which they introduce is narrowed down to one single entity (Farkas 2000). An ADOC has unique reference.

In 3.2 we saw that ADOCs must surface whenever the direct object position is empty and an overt antecedent surfaces in the left periphery of the clause. We have no reason to believe that ADOCs in such contexts are different from those in clitic constructions with no overt “associate” in the same clause. Following Delfitto⁴ (2002), we assume that all clitic constructions contain both a clitic and the antecedent of the clitic, i.e. the reference of the clitic is mediated by a sentence-internal topic, placed in the left periphery, plausibly in SpecTopP⁵. In the clitic constructions discussed in 3.2 the topic is overt. In single clitic constructions, it is a sentence-internal phonetically null topic placed in the same root position, SpecTopP, which cannot be c-commanded; consequently, its identification will occur through discourse. The advantage of this analysis is that it captures the uniformity of ADOCs across all the identified obligatory contexts. They all show that the ADOC is obligatory when its clausal antecedent (null or overt) is a referentially stable topic and the direct object position is null. The antecedent of the null direct object is salient, and theoretically retrievable at the interface. However, the derivation does not converge with null objects. This indicates that Romanian bans null objects which are referentially stable topics.

⁴ It is important to mention that Delfitto’s own analysis builds on results on the acquisition of clitics reported in Baauw (2000).

⁵ According to Delfitto (2002), there is an inherent link between pronominal clitics and clitic resumption of left-dislocated topics. The same line of investigation is taken for Greek clitics in Androulakis (2001), where it is argued that clitics in dependencies should be analyzed in the same way as simple clitics.

- (11)(O) avem aici pe Ruxi.
 Acc clitic 3rd sg fem have here *pe* Ruxi
 “We have Ruxi here”.

These contexts are *permissible* clitic contexts, the clitic can surface but it is not obligatory. The presence of *pe* does not require the use of the clitic nor does the presence of the clitic force the use of the preposition *pe*. If this is indeed the case, it follows that the role of *pe* is not merely that of an Accusative case assigner as standardly assumed in the literature. Its presence cannot be directly related to the obligatory use of the clitic. Farkas (1978), Farkas and Heusinger (2003) analyze *pe* as a realization of differential object marking in Romanian, a language where differential object marking is sensitive to referential stability and topicality.

What the analysis of both single clitic constructions and of the so-called clitic-doubling constructions reveals is that the clitic is always associated with a contextually salient element. All the identified obligatory clitic contexts in Romanian are structures in which the clitic has the same properties. Clitic doubling, clitic dislocation structures and single clitic constructions are less different than previously assumed (see also Kayne 1994; Androulakis 2001; Delfitto 2002 for similar proposals for other languages). The contribution of the ADOC is systematic across all these constructions: it rescues the referential properties of a D-linked null direct object and, consequently, it also has a topicalization effect (Farkas 1978).

Two possibilities present themselves with respect to the position of the lexical double in clitic doubling constructions. Both have been proposed in the literature. In principle, the *pe* phrase could occur in argument position (as argued, for example, in Kayne 1994) or in adjunct position (as argued for clitic doubling structures in Greek by Androulakis 2001). If our analysis so far is on the right track, then the postverbal complement position is occupied by the null object, the internal argument of the verb. Evidence in favour of such an analysis comes from clitic doubling constructions like the one in (12), where the clitic and the lexical “double” are both in post-verbal position:

- (12)A desenat-o pe Maria.
 has drawn Acc clitic 3rd sg fem *pe* Maria
 “He has drawn Maria.”

We do not commit ourselves with respect to the exact position which the *pe* marked phrase occupies. It might be the Topic projection (the clause-internal topic position) in the low IP area, i.e. the area immediately above vP, as proposed in Belletti (2004). Since verbs move to Inflection in Romanian, the *pe* phrase will surface in postverbal position. Alternatively, one could propose that the prepositional phrase occupies whichever position is assumed for clitic-resumed right-dislocated elements. But the *pe* phrase cannot occur in argument position.

3.5 Clitics as Last Resort

The data analyzed in the previous subsections show that referential null objects are illicit in Romanian. The question is what exactly bans such null objects, requiring the presence of the clitic. Like any null element, they require licensing and identification. They are licensed by syntax⁶, they are the internal argument of a transitive verb. Agreement object is not pronominal in Romanian, so the null object cannot be locally identified through Agreement. In this it differs from *pro* in subject position, which can be identified via Spec-head agreement with Agreement subject. The null object occupies the internal argument position, also a case marking position. Case assignment at Merge in Romanian has been proposed in various studies (Avram and Coene 2000, Alboiu 2002). A tension is created between the case-assigning properties of the structural position and the null status of the internal argument. As a Last Resort strategy, the features of the null object are spelled-out by the clitic, becoming visible for the computation. In a nutshell, the clitic is the spell-out of the features of a null object, created in the derivation. What actually becomes visible is the Person feature, spelled-out as a clitic which, like any pronominal element, inherits the features of the R-expression whose substitute it is. It obviously follows that, being a copy of the null object, the clitic also inherits its referential stability as well as its topic feature.

The question which arises at this point is how the system identifies the *phi*-features (number and gender) of the null object, its complete Person index. We saw that the antecedent of the clitic is in an A' position, at the left periphery of the clause, where it can be linked into discourse. The identification of the null object requires Person feature matching with its antecedent. But matching is possible only if there is no intervening Person feature. In Romanian the Person feature of Inflection is strong (Agreement is pronominal), allowing *pro* subjects; there will always be a potential barrier between the features of the antecedent in the left periphery and those of the null object⁷. The Person feature of Inflection disrupts the feature matching relation between the antecedent and the null object. That is why the clitic will have to move to a position higher than the intervening blocking feature; it moves to Top for identification reasons, i.e. so that the referential index of the null object be rescued via matching with its antecedent.⁸ In finite constructions, ADOCs must surface at

⁶ Pîrvulescu and Roberge (2005) argue that the direct object position is always projected. It merges to all verbs in the syntax as a property of Universal Grammar. This position can be occupied by an overt constituent or it can be phonetically null. For Romanian, the position must be occupied by an overt constituent when the direct object has *no choice* reference.

⁷ Compare the ban on null objects in Romanian to the availability of a restricted set of null objects in English, otherwise a non null object language. In the so-called recipe context, English allows null objects (Massam and Roberge 1989) but only when the subject is omitted. English does not have pronominal Inflection; but the Person features of the overt subject would block feature matching between the null object and its antecedent.

⁸ A reviewer correctly points out that this analysis leaves the structures with the feminine clitic *o* placed in post-verbal position (illustrated in 1) unaccounted for. In Avram (2000) the

the left periphery of the clause. In non-finite constructions Inflection does not have a strong Person feature, i.e. it is not pronominal, so it will not intervene between the antecedent in the left periphery and the clitic, making movement to a higher projection unnecessary.

It is also important to stress that the clitic is created across two phases: vP and CP. Case is checked *in situ*, making the feature Person visible, but gender and number features can be “copied” only after the identification of the referential index of the null object, i.e. after movement to the Topic projection.

To sum up, we propose an approach to ADOCs as a syntactically created copy of features. This bundle of features receives case inside the null DP whose copy it is and a referential index (reflected in number and gender agreement) in a higher projection at the left periphery. Both feature spell out and movement are related to the value of the Person feature.

4. Predictions for acquisition

Following our analysis of ADOCs in Romanian we can say that what children are required to know in order to acquire clitics is that the target language bans D-linked null objects. This ban is enforced by the value of the Person feature in the language. It means that valuation of the Person feature across domains is a prerequisite for the acquisition of clitics. The prediction is that before the full valuation of the Person feature, the early grammar can contain target-deviant clitics. An unvalued Person feature allows the identification of the null object via direct recourse to discourse (very much along the line of Schaeffer 2000), resulting in clitic drop. It also allows a target-deviant grammaticalized Person index, leading to gender and number substitutions and possibly non-moved clitics. The identification of the Person index of the null object implies movement to the left periphery. This indicates that target-like clitics presuppose the existence of an active C-domain, as

exceptional behaviour of the singular feminine clitic (which remains in post-verbal position in the periphrastic past tense with the auxiliary *avea* ‘have’ but surfaces in pre-verbal position in all the other finite constructions) is argued to be due to a coalition of factors. One of them is phonological. Avram (1986) points out that sandhi is optional with the masculine clitic when it precedes the lexical verb *avea* ‘have’ (i) whereas in front of the auxiliary verb *avea* ‘have’ the sandhi variant (the one which is also used in post-verbal position) is the only one which is allowed (ii) :

- (i) \bar{i} am / \bar{i} -am
 Acc clitic 3rd sg masc have/ Acc clitic 3rd sg masc have
 “I have it.”
- (ii) \bar{i} -am văzut / * \bar{i} am văzut
 Acc clitic 3rd sg masc have seen
 “I have seen him.”

Building on this observation, Avram (2000) suggests that the impossibility of placing the feminine clitic *o* in front of the auxiliary may indicate that it lacks a sandhi variant. *O* is the only pronominal which does not have different forms for the class of weak pronouns and for that of clitics within the tripartition proposed in Cardinaletti and Starke (1995).

already argued in the literature for weak pronouns in languages such as Dutch (Haegeman 1996) within a truncation approach (Rizzi 1995).

In the analysis which we adopt, clitics are Last Resort elements⁹. Last Resort operations are costly, language specific and consequently subject to delayed acquisition.

On the other hand, one of the properties which make clitics a special class is their contribution to the meaning of the structure in which they occur. The analysis which we propose takes clitics to be associated with D-linked topicality. In terms of acquisition, interpretable features are available at the interface and can be important facilitators. Previous acquisition studies showed that children are sensitive to information partitioning at an early age (see, for example, de Cat 2002, Okada and Grinstead 2003). If ADOCs have a D-linking function we can predict that children might identify their role early, and consequently we expect early emergence. Moreover, the presence of ADOCs in the input may represent an important clue with respect to the non-availability of null objects. ADOCs in Romanian could be important triggers in the setting of the null object parameter.

We have also seen that in the so-called clitic doubling constructions the presence of the clitic is optional with most DPs. There is variation in the input; the clitic may or may not surface with the same co-indexed DP. The immediate question is whether such variation in the input is helpful or, on the contrary, a delaying factor. Actually, only 3rd person Accusative clitics are subject to optionality. 1st and 2nd person ADOCs cannot be omitted, because they can only have a pronominal “double”¹⁰:

- (13) a. *(Te-) a văzut și pe tine la teatru.
 Acc clitic 2nd sg) has seen and *pe* you.Acc at theatre
 “He saw you too at the theatre.”
 b. *(Mă) cunoaște numai pe mine.
 Acc clitic 1st sg knows only *pe* me
 “He knows only me.”

The acquisition pattern will be different because the two classes of clitics are subject to different licensing conditions. 3rd person ADOCs are dependent for

⁹ The proposed unifying view of Romanian ADOCs as Last Resort elements across all the identified obligatory clitic contexts may lead to the prediction that all these clitic constructions are acquired simultaneously. This is actually the claim in Torrens and Wexler (2000) for Spanish. But our analysis does not imply that all the obligatory clitic constructions involve the same underlying structure and by no means do they involve the same degree of computational complexity. Nothing in our analysis forces the prediction that all the structures emerge simultaneously. Actually, if one assumes that computational complexity matters in the acquisition process, being intricately related to phase derivation and phase memory (Chomsky 1999), we do expect to find non-simultaneous emergence of these constructions. In the present study, we do not verify this prediction.

¹⁰ The difference between a clitic doubling and a single clitic construction in the case of 1st and 2nd person clitics is one of contrast; only the former will display a contrastive effect.

referential indexicality on a linguistic antecedent and are licensed syntactically; they inherit the Person feature of the null object whose Person index they spell out. Only 3rd person clitics are Last Resort copies of a D-linked null object created in the derivation. They are the only “determiner pronouns” (Kayne 2000). Identification of the Person index in their case is achieved via feature matching. The valuation of the Person feature in Inflection, which may disrupt the feature matching relation, is relevant in their case. This predicts delayed acquisition. 1st and 2nd person ADOCs are weak pronouns whose reference is determined by changing discourse roles. We take 1st and 2nd person clitics in Romanian as the weak variants of 1st and 2nd person pronouns when surfacing in a functional projection. Since they are pronouns, they do not have an antecedent in their local domain. Whether the child has or has not valued the Person feature of Inflection is not directly relevant for their referential index because they are inherently discourse licensed. This is why we predict a low omission rate in the case of 1st and 2nd person clitics. 1st and 2nd person ADOCs also display a lower number of specified features; unlike 3rd person clitics, they lack gender specification. The data suggest that morphological complexity might also play a part. This is why we expect agreement errors with morphologically complex clitics but no or very few errors with 1st or 2nd person ADOCs.

In a nutshell, our analysis of ADOCs makes at least the following predictions with respect to acquisition: (i) ADOCs emerge early; (ii) emergence is followed by a stage with a significant omission rate in the case of 3rd person ADOCs but low omission rate in the case of 1st and 2nd person ADOCs; (iii) (gender and number) substitution errors should be attested only in the case of 3rd person ADOCs.

5. The emergence of ADOCs

5.1 The Data

The data used in the analysis come from two longitudinal corpora of monolingual Romanian in electronic format. The two children, a girl (B.) and a boy (A.), are both monolingual speakers of Romanian. The data were collected for a period of 20 months at the children’s home on a weekly basis. Each session lasted approximately 60 minutes. The data were transcribed according to the CHAT format (as described in MacWhinney 2000). The recordings include situations of free interactions with the mother (or, occasionally another caretaker) and the investigator. The overall number of files examined for the analysis of ADOCs are given in Table 1:

Child	Number of files	Age
B.	26	1;05 - 2;10
A.	11	1;09 - 3;05

Table 1: Longitudinal data used in the analysis

5.2 Procedure

A detailed examination of each file was conducted in order to identify all the obligatory clitic contexts on the basis of the findings in section 3. For the structures with a null topic, the presence/absence of the clitic was evaluated on the basis of discourse situation. The presence of the clitic was evaluated as obligatory if the antecedent was mentioned in the previous 5 lines. Clitic structures with no overt antecedent in the same clause evaluated as target-like are illustrated in (14) and illicit clitic omissions are illustrated in (15):

(14) Am luat creion. Uite-l.
“I have taken pencil. Look at it.” [B. 2; 6.08]

(15) Adult: Da’ ce-ai făcut cu ea?
“But what did you do to it?”
Child: Am dezlipit *[o] așa.
have taken off [] like this
“I have taken [] off like this.” [A. 3;5.04]

Clitic omission with left dislocated direct objects was counted as illicit:

(16) uite, pe tati nu *[-l] împuşc.
look *pe* daddy no [] shoot
“Look, I’m not shooting daddy.” [B. 2;7.20]

In the case of (the rare) relative clauses with a relativized direct object, clitic omission counted as illicit:

(17) Adult: ăsta ce e?
“What is this?”
Child: Un brăduţ care *[-l] a adus Moş Crăciun.
a tree.diminutive which [] has brought Santa Claus
“A little tree which Santa Claus has brought .” [B. 2;7.20]

The omissions which involved a definite pronoun antecedent (clitics included) were the only ones counted as deviant in the analysis of clitic doubling constructions. All the other situations (proper names included) were evaluated as adult-like even when the same structure in the input (sometimes in the immediately preceding or following line) contained a clitic:

(18) Child: a luat pe babă.
has taken *pe* old woman
“He has taken the old woman.”

Adult: a luat- o pe babă?
 has taken clitic 3rd fem sg *pe* old woman
 “Has he taken the old woman?” [B. 2;3.23]

Clitics occurring in imitations or poetry lines, clitics occurring in formulaic structures as well as cases where the verb was not clear or null (even though, in principle, retrievable from the context) did not enter the countings.

Omission rates as well as rates of used clitics were calculated against the number of identified obligatory clitic contexts. Erroneous substitutions (gender, number) were calculated against the total number of used clitics.

5.3 Results and discussion

5.3.1 Early emergence and clitic drop

We examined the data with a view to testing the predictions of our analysis (subsection 4). The general results concerning ADOC emergence and early omission are given in Figures 1 and 2:

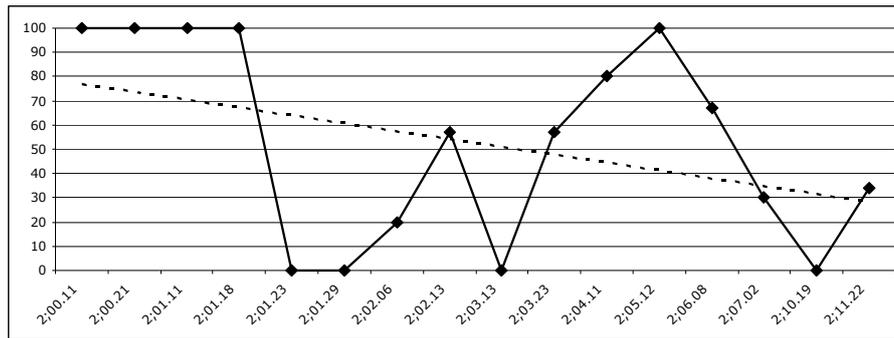


Fig. 1: ADOC omission in Romanian: the B. corpus

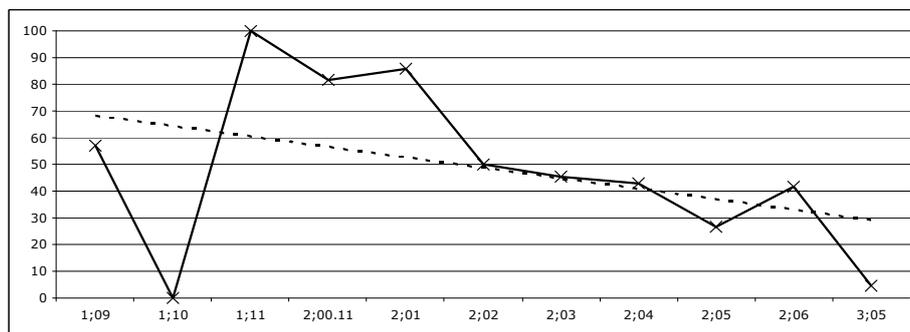


Fig. 2: ADOCs in Romanian: the A. corpus

The data indicate (i) early emergence of clitics followed by (ii) a high rate of omissions in obligatory contexts, just as predicted. In the B. corpus (Figure 1), where the recordings began early enough, at 1;3, there is a very short stage when clitics are completely absent. The first ADOC is attested at 2;0.03. In the A. corpus ADOCs are attested from the first recording session, at 1;9 (Figure 2). But in both corpora early emergence is followed by a clitic omission stage.

The Romanian data are comparable to previous findings with respect to the existence of an omission stage in the acquisition of ADOCs. The reported omission rates differ from one language to another, or from one study to another, but omissions have been reported for many languages (French, Italian, Spanish, Catalan, European Portuguese, Dutch). The data also reveal that as early as 2;3, both in the B. corpus and in the A. corpus, there is growing evidence that the children know that D-linked null objects are banned in the target language. Omissions are still attested but the number of target consistent constructions gradually increases, reaching a rate of over 90% at 2;10 in the B. corpus and at 2;11 in the A. corpus. The number of agreement errors is practically null. In this, our findings are again similar to those reported for other languages for which clitic omission in early language has been attested with 2 year olds (Schaeffer 2000, Ticio and Reglero 2002, Hamann 2002, Rasetti 2003, van Kampen 2004, Pîrvulescu 2006). One has to mention that, unfortunately, percentages are usually difficult to compare across languages, because of the different procedures adopted for the countings in the various studies. However, if we compare the developmental path of the two Romanian children in our study to the data reported for the French children Max and Anne (the York corpus) in Pîrvulescu (2006) (where a similar counting methodology was used), the percentages for the early recording sessions reveal a low rate of used ADOCs with all the children. The French children reach a minimum rate of used ADOCs of 50% at around 2;4 and 2;5, respectively, when their MLU is higher than 3. The Romanian children reach a 50% rate slightly earlier, when their MLU is lower than 2. The increase in ADOC use coincides with the improvement/growth of the MLU. The Romanian children reach an adult-like rate of correct use in obligatory context of minimum 90% at around 2;10, when their MLU is still lower than 3, whereas the French children reach this rate at 2;5 and 2;10, respectively, when their MLU is approximately 3;5. It is not irrelevant to add that in some cases the MLU can be misleading. B.'s omission rate drops to 50% at 2;2.13, when her MLU is 1.819. According to MLU5, though, at this stage B. is already in the multi-word stage. The recording contains utterances which are at least 5 free morpheme long:

(19) vreau să strig la rața # rața!
 want subj. marker shout at duck.the duck.the
 "I want to call the duck # hey, duck!"

These differences, if real, indicate different acquisition speed for various languages, which may reflect the different status of clitics in the two systems.

5.3.2 Substitution errors

The first attested clitic is, in both corpora, the feminine singular clitic *o* ‘her’ placed in postverbal position:

- (20) Adult: ce-ai făcut cu ligheanul?
 “What have you done to the bowl?” (bowl = **masc**)
 Child: spart- **o**.
 broken Acc clitic 3rd sg **fem** [A. 1;9]

A closer look at the distribution of clitics reveals that actually postverbal *o* is the only one which is used until 2;3.14 in the A. corpus and until 2;1.23 in the B. corpus, irrespective of the features of the antecedent (see Tables 2 and 3).

Age	1;9	2;0	2;1	2;2	2;3	2;4	2;5	2;6	2;6.30	2;7
Total ADOCs	6	4	2	2	11	5	6	7	21	23
Post-verbal <i>o</i>	6	4	2	2	8	5	5	2	9	9
Post-verbal other than <i>o</i>	0	0	0	0	0	0	1	0	0	0
Pre-verbal	0	0	0	0	3	0	5	5	12	14

Table 2: Romanian: pre- and post-verbal ADOCs in the A. corpus

Age	2;0	2;0.21	2;1.11	2;1.23	2;1.29	2;2	2;3
Total ADOCs	2	4	2	5	1	2	2
Post-verbal <i>o</i>	2	4	2	0	0	0	0
post-verbal other than <i>o</i>	0	0	0	1	0	0	0
Pre-verbal	0	0	0	4	1	2	1

Table 3: Romanian: pre- and post-verbal ADOCs in the B. corpus

The fact that *o* is exceptional among the other Romanian ADOCs (it is the only one which can occur in postverbal position in a limited number of finite constructions), corroborated with the exclusive use of *o* in (canonical) argument position during this early stage, might indicate that this has to be taken as a language specific phenomenon, rooted in the properties of the feminine clitic *o*. However, a similar phenomenon is discussed in Dominguez (2003). She observes that the Spanish child Maria goes through a stage when one clitic is used as a “substitute”; the masculine *lo*¹¹ is used instead of the feminine *la*, instead of the plural *los*, and even when no clitic is required. Dominguez reports that at 2;3 Maria uses 21 clitics *lo*, out of which 8 are substitutes. The data coming from Romanian and Spanish are very similar in this respect. One has to notice that the corpus which Dominguez

¹¹ *Lo* is the default form in adult Spanish as well.

pronominals) we also analyzed the available data with a view to testing whether their different properties are reflected in their acquisition pattern. Tables 4 and 5 below compare the early omission rate in the case of 1st and 2nd person clitics against the general omission rate in the two corpora.

Age	2;1.18	2;1.23	2;1.29	2;2.13	2;3.13	2;4.11	2;5.12	2;5.18
1 st -2 nd person ctx	5	5	1	1	3	4	6	3
1 st -2 nd omissions	0	0	0	0	0	0	0	0
General omissions	0	57.14 %	80%	50%	30%	25%	18.75 %	21%

Table 4: General ADOC omission vs. omission of 1st and 2nd person ADOCs: the B. corpus

The data reveal an obvious difference between the acquisition pattern of 1st/2nd person ADOCs and 3rd person ADOCs.¹³ While the omission rate in the case of the latter is relatively high in the early files and gradually decreases, the rate of attested omissions in the case of 1st and 2nd person ADOCs is almost 0. Our results are not singular. They confirm previous findings in the literature. van Kampen (2004) argues, on the basis of data coming from Dutch and French, that 1st and 2nd person pronouns are acquired earlier because they do not require a linguistic antecedent in the discourse; they have a referent in the speech situation.

5.3.5 Summing up

The data indicate that the acquisition of ADOCs in Romanian comprises three stages: (i) a very short ‘no clitic’ stage; (ii) early emergence of clitics [1;9 in the Antonio corpus, i.e. from the first recording session, and 2;0 in the Bianca corpus] used first as “fillers” or default forms; (iii) after the emergence of clitics, children randomly omit them at a rate higher than 50%, they make (very few) agreement or case errors, and there is an obvious asymmetry between the omission rate of 3rd person clitics, on the one hand, and 1st and 2nd person clitics on the other hand. The attested agreement errors refer only to 3rd person clitics. The rate of clitic omission gradually decreases with age. At about 2;10 (MLU still < 3) in the B. corpus and at approximately 3 (MLU<3) in the A. corpus, the omission rate decreases significantly, dropping below 10%.

¹³ Notice that this may be one of the reasons for which omission rates in elicited production studies are higher. The only elicited form is the 3rd person singular, which is omitted at a higher rate in the naturalistic data as well (vs. 1st and 2nd person clitics).

6. Conclusions

In the present paper the hypothesis that ADOCs reflect a ban on null D-linked direct objects in Romanian has been put forth. This property of Romanian has been argued to derive from the pronominal nature of Inflection. The Person feature of Inflection blocks identification of the Person index of a null object in postverbal position via feature matching with the antecedent placed in SpecTopP. 3rd person ADOCs have been defined as Last Resort elements created in the derivation, which have to move to a position higher than Inflection in order to allow the identification of the Person index of the null D-linked object. Such an analysis allowed us to make several predictions with respect to acquisition. ADOCs are associated with referential stability and topicality: they are obligatory when the antecedent of the null object has been previously mentioned, either in the same clause or in the preceding discourse. Since children seem to be sensitive to discourse information at an early stage, we predicted that this property will be helpful in the acquisition process and lead to early emergence of clitics. This prediction has been shown to be on the right track. ADOCs emerge very early (1;9 in the case of A., 2;0 in the case of B., $MLU < 2$).

On the other hand, the analysis of clitics as Last Resort elements makes the prediction that their acquisition requires investigation of the input, being a language specific property. This predicts delayed acquisition. This prediction, again, is borne out by our data. The early emergence is followed by non-target ADOCs: (i) an early stage, extremely short, when a default clitic is placed exclusively in postverbal position; (ii) a stage when other clitics, placed both in pre- and post-verbal position are attested but when children randomly omit these pronominal elements. It is also during this stage that rare substitution errors are attested. ADOCs begin to be used in a target-like way at the age of approximately 3;0, when the MLU is almost 3. The data reported in Avram and Coene (2006) provide evidence that for Romanian, the activation of the C-system seems to be a prerequisite for the acquisition of ADOCs, which emerge only after the emergence of those elements which indicate the availability of an active C-domain. This is also predicted by our analysis, since the valuation of the Person feature is only possible when the C-domain is active. The specification of Person features (the *phi*-features on T) is inherited from C (Chomsky 2005).

Our analysis made one more important prediction, which built on the difference between various ADOCs. Only 3rd person clitics can be defined as Last Resort copies of a D-linked null object created in the derivation. It is only in their case that the identification of the Person feature is achieved via matching. The reference of 1st and 2nd person pronouns is determined by changing discourse roles. Whether the child has or has not valued the Person feature of Inflection is not directly relevant for their referential index because they are inherently discourse licensed. This predicts delayed acquisition of 3rd person ADOCs, but not of 1st and 2nd person ADOCs. The prediction has been borne out by the data. There is practically no omission rate in the case of 1st and 2nd person clitics and there are no gender/number substitutions in their case. This finding may shed light on why clitic acquisition

studies relying on experimental data report a higher rate of omissions than those relying on longitudinal data. The former usually look only at the use of 3rd person clitics, whereas the latter examine all types of ADOCs.

The data examined in the present study indicate that the acquisition of ADOCs is problematic and hence delayed in Romanian. Children's use of clitics is not constant for all stages and not comparable to that of adults until the age of 3. One has to mention though that, in spite of the obvious delay, when one looks at the emergence and acquisition of ADOCs against the complexity of the construction, the accomplishment seems extraordinary. Around the age of 3 Romanian children use ADOCs correctly in over 90% of the obligatory clitic contexts.

References

- Alboiu, G. (2002) *The Features of Movement in Romanian*. Bucharest: Editura Universității din București.
- Androulakis, A. (2001) 'Clitics and doubling in Greek', in: M. Georgiafentis, M. P. Kerswill and S. Varlokosta (eds.) *Reading Working Papers in Linguistics*, 5, 85-111.
- Avram, A. (1986) 'Sandhi phenomena in Romanian', in: A. Henning (ed.) *Sandhi Phenomena in the Languages of Europe*. Berlin/New York/Amsterdam: Mouton de Gruyter, 551-574.
- Avram, L. (2000) 'From possessive clitics to object clitics: a unifying analysis', in: L. Tasmowski (coord.) *The Expression of Possession in Romance and Germanic Languages*. Cluj-Napoca: Clusium, 83-100.
- Avram, L. (2001) 'Remarks on the optional clitic stage in child Romanian'. *Bucharest Working Papers in Linguistics*, Vol. III, Nr.1: 16-28.
- Avram, L. and M. Coene (2000) 'Genitive/Dative clitics as Last Resort', in: M. Dimitrova-Vulchanova and L. Hellan (eds.) *Proceedings of the 3rd FASSBL Conference. Trondheim Working Papers in Linguistics*, 34, 157-169.
- Avram, L. and M. Coene (2003) 'From feature selection to feature valuation: the view from early clitics and determiners'. Paper presented at GLOW 2003, Workshop II: Learnability from a generative perspective, Lund, April 12.
- Avram, L. and M. Coene (2004) 'What early clitics can tell us about early subjects', in: J. van Kampen and S. Baauw (eds.) *Proceedings of Gala 2003*. LOT, Utrecht, 93- 102.
- Avram, L. and M. Coene (2005) 'On the acquisition of finiteness', in: M. Coene and L. Tasmowski (eds.) *On the Expression of Time and Space*. Cluj-Napoca: Clusium, 151-172.
- Baauw, S. (2000) 'Grammatical Features and the Acquisition of Reference. A Comparative Study of Dutch and Spanish'. PhD dissertation, LOT Utrecht University.
- Cardinaletti, A. and M. Starke (1995) 'The typology of structural deficiency: On the three grammatical classes'. *FAS Papers in Linguistics*, vol. I, 1-55.
- Chomsky, N. (1977) 'On *wh* movement', in: P. Culicover, T. Wasow and A. Akmajian (eds.) *Formal Syntax*. New York: Academic Press, 71-132.
- Chomsky, N. (1999) 'Derivation by Phase'. *MIT Occasional Papers in Linguistics* 18. Cambridge, MA: MITWPL.
- Chomsky, N. (2000) 'Minimalist inquiries: the framework', in: R. Martin, D. Michaels and J. Uriagereka (eds.) *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik*. Cambridge MA: MIT Press, 89-155.
- Chomsky, N. (2005) 'On phases', ms. MIT.

- Cinque, G. (1990) *Types of A'-dependencies*. Cambridge, MA: MIT Press.
- Coene, M. and L. Avram (2004) 'Romanian early root verbal inflection is optionally /-finite/'. Paper presented at *Going Romance 2004*, Utrecht, December 10.
- Costa, J. and M. Labo (2005) 'Clitic omission, null objects or both in the acquisition of European Portuguese?'. Paper presented at *Gala 2005*, Siena, September 8.
- De Cat, C. (2002) *French Dislocation*. PhD dissertation, University of York, UK.
- Delfitto, D. (2002) 'On the semantics of pronominal clitics and some of its consequences'. *Catalan Journal of Linguistics* 1, 29-57.
- Dominguez, L. (2003) 'Interpreting reference in the early acquisition of Spanish clitics', in: S. Montrul and F. Ordóñez (eds.) *Linguistic Theory and Language Development in Hispanic Languages*. Somerville, MA: Cascadilla Press, 212-228.
- Farkas, D. (1978) 'Direct and indirect object reduplication in Romanian.' *Papers from the 17th Regional Meeting of the Chicago Linguistic Society (CLS)*. University of Chicago, 88-97.
- Farkas, D. (2000) 'Varieties of definites'. Paper presented at *From NP to DP. Syntax and the Pragsemantics of the Noun Phrase*. Antwerp University, February 11.
- Farkas, D. and K. Heusinger (2003) 'Referential stability and differential case marking in Romanian'. Paper presented at the XV European Summerschool in Logic, Language and Information (ESSLLI), Workshop *Direct Reference and Specificity*, Vienna, Austria, August 21.
- Guasti, T. (1993/1994) 'Verb syntax in Italian child grammar: finite and non-finite verbs.' *Language Acquisition* 3, 1-40.
- Guțu-Romalo, V. (ed.) (2005) *Gramatica limbii române*. Volume II: *Enunțul*, Bucharest: Editura Academiei Române.
- Haegeman, L. (1996) 'Root infinitives, clitics and truncated structures', in: H. Clahsen (ed.) *Generative Perspectives on Language Acquisition*. Amsterdam/Philadelphia: John Benjamins, 271-308.
- Hamann, C. (2002) *From Syntax to Discourse. Pronominal Clitics, Null Subjects and Infinitives in Child Language*. Dordrecht/Boston/London: Kluwer Academic Publishers.
- Hamann, C., L. Rizzi and U.H. Frauenfelder (1996) 'The Acquisition of subject and object clitics in French', in: H. Clahsen (ed.) *Generative Perspectives on Language Acquisition*. Amsterdam/Philadelphia: John Benjamins, 309-334.
- Hurtado, A. (1982) 'Clitic chains', ms. MIT.
- Jakubowicz, C., L. Nash, C. Rigaut and C.-L. Gerard (1998) 'Determiners and clitic pronouns in French-speaking children with SLI'. *Language Acquisition*, 113-160.
- Kampen, J. van (2004) 'Learnability order in the French pronominal system', in: R. Bok-Bennema, B. Hollebrandse, B. Kampers-Manhe and P. Sleeman (eds.) *Selected Papers from Going Romance 2002*. Amsterdam/Philadelphia: John Benjamins, 163-183.
- Kayne, R. (1994) *The Antisymmetry of Syntax*. Cambridge, MA.: MIT Press.
- Kayne, R. (2000) 'A note on clitic doubling in French', in: R. Kayne *Parameters and Universals*. Oxford: Oxford University Press, 163-184.
- Landau, B. and L. Gleitman (1985) *Language and Experience*. Cambridge, MA.: Harvard University Press.
- Longobardi, G. (2004) 'On the syntax of denoting'. Paper presented at *Going Romance 2004*. Leiden, December 10.
- MacWhinney, B. (2000) *The CHILDES Project: Tools for Analyzing Talk*. Volume 1 *Transcription Format and Programs*. Volume 2: *The Database*. 3rd edition. Mahwah, NJ: Lawrence Erlbaum.

- Massam, D. and Y. Roberge (1989) 'Recipe context null objects in English'. *Linguistic Inquiry* 20(1), 134-139.
- Okada, K. and J. Grinstead (2003) 'The emergence of the CP in child Japanese', in: J. M. Liceras et al. (eds.) *Proceedings of the 6th Generative Approaches to Second Language Acquisition Conference (GASLA 2002)*. Somerville, MA.: Cascadia, 213-218.
- Perez-Leroux, A. T., M. Pîrvulescu and Y. Roberge (2006) 'Early object omission in child French and English', in: C. Nishida and J.-P. Montreuil (eds.) *New Perspectives in Romance Linguistics*. Amsterdam/Philadelphia: John Benjamins, 213-220.
- Pîrvulescu, M. (2006) 'The acquisition of object clitics in French L1: spontaneous vs. elicited production', in: A. Belletti, E. Bennati, C. Chesi, E. DiDomenico and I. Ferrari (eds.) *Language Acquisition and Development. Proceedings of GALA 2005*. Newcastle: Cambridge Scholars Press, 450-462.
- Pîrvulescu, M. and Y. Roberge (2005) 'Licit and illicit null objects in L1 French', in: R.S. Gess and E.J. Rubin (eds.) *Theoretical and Experimental Approaches to Romance Linguistics*. Amsterdam/Philadelphia: John Benjamins, 96-109.
- Postal, P. (1966) 'On so-called "pronouns" in English', in: D. Reibel, D. and Schane, S. (eds.), *Modern Studies in English*, Englewood Cliffs, N.J.: Prentice-Hall, 201-223.
- Rasetti, L. (2003) *Optional Categories in Early French Syntax: a developmental study of root infinitives and null arguments*. Unpublished PhD, University of Geneva, Switzerland.
- Rizzi, L. (1993/1994) 'Some notes on linguistic theory and language development.' *Language Acquisition* 3, 371-394.
- Schaeffer, J. (2000) *The Acquisition of Direct Object Scrambling and Clitic Placement. Syntax and Pragmatics*. Amsterdam/Philadelphia: John Benjamins.
- Schaeffer, J. and D. Ben Shalom (2004) 'On root infinitives in child Hebrew'. *Language Acquisition* 12(1), 83-96.
- Ticio, E. and L. Reglero (2002) 'Evidence for AgrO in child grammar: English vs. Spanish', J. Costa and M. J. Freitas (eds.) *Proceedings of Generative Approaches to Language Acquisition*. Lisboa: Associação Portuguesa de Linguística.
- Torrens, V. and K. Wexler (2000) 'The acquisition of clitic doubling in Spanish', in: Powers, S. and C. Hamann (eds.) *The Acquisition of Scrambling and Cliticization*. Dordrecht/London/Boston: Kluwer Academic Publishers, 279-298.
- Wexler, K., A. Gavarró and V. Torrens (2004) 'Feature checking and object clitic omission in child Catalan', in: R. Bok-Bennema, B. Hollebrandse, B. Kampers Mahne and P. Sleeman (eds.) *Romance Languages and Linguistic Theory 2002*. Amsterdam/Philadelphia: John Benjamins, 253-268.

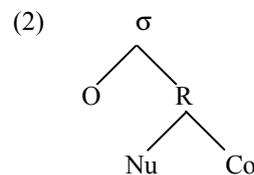
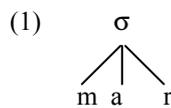
The Role of Syllable Structure in the Acquisition of Brazilian Portuguese

Giovana Bonilha, Carolina Lisbôa Mezzomo, Regina Ritter Lamprecht
Universidade Federal de Santa Maria, FFFCMPA, Pontifícia Universidade Católica
do Rio Grande do Sul

1. Introduction

According to Fery and Van de Vijver (2003), the 70's were characterized, in the phonological studies, as the time when the syllable was considered as a prosodic unit, especially in the studies of Vennemann (1974), Hooper (1976) and Kahn (1976).

There are two main lines of thought concerning the syllabic structure analysis: the first is the one developed by Kahn in 1976 and the second by Selkirk in 1982. For Kahn, the segments are directly connected to the syllabic node, then the rules act on the syllable as a whole, as shown in (1). According to Selkirk, the relationship among the constituents takes place in a distinguished way, as the rules are applied to just one of the elements: *onset* (O), *rime* (R), *nucleus* (Nu) or *coda* (Co), as shown in (2).



According to Freitas (1997), Vigário and Falé (1993) and Mateus (1993) proved the working of the model proposed by Selkirk (1982) for the description and analysis of the syllable in European Portuguese. The same can be said in relation to the acquisition of Dutch (Fikkert, 1994).

The acquisition of the syllabic constituents of Brazilian Portuguese has been the research focus for many papers such as Mezzomo (2004), Bonilha (2000) and Ribas (2002), among others. Most papers make use of cross-sectional data, offering results about the ordering in the acquisition of syllabic structure and trying to contribute to the description of Portuguese.

Nevertheless, the study of longitudinal data and the possibility of observing the interaction of different phonological units in the acquisition process, like segment, syllable and foot structure, for example, makes it necessary to rethink the syllable structure acquisition in Brazilian Portuguese.

On the bases of data of two longitudinal children aged 1;1 to 3;9, the present paper rethinks the role of syllabic structure in the segmental acquisition of Brazilian Portuguese by making use of Optimality Theory (Prince & Smolensky, 1993) and by

the application of the gradual learning algorithm proposed by Boersma and Hayes (2001).

The present study will focus on three main questions:

(i) What is the role of the syllabic structure in the phonological acquisition of Portuguese? Which evidence found in the child data indicates that these structures have been acquired?

(ii) Does the segmental acquisition drive the syllabic structure acquisition (bottom-up) or is it the syllabic structure acquisition which prevents the segmental acquisition (top-down)?

(iii) In terms of OT, is there any evidence in the data for the demotion of constraints related to the syllabic structure, such as NoCoda and NotComplex(nucleus)? Does the demotion of these constraints happen in a gradual way, as it does with the feature constraints, or does the raking just show the gradual demotion of feature constraints below the faithfulness constraints?

2. The syllable in Brazilian Portuguese

According to the literature, there are 15 syllabic patterns in Portuguese.

(3) a. Open syllables	b. closed yllables
V <u>á</u> gua	CVC <u>lar</u>
CV <u>pá</u>	VC <u>ar</u>
CCV <u>abre</u>	CCVC <u>três</u>
CVV <u>pai</u> ¹	CVCC <u>monstro</u>
CCVV <u>grau</u>	CCVCC <u>trens</u>
VV <u>oi</u>	VCC <u>instante</u>
	CCVVC <u>graus</u>
	CVVC <u>dois</u>
	VVC <u>austero</u>

The patterns shown in (3) are formed by the following syllabic constituents: simple onset, complex onset, simple nucleus, complex nucleus, simple coda and complex coda.

The medial onset can be formed by 19 consonants and the initial one by just 16, as the language has a filter that prevents the distribution of the elements /p/ and /k/ and /t/ at the beginning of the word.

In coda position, four consonants can be produced, the nasal, which assimilates the point of articulation of the following consonant, the coronal fricative, which

¹ A syllable that presents a falling diphthong is not considered closed because the glide is in a complex nucleus, according to Câmara Jr. (1977), Cristófaró-Silva (1999) and Bonilha (2000).

assimilates the sonority value of the subsequent consonant, a coronal lateral liquid, which is produced as a dorsal glide in many dialects, and the non-lateral liquid.

Another limitation lies in the distribution of the segments in complex onsets. According to Bisol (1999), the onset position is formed of up to two elements. According to the author, in Brazilian Portuguese, the complex *onset* must be formed by elements that have a minimum distance of two points in the sonority scale, as the first consonant will always be a plosive or a labial fricative, while the second one will always be a lateral or non-lateral liquid.

It is worthwhile mentioning that some authors put the glide in a complex nucleus in Brazilian Portuguese, such as Câmara Jr. (1977).

In (4) the segments and consonantal sequences allowed by Portuguese in a simple onset, complex onset, simple coda and complex coda are presented.

(4)	Constituents	Segments and consonantal sequences
	Initial onset	/p/, /b/, /t/, /d/, /k/, /g/, /f/, /v/, /s/, /z/, /ʃ/, /ʒ/, /ç/, /m/, /n/, /l/
	Medial onset	/p/, /b/, /t/, /d/, /k/, /g/, /f/, /v/, /s/, /z/, /ʃ/, /ʒ/, /ç/, /m/, /n/, /ɲ/, /l/, /ʎ/, /r/
	Coda	/N/, /l/, /S/, /r/
	Complex coda	ns, rs, ls
	Complex onset	pr, br, tr, dr, kr, gr, pl, bl, tl, kl, gl, fl, fr, vr

3. Syllabic structure acquisition

Fikkert (1994) presents the first acquisition model of prosody that integrates the syllable as a unit composed of hierarchically organized constituents. The author proposes, based on Principles and Parameters framework, universal stages for the acquisition of the syllable. In (5) we give the syllable acquisition stages proposed for Dutch:

- (5) a. Acquisition stages of simple onsets
 Stage I: simple onsets (occlusives)
 Stage II: empty onsets
 Stage III: simple onsets (other consonants)
 Stage III a: nasals in simple onset
 Stage III b: other articulation manners in simple onsets
- b. Acquisition stages of complex onsets
 Stage 0: selection strategy – there are no targets formed by complex onsets
 Stage 1: deletion of the 2nd element that constitutes the onset
 Stage 2: deletion of the 1st element that constitutes the onset
 Stage 3: production of complex onset

- c. Acquisition stages of rime
 - Stage I: simple nuclei
 - Stage II: branching rime (final obstruent)
 - Stage III: branching nuclei
 - Stage IV: extra-rime consonants

Freitas (1997) analyses European Portuguese data and notices that Portuguese children exhibit basically the same stages proposed by Fikkert (1994), which corroborates the universality of these developmental stages.² Some differences in relation to the acquisition stages of simple onset were found, as in EP stage 1 includes the production of empty onsets and nasals.

According to the author, the data observed in European Portuguese acquisition provided evidences in favor of the syllabic representation model that sees the syllable as a hierarchically organized unit composed of internal constituents.

According to Freitas (1997), the segmental emergency depends on the availability of the functioning of the syllabic constituents, as some segments appear firstly in specific positions and only later on in others. In (6) we give the results found by the author, based on longitudinal/cross-sectional data.

- (6) a. Fricatives
 - 1. Simple onset – MINIMUM ONSET PARAMETER
 - 2. Coda – BRANCHING RIME PARAMETER
- b. Liquids
 - 1. Simple onset – MINIMUM ONSET PARAMETER
 - 2. Syllable end – BRANCHING NUCLEUS PARAMETER
 - 3. Complex onset – MAXIMUM ONSET PARAMETER

As can be observed in (6a), the coronal fricative, /s/, is first produced in simple onsets, as the YES value of the BRANCHING RIME PARAMETER has not yet been activated. According to (10b), the liquids, /l/ and /r/, are also first produced in simple onsets, and later on, with the activation of the YES value for the BRANCHING NUCLEUS PARAMETER, they are produced at the end of syllable. Later on, with the activation of the YES value for the MAXIMUM ONSET PARAMETER, the liquids are produced in complex onsets.

The emerging order of the fricatives and the liquids follows, therefore, the activation order of the parameter. It is not because the production of a segment is possible in the child's system that it appears simultaneously in different syllabic constituents. Then, according to Freitas (1997), the fact that the segmental emergence depends on the activation of the parameters relates to the syllabic

² According to Freitas (1997:280), stage IV should not be considered in the Rime acquisition of EP, as the parameter of the position extra-Rime is not activated in the target system.

structure, constitutes evidence in favor of a top-down type of phonological acquisition model.

In terms of OT, it is as if we stated that the constraints related to the features have already been demoted in the hierarchy of the child, but specific segments are not produced in specific syllabic positions because constraints such as NoCoda, NotComplex(nucleus) and NotComplex(onset) continue ranked above faithfulness constraints.

The idea that the acquisition of a segment is intrinsically connected to the position it occupies in the syllable and in the word is shared by many studies on the phonological acquisition of Brazilian Portuguese based on cross-sectional data. Lamprecht (1986, 1990) and Miranda (1996) showed in their studies that the syllabic position occupied by the segment is decisive in considering the segmental acquisition process finished.

Miranda's (1996) study on the acquisition of /r/ in coda position in Brazilian Portuguese, in which they analyzed the cross-sectional data of 110 children aged 2:0 to 3:9, found that the position that /r/ occupies in the syllable and also in the word is the most important factor for its acquisition. According to the author, a non-lateral liquid is first acquired in the coda, then in simple onsets and finally in complex onsets. The early acquisition in coda position is justified by the phonic salience of segments at the end of the word.

It is interesting to point out that such observation does not mean that the segmental acquisition is driven by the acquisition of the syllabic structure, as was clearly observed by Fikkert (1994) and Freitas (1997) for the Dutch and the European Portuguese data respectively. In this case there is no non-production of the segment in a specific position, which would corroborate the role of the syllabic structure in the acquisition process. The reasons that motivate the percentage variations across the different age groups can, however, be different ones. Among them, we can consider possible interferences of the cross-sectional data in the interpretation of the results.

The results found in relation to the acquisition of the syllabic structures in Portuguese do not always converge. Bonilha (2000) proposes four stages in the acquisition of syllabic constituents – simple onset, complex nucleus, coda and complex onset – as she considers that the glides are positioned in complex nucleus in Portuguese, based on arguments of the acquisition process of the language. The author's proposal, however, is not in agreement with the universal developmental stages defended in Fikkert (1994) and Freitas (1997).

According to Mezzomo (2004), the glide that forms the falling diphthongs is in syllabic codas in Portuguese, in agreement with Collischonn (1997) and Bisol (1999). Therefore, the second acquisition stage proposed by Bonilha does not exist. So, the universal ranking proposed by Fikkert (1994) is maintained. For Mezzomo (2004), the glides are positioned in syllabic coda, as they appear as the first filling elements of this constituent, due to aspects related to sonority scale. According to the author, and in agreement with the findings of the referred study, the coda acquisition is linked to the segmental emergence, hence there are no specific parameters acting in the acquisition of the different codas of Portuguese. In order to

make this acquisition occur, it is just necessary that the Branching Rime Parameter have its YES marked value activated.

Bonilha (2000) tries, adopting the falling oral diphthongs acquisition analysis, to corroborate the positioning of the glide in the syllabic nucleus. Some observed aspects had a special relevance: (i) the fact that the acquisition of the phonological diphthong [aw], derived from the sequence /au/, and of the diphthong [aw], derived from the semi vocalization of /l/, shows many significant differences,³ and (ii) the fact that the stabilization of the falling diphthongs is connected to the difference in height between the base vowel and the glide that form the diphthong.⁴ According to Booij (1989) the glide will be positioned in the coda or in a complex nucleus depending on the relationship that it will have with the nucleus or with the syllabic coda. In the acquisition data, there is a connection between the syllabic nucleus and the glide, since the acquisition of the diphthongs is connected to the different sequences of segments that form them.

4. Methodology

The subjects of the current study are two children, from now on S1 and S2, with normal phonological development, in the age group of 1;1 to 3;9 years old. Both are Portuguese-speaking monolingual girls.

The data used in this study were selected from the *INIFONO databank*. The selected subjects present phonological systems considered normal until the time of the data recollection and have normal hearing, perception, neurological, motor and cognitive abilities. The data that form the sample of this study were obtained from interviews, at the informants house, and recorded on tape. The data gathering happened spontaneously, while the children were interacting with the family and the interviewers. The analysis was based on the words produced spontaneously.

In line with other studies in the field, a segment was considered to be acquired if its correct use reaches 80%.

In relation to S1, the data were recollected in 32 moments considering periods of 15 days, from 1;02 to 1;08, and 1 month in the other age groups.

In line with Fikkert (1994), the S1 data were transcribed just on a perceptual basis. The absence of the acoustic analysis, however, does not seem to have a significant effect on the results found.

A longitudinal approach was chosen instead of a cross-sectional approach, because we focus on the presence of markedness constraints related to different

³ According to Bonilha (2000), the phonological diphthong [aw] emerges and stabilizes in a stage before the diphthong [aw], derived from the semi vocalization of /l/ in the coda. It seems to show that they occupy different syllabic positions, in the complex nucleus and the coda, respectively.

⁴ The data analyzed by the author showed that the acquisition of diphthongs is basically related to the height of the base vowel that constitutes them, that is, diphthongs formed by low vowels are acquired in a previous stage to those formed by medium and high vowels.

phonological units as segments, and on syllabic structure. Only by studying longitudinal data it is possible to establish whether some segments, although acquired, are not produced by the presence of constraints related to the syllabic structure.

The minimum age of 1:0 for this study was set because around this age children start producing their first words; the maximum age, 3:9, is justified because at this age the phonological system of Portuguese is practically acquired. Lamprecht (1990) observed that except for the consonant clusters, all the phonological processes are acquired by the age of 4:1 or 4:2. It is also important to mention that other studies involving the syllabic structure acquisition used similar age groups, e.g., Fikkert (1994) and Freitas (1997).

5. Data Analysis

In order to answer questions (i) and (ii), we must consider the segments that can form different syllabic constituents in Portuguese, such as the consonants /l/, /s/, /r/ and /n/, which can occupy the positions of onset and medial coda in Brazilian Portuguese.

If there is a role of the syllabic structure in the segmental acquisition of S1, it is expected, for example, that emergence and acquisition of the segments take place first in specific syllabic positions, in line with what was observed by Freitas (1997) for European Portuguese data.

In table 1 we provide the results obtained in relation to the segmental emergence and acquisition of S1.

Segments	Emergence					Acquisition				
	Final coda	Med coda	Init onset	Med onset	Comp onset	Final coda	Med coda	Init onset	Med onset	Comp onset
N	-	1:4	1:5	1:5	-	-	1:7	1:7	1:7	-
L	1:6	2:1	1:4	1:4	2:2	1:6	2:7	1:8	1:8	2:2
R	1:10	2:9	-	1:10	2:7	3:0	3:2	-	3:2	3:2
S	1:6	1:10	1:6	1:6	-	3:2	3:2	3:2	3:1	-

Table 1 – Segmental emergence and acquisition

The final nasal coda was not considered in the results in table 1, as in this position it is produced as a diphthong in Portuguese. Mezommo (2004) argues that the coronal coda is the most frequent one in the data. According to this author, out of 441 production possibilities of the nasal coda in medial position produced by the subjects of her study, 297 corresponded to the coronal coda, that is, 67,34%. The same was observed in relation to the data of S1.

According to table 1, the coronal nasal emerges at around 1:4 in the onset and in syllabic codas, and the acquisition age was exactly the same in onsets and codas, namely 1:7. Observe the data in (7).

- (7) Some productions of the nasal in onset and medial coda by S1
- a. tampar [ko'pa] ~ [kõ^m'pa] – 1:4:22 – *to buy*
 - b. banana ['ba] – 1:5:7 – *banana*
 - c. boneca [ta'tɛka] – 1:5:20 – *doll*
 - d. pronto ['potu] – 1:5:20 – *it's done*
 - e. conto [kõⁿtu] – 1:6:3 – *tale*
 - f. não [nãw̃] – 1:5:20 – *no*
 - g. nenê [ne'ne] – 1:5:20 – *baby*
 - h. banana [nana] – 1:6:17 – *banana*
 - i. senta [ʃẽⁿta] – 1:6:17 – *sit down*
 - j. conto [kõⁿtu] – 1:7:1 – *tale*
 - k. boneca ['neka] – 1:7:1 – *doll*

In the acquisition of the nasal, therefore, there was no evidence concerning the role of syllabic structures as simple onsets and codas in the analyzed data.

The acquisition of the non-lateral liquid /r/ also occurs simultaneously in the different syllabic constituents, at around 3:2. In relation to this segment, two aspects must be highlighted. The first one is that /r/ emerges only at 2:9 in medial coda and at 1:10 in final coda. The emerging age in medial coda also goes beyond the emerging age in complex onsets that occurs at 2:7. Such fact has been explained in the literature – see Yavas (1988), Miranda (1996) and Mezzomo (1999) for data of Brazilian Portuguese. The end of the word is the most favorable position for the production of 'r', as can be observed in (8).

- (8)
- a. gordo ['gudu] – 1:7:1 – *fat*
 - b. garagem [ga'aʒi] – 1:9:9 – *garage*
 - c. quero ['kelu] – 1:9:9 – *I want*
 - d. agora [a'gɔla] – 2:1:27 – *now*
 - e. comprei [kõmpej] – 2:7:12 – *I bought*
 - f. dormi ['dumi] – 2:7:12 – *I slept*
 - g. jogador [ʒoga'dor] – 2:7:12 – *player*
 - h. três ['tres] – 2:7:12 – *three*
 - i. porque [pur'ke] – 3:2:28 – *because*
 - j. abri [a'bri] – 3:2:28 – *I opened*
 - k. estrela [is'trela] – 3:2:28 – *star*
 - l. agora [a'gɔra] – 3:2:28 – *now*

The simultaneous acquisition of the non-lateral liquid in all syllabic positions does not make it possible to verify the role of the syllabic constituents as a simple onset, coda and complex onset in the data of S1.

The data of another longitudinal subject, S2, can corroborate how much the segmental acquisition seems to drive the acquisition of the syllabic constituents. In the analysis of spontaneous data recollected at 3:1:26 e 3:2:15, it was observed that

S2 produces the complex onsets formed by the lateral liquid and reduces the ones formed by the non-lateral liquid⁵, as can be observed in (9).

- (9) a. florzinha [ˈflorˈziɲa] – 3:1:26 – *little flower*
 b. outra [ˈota] – 3:1:26 – *other*
 c. brincar [bĩnˈka] – 3:2:15 – *to play*

Considering the role of syllabic structure in segmental acquisition, it is expected that the consonant clusters formed by the lateral liquid and by the non-lateral liquid were acquired together. This does not happen simply because the acquisition of /r/ in the system of S1 and S2 occurs after the acquisition of /l/, at 3:2, in all the syllabic constituents. This can be easily explained by Optimality Theory, as shown in (10).

(10) a.

/ˈblu.za/	Not Complex Onset	MAX I/O	Ident I/O	Markedness
☞ /ˈbu.za/		*		*
/ˈblu.za/	*!			*
/ˈbru.ʃa/				
☞ /ˈbu.ʃa/		*		*
/ˈbru.ʃa/	*!			*
/ˈblu.ʃa/	*!			*

b.

/ˈblu.za/	[*[+aproximante] [-vocóide]]	MAX I/O	Ident I/O	Markedness	Not Complex
/ˈbu.za/		*!		*	
☞				*	*
/ˈbru.ʃa/					
☞		*		*	
/ˈbru.ʃa/	*!			*	*
☞			*	*	*

⁵ The reduction is not categorical if there are cases of substitution of the lateral liquid, metathesis, and epenthesis.

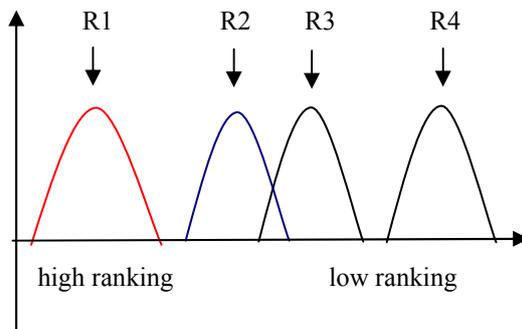
c.

/ˈblu.za/	MAX I/O	Ident I/O	[* [+aproximante] & * [+continuo] & [-vocóide]]	Markedness	Not Complex Onset
/ˈbu.za/	*!			*	
☞				*	*
/ˈbru.ʃa/					
/ˈbu.ʃa/	*!			*	
☞			*	*	*
/ˈblu.ʃa/		*!		*	*

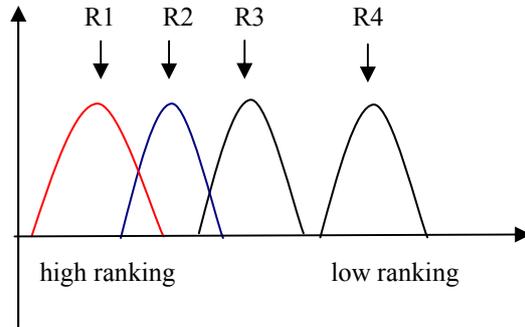
The tableaux in (10) express the three stages evidenced by S1 in the acquisition of onsets formed by liquids: (i) reduction of consonant clusters; (ii) production of onsets with lateral liquids, reduction of clusters with non-lateral liquid and substitution of liquids; (iii) acquisition of onsets formed by non-lateral liquids.

It is interesting to notice that, according to a probabilistic grammar (Boersma and Hayes, 2001), the floating ranking between Max I/O and Ident I/O shows why the second acquisition stage does not postulates just the violation of Ident I/O, with the substitution of liquids. As can be observed in (11), the gradual reordering among faithfulness constraints and the conjoined constraint [* [+aproximant] & * [+continuous] & [-vocoid]]_(seg) relates to segmental features.

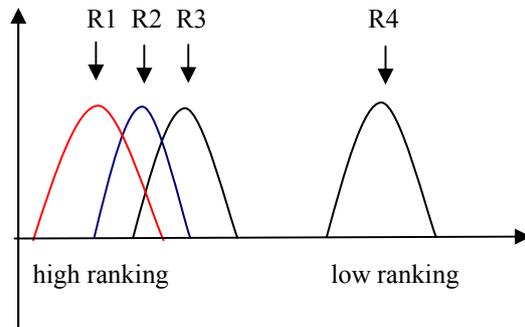
(11)a.



b.



c.



- R1 = [*[+aproximant] & *[+continuous] & *[-vocoid]]_(seg)
- R2 = Max I/O
- R3 = Ident I/O
- R4 = Markedness

In (11a), one can see a fragment of the grammar between 2:8:16 and 2:9:16, with the use of reduction strategies of consonant cluster and substitutions of non-lateral liquids. But (11b) shows the increase in the rate of substitutions made at the age of 2:10:17 and 3:0:21, and the beginning of the production of some consonant clusters formed by the non-lateral liquid, since the conjoined constraint is, in this stage, sharing stratum with Max I/O and Ident I/O.

The re-orderings seen in (11) therefore express the acquisition of consonant clusters driven by the segmental acquisition of a non-lateral liquid. It must be highlighted that the tableaux in (10) present the constraint NotComplex(onset), so the constituent has a role in the phonological acquisition of S1, as an integrant part of the syllable, but its connection with the emergence and acquisition of non-lateral liquids is not observed. This depends only on the demotion of the conjoined constraint [*[+aproximant] & *[+continuous] & [-vocoid]]_(seg).

In relation to the coronal fricative, the emergence of this segment occurs simultaneously in onset and coda position in the data of S1, at 1:6. The same can be observed for the development that takes place at around 3:2. Just as in case of the emergence of the non-lateral liquid, the fricative emerges first in initial coda and later on in medial coda position. Acquisition data of the fricative by S1 are displayed in (12).

- (12) a. sapato [pa'pa] – 1:1:22 - *shoe*
 b. sentar [te'ta] – 1:4:22 – *to sit down*
 c. pescoço – [pi'ko] – 1:5:7 - *neck*
 d. sabe [ˈsabi] ~ [ˈʃabi] – 1:6:17 – *she knows*
 e. sentar [sɛn'ta] – 1:6:17 – *to sit down*
 f. dois [ˈdojs] – 1:6:17 – *two*
 g. seis [ˈsejs] – 1:6:17 - *six*
 h. dois [ˈdojʃ] – 1:8:12 - *two*
 i. você [bo'se] – 1:9:9 - *you*
 j. sete [ˈʃɛtʃi] – 1:9:9 - *seven*
 k. maçã [ma'sã] – 2:3:17 - *apple*
 l. sentar [ʃin'ta] – 2:3:17 – *to sit down*
 m. três [ˈtlejs] – 2:3:17 - *three*

Just like the nasal and the non-lateral liquid, the coronal fricative is acquired simultaneously in all the syllabic positions by S1.

It is interesting to observe that the difference between the final coda and medial coda occurs just in the emergence of the consonantal segments in the data of S1, as the acquisition occurs simultaneously in the two types of coda.

The emergence and acquisition of /l/ presents, in opposition to the other consonants displayed in table 1, a distinguished behavior, as this segment is first acquired in final coda and medial onset, at around 1:8, and only at 2:2 in complex onsets. The acquisition of /l/, consequently, as opposed to the other consonants, seems to show the presence of the syllabic constituent of a complex onset, given that, although acquired, the lateral liquid is not produced in all syllabic positions.

However, it should be pointed out that the production possibility index of consonant clusters formed by the lateral liquid is extremely low in Portuguese, i.e., the frequency of this structure is very low in the language. See the data in table 2.

Age	Onset – /l/		Onset - /r/		Substit. by /l/
	Pos	Oc	Pos	Oc	
1:7:1	0	0	7	0	
1:7:15	1	0	10	0	
1:7:28	1	0	7	0	
1:8:12	1	0	7	0	
1:8:27	1	0	7	0	
1:9:9	0	0	16	0	
2:1:27	3	0	21	0	
2:2:19	1	0	13	0	
2:3:17	1	1	37	0	1
2:5:24	0	0	32	0	
2:7:12	0	0	33	1	
2:8:16	2	2	26	0	2
2:9:16	0	0	32	0	13
2:10:17	1	1	47	3	16
3:0:21	2	1	50	8	27
3:1:20	0	0	39	12	6
3:2:28	1	1	41	39	3
3:3:27	0	0	24	19	
3:4:27	1	1	28	26	
3:5:28	0	0	37	31	
3:6:28	0	0	18	13	
3:8:14	2	2	21	18	
3:9:13	2	2	25	25	

Table 2 - Consonant clusters formed by the lateral liquid

As can be seen in table 2, the production possibilities of complex onsets formed by the lateral liquid are extremely reduced if compared to the possibilities of the complex onset formed by the non-lateral liquid⁶. The frequency of this sequence in the speech of S1 is in accordance with the frequency rates of the segments /l/ and /r/ in onset position, referred to as in Albano (2001): 1,28 and 2,53, respectively, for adult data.

Table 3 shows the occurrence of the branching onset formed by the lateral in Brazilian Portuguese.

⁶ The same was observed by Ribas (2002) based on cross-sectional data of 134 children aged 2:0 and 5:3.

Age	/l/ - %		
	Init	Med	Final
1:1:22			0
1:2:13 - 1:2:27		0	50
1:3:10 - 1:3:34	0		0
1:4:09 - 1:4:22	75	33,3	
1:5:07 - 1:5:20	83,3	33,2	0
1:6:03 - 1:6:17	75	0	60
1:7:01 - 1:7:15	66,6	66,6	23
1:8:12 - 1:8:27	81,2	72,7	90
1:9:09	100	100	50
2:1:27	80	100	92,3
2:2:19	100	100	26,6
2:3:17	100	87,5	100
2:5:24	100	100	100
2:7:12	100	100	100

Table 3 – Correct productions of lateral liquids

Although it is not produced in complex onsets, the lateral liquid can already be considered acquired at 1:8:12, as the correct production percentage goes beyond the predicted 80%. So, once more, just like in the acquisition of the other segments analyzed, the syllabic structure does not drive the segmental acquisition in the analyzed data.

The absence of asymmetries in the acquisition of the coronal nasal, the coronal fricative, and the non-lateral liquid may indicate that:

(i) glides are positioned in complex nucleus in Portuguese, according to Bonilha (2000). So, the absence of asymmetries in the acquisition of the coronal nasal and the non-lateral liquid seems to show, according to Freitas (1997), that nasals also occupy the position of syllabic nucleus in Portuguese.

(ii) glides are positioned in syllabic codas in Brazilian Portuguese, being acquired too early due to the presence of the sonority scale Mezzomo (2004). The segments in the coda emerge according to the sonority between the nucleus and coda, that is, the smaller the sonority distance, the earlier the acquisition. Because of this, the coronal nasal is acquired at the same time in onset and coda. Consequently the coronal fricative and the non-lateral liquid will not evidence asymmetries between onset and coda either.

(iii) glides are positioned in complex nucleus in Portuguese, and the data analyzed show that the syllabic structure acquisition is driven by segmental acquisition.

In order to validate the hypothesis in (i), the emergence of asymmetries in the acquisition of the coronal fricative is expected, as observed in the data analysis of the European Portuguese. The acquisition of the coronal fricative by S1 in simple onset and coda occurs, however, in the same age group. Hence, the similar behavior

of the three consonants /n/, /r/ and /s/ does not allow the hypothesis in (i) to be considered possible.

The hypothesis in (ii) would be valid only for the absence of asymmetries between onset and coda, but not between a simple onset and a complex onset. The data of S1 confirm this fact, as the acquisition of the lateral liquid evidences the asymmetries between simple onset and complex onset. However, the acquisition of the lateral liquid at 2:2 in the complex onset, six months after the acquisition of the lateral in simple onsets, does not indicate, as shown by Freitas (1997), the role of complex onset in the segmental acquisition of Brazilian Portuguese, as the liquid is already acquired at 1:8.

The problem of the hypothesis in (ii) is that it is not able to explain why the falling diphthongs formed by low and mid-low vowels are acquired so early by S1, as observed by Bonilha (2000, 2004). If the smallest sonority distance between the nucleus and glides is what guarantees that the diphthongs be acquired in the coda before the nasal, fricative and non-lateral liquid, how can we explain that sequences such as [aj] and [aw] are acquired before sequences like [ej] and [iw]? The highest sonority distance between low and mid-low vowels and glides would favor the early acquisition of [ej] and [iw].

According to Albano (2001), *the inherent shortening of the semi vowels* characterizes these segments as constituents of the syllabic margins, but not specifically as codas, which makes it possible to classify them as a second element of a complex nucleus.

Bonilha (2000), applying a cross-sectional data analysis, observed that the falling diphthongs are produced in a precise way from the age of 1:3:24, with the production of the diphthong [aw]. The data of S1 prove the early acquisition of the diphthongs formed by the low vowel, [aw] and [aj]. It was also observed by the author that the diphthong formed by the mid-low vowel, [ɛw], is produced in a correct way in all tokens. Concerning the diphthongs formed by the mid-high vowels, [oj] and [ej], they not only emerge at 1:5:20 in the speech of S1, but also present instability in the production.

Such instability is observed mainly in the production of [ej], whose percentages vary from 33,3% to 100%. It should be noted that the diphthong [ej], in the analysis of Bonilha (2000), did not reach a satisfactory production percentage until the age of 2:5.

Following the author, it is possible to postulate the role of the constraints displayed in (13), to explain the acquisition of the falling diphthongs by S1.

- (13) NoSequence (nucleus) (+low...+high): a complex nucleus must not present a vowel sequence [+low], [+high].
NoSequence (nucleus) (-low...+high): a complex nucleus must not present a vowel sequence [-low], [+high].
NotTwice (coronal): two coronal elements must not appear in sequence.

As was proposed by Bonilha (2000), in the first acquisition stage the demotion takes place of NoSequence (nucleus) (+low...+high) below the faithfulness

constraints, which, in agreement with the gradual acquisition algorithm, are promoted in the hierarchy. Therefore, such movement allows the production of diphthongs formed by low and mid-low vowels

In the second stage, there is the demotion of NoSequence (nucleus) (-low...+high), making it possible for [ej], [oj], [ew] and [iw] to emerge. The instability in the production of [ej] is shown by the role of NotTwice (coronal), that shares stratum with the faithfulness constraints. In (14) we present the constraint hierarchies that allow the acquisition of the falling diphthongs by S1.

- (14) a. 1st acquisition stage
 H1 = NotTwice(coronal), NoSequence (nucleus) (-low...+high) >> Max I/O >> NotComplex(nucleus), NoSequence (nucleus) (+low...+high)
- b. 2nd acquisition stage
 H2 = NotTwice(coronal) >> Max I/O >> NotComplex(nucleus), NoSequence (nucleus) (+low...+high), NoSequence (nucleus) (-low...+high)
- c. 3rd acquisition stage
 H3 = Max I/O >> NotComplex(nucleus), NoSequence (nucleus) (+low...+alto), NoSequence (nucleus) (-low...+high), NotTwice(coronal)

The data of S1 show not only the acquisition of the diphthongs at a stage before the acquisition of codas, but also express the early emergence, even with unstable values, of the pattern CVVC.

Considering the early acquisition of the final coda /l/, we postulate that in final position of the word /l/ is interpreted by S1 as a constituent of the complex nucleus.

6. Conclusion

According to Fikkert (1994), Miranda (1996) and Freitas (1997), among others, the acquisition of a segment depends on the setting of parameters related to the syllabic patterns of a language. The age of emergence and acquisition of the segments in the different syllabic constituents, presented in Table 1, show, however, the absence of the role of the syllabic constituents in the segmental acquisition of S1. What seems to happen, in fact, is that the acquisition of specific syllabic constituents depends on the demotion of feature constraints.

The fact that the role of the syllabic structure in the segmental acquisition of S1 was not evidenced does not mean that constituents such as Onset, Coda and Rime should not be considered in the phonological analyses of Portuguese. What is defended here is not the absence of syllable structure in the phonological acquisition of S1 -the syllable linguistic unit is present since the beginning in production (Fikkert, 1994; Freitas (1997)), assuring the emergence of the different syllabic patterns of Portuguese - but the lack of evidence on the role played by syllabic structure in the segmental acquisition process of this child. In this sense, the syllabic representation proposed by Kahn (1976) seems sufficient to explain the data.

Having as a basis the data of a longitudinal subject, the results of the present study seem to show that analyses concerning the interaction between segmental acquisition and syllabic structure based on cross-sectional data must be rethought.

References

- Albano, E. (2001) *Os Gestos e Suas Bordas – Esboço de Fonologia Acústico-Articulatória do Português Brasileiro*. Campinas: Mercado de Letras/ALB.
- Bisol, L. (1999) ‘A sílaba e seus constituintes’, in: M. H. de M. Neves (ed.). *Gramática do Português Falad., volume VII*. Campinas: Editora da Unicamp.
- Boersma, P. and B. Hayes (2001) *Empirical Tests of the Gradual Learning Algorithm*. ROA.
- Bonilha, G. F. G. (2000) *Aquisição dos Ditongos Orais Decrescentes: Uma Análise à Luz da Teoria da Otimidade*. MA thesis, UCPel.
- Câmara Jr, M. J. (1977) *Estrutura da Língua Portuguesa*. Petrópolis: Vozes.
- Collischonn, G. (1997) *Análise Prosódica da Sílaba em Português*. Doctoral dissertation, Instituto de Letras e Artes, PUCRS.
- Fikkert, P. M. (1994) *On the Acquisition of Prosodic Structure*. Doctor dissertation, Holland Academic Graphics.
- Freitas, M. J. (1997) *Aquisição da Estrutura Silábica do Português Europeu*. Doctoral dissertation, University of Lisbon.
- Hooper, J. (1976) *An Introduction to Natural Generative Phonology*. New York: Academic Press.
- Kahn, D. (1976) *Syllable-Based Generalizations in English Phonology*. Cambridge MA: MIT Press.
- Lamprecht, R. R. (1990) *Perfil da Aquisição Normal da Fonologia do Português. Descrição Longitudinal de 12 Crianças: 2:9 a 5:5*. Doctoral dissertation, Instituto de Letras e Artes, PUCRS.
- Mezzomo, C. (2004) *Aquisição da Coda no Português Brasileiro: Uma Análise via Teoria de Princípios e Parâmetros*. Doctoral dissertation, Porto Alegre: PUCRS.
- Miranda, A. R. M. (1996) *A Aquisição do “r”: Uma Contribuição à Discussão sobre Seu Status Fonológico*. MA thesis – Instituto de Letras e Artes, PUCRS.
- Prince, A. and P. Smolensky (1993) *Optimality Theory - Constraint Interaction in Generative Grammar*. RuCCs Technical report 2.
- Ribas, L. (2002) *Aquisição do Onset Complexo*. MA thesis – Instituto de Letras e Artes, PUCRS.
- Selkirk, E. (1982) ‘The syllable’, in: H.V.D Hulst and D. Smith (eds.) *The Structure of Phonological Representations (part II)*. Dordrecht: Foris.
- Vennemann, T. (1974) ‘Words and syllables in Natural Generative Grammar’, in: A. Bruck (ed.) *Papers from the Parasession on Natural Phonology*. Chicago: CLS.

The L2 Acquisition of Spanish Focus

A case of incomplete and divergent grammars

Laura Domínguez
University of Southampton

1. Introduction

This paper explores the second language acquisition of Spanish word order with the aim of analysing the availability of optional forms in advanced non-native grammars. In contrast to English, Spanish word order is flexible and the elements of a sentence can appear in more than one configuration (e.g. SV, VS). This may appear to be a case of free alternation. However, such optionality is only apparent as each of the configurations is constrained by defined syntactic rules (depending on the type of verb) and pragmatic rules (depending on the type of information encoded in the sentence). Consequently, this phenomenon is ideal for testing hypotheses about optionality in endstate L2 grammars since two variations of the same structure are present in the input which in turn means that L2 learners need to figure out both the rules that constrain each of the forms and their context of use. In this paper we explore whether this ambiguity in the input may delay the acquisition of these forms until an advanced stage such that alternate forms will still be present in the grammar of near-native speakers of Spanish. Our hypothesis builds upon the assumption that in developing grammars the emergence and persistence of optional forms are highly dependent on the level of systematisation and robustness of the input subjects receive. Consequently, we predict that in the case of word order variation in Spanish advanced learners would go through a persistent stage of optionality since the evidence they receive is not systematic enough to make proper generalisations that would allow them to map each of the forms with its context of use.

In this study we also focus on the source of such optionality arguing that certain errors found in advanced non-native grammars cannot be sufficiently accounted for as simple transfers from the learner's L1. In this respect we also investigate whether optionality in advanced L2 grammars is only derived from features available in the L1 or whether subjects will allow grammars that are unlike their L1 but also divergent from the L2.

2. Optionality in advanced L2 grammars

It has been widely observed that non-native optionality, i.e. where two competing grammars exist in the mental representation of L2 learners, is a common feature of developing grammars even at advanced proficiency levels (White 1991, 1992; Eubank 1994; Sorace 1993, 1999, 2000; Prévost and White 2000). The standard view on optionality is that the optional use of a particular form reveals that L2 learners may be considering grammatical representations that are not exactly target-like but nevertheless may still be congruent with their own interlanguage. The

question that we focus on is why optionality is still permitted at very advanced stages of acquisition, exploring whether the evidence found in the input can affect the stage of optionality in developing grammars. It has been observed that the nature and systematisation of the input is a factor allowing optional forms to linger in advanced L2 grammars. More precisely, Papp (2000) argues that L2 learners need to figure out the status of a rule in the second language and in doing so they need to establish whether that rule is categorical, optional or quasi-optional. The more ambiguous the target language data is, the longer it takes for the learners to learn the rule facilitating lingering stages of optionality. In the case of Spanish word order the input is highly ambiguous since learners are exposed to pairs of structures which are almost identical and which may seem interchangeable in the same contexts:

- (1) a. Ha comprado el libro Juan
has bought the book Juan
“Juan has bought the book”
- b. Juan ha comprado el libro

In example (1) the same elements appear in two different configurations and bear different informational status. Specifically, sentence (1a), with a postverbal subject, necessarily implies that the subject is the only new information (i.e. the focus) of the sentence, whereas sentence (1b) where the canonical word order is preserved is necessarily an all-focus sentence. Consequently, these two examples are felicitous in two different contexts – such as in answer to the questions ‘who has bought the book?’ and ‘what happened?’ respectively.

Optionality is a widespread phenomenon in both first and second language developing grammars and is still the subject of much debate in the field of acquisition. Most of the studies in the optionality literature have focused on the emergence of functional categories. For instance, Lardiere (1998, 2006) analyses the case of a Chinese learner of English showing how in the subject’s oral data the past tense morpheme in English is optionally expressed even at an advanced stage where other similar structures, (like definite articles), and relatively more complex structures, (like I-to-C movement), are correctly acquired. In first language acquisition, Poeppel and Wexler (1993) and also Wexler (1994, 1998) have shown how children use both inflected verbs and root infinitives during a stage at around two years of age. Although optional root infinitives should not be allowed in the child’s grammar it is also evident that the nature of this optional stage is not random and both options, in this case the inflected and non inflected verb forms, are legitimate in the child’s grammar at this early stage. Unlike second language learners, though, children seem to go through this optional infinitive stage quite quickly which seems to support the claim that the nature and amount of input plays a crucial role in acquisition even at later stages where interlanguage grammars are stable.

From a theoretical point of view the availability of optional forms in both native and non-native grammars is problematic since in a framework such as Minimalism (Chomsky 1995) there should only be one output for each single derivation. Recent studies on optionality have shown, however, that optionality in L1 grammars is in fact only apparent since the distribution of two or more optional forms may be in

fact be constrained by different discourse conditions (Parodi and Tsimpli 2005) implying that even if a system allows more than one form for the same derivation these forms may not be exactly identical. Robertson's (2000) study of Chinese's L2 use of English articles reached a similar conclusion. He argues that even if two possible variations of the structure may coexist in learner grammars (in this case the overt realisation of articles) the contexts in which the different forms are used are not exactly identical, which means that optionality may only be apparent (and learner's use of two forms is not arbitrary), and that there may be more behind the availability of optional forms than an incorrect representation of the target grammar.

The second issue that we are concerned with in this study is the source of optional forms in second language grammars. It is generally assumed that learners may revert to their native language when they find difficulty in inducing the rules of the target grammar (Sorace 1993; Papp 2000). Consequently, the first language can be the starting point from which learners build their L2 representations but it continues to influence the acquisition process even at an advanced stage. More specifically, if second language learners find the input too vague to build their knowledge of the target language upon it, they may opt for reverting to the setting available in their first language and use it to build representations of the target grammars. The acquisition of focus in Spanish is a good testing ground because focus alters the canonical word order and allows for the same elements to be reorganised in different orders (Zubizarreta 1998; Domínguez 2004) in what apparently constitutes a case of optional word order. In this study we analyse the acquisition of pairs of structures which are apparently identical except for the ordering of their elements (see example 1). Crucially, only one of the two structures that is analysed in our study is available in the L1 (i.e. the non inverted option (1b)). Based on the input that the subjects in our study are exposed to we observe that L2 learners have enough evidence to assume that two optional structures exist in the target language until they learn the discourse-pragmatic constraints of each of the available forms. Also, the linguistic evidence from which L2 learners create grammatical assumptions can be quite ambiguous. Given such an obvious lack of robustness in the input, the learning task is made considerably more difficult and presumably learners will face longer periods of grammatical indeterminacy even at advanced levels of proficiency. It may be possible that during this stage of indeterminacy learners revert to their L1 favouring the one option available in Spanish that is available in English as well. In line with these assumptions, previous studies on the acquisition of Spanish word order have shown that advanced second language learners encounter problems acquiring the pragmatic conditions that constrain word order alterations (Ocampo 1990; Hertel 2003; De Miguel 1993; Lozano 2006).

In this study we follow Sorace's (1993) three types of representations allowed in near-native grammars: convergent, divergent and incomplete L2 end-states. For Sorace, if non-native representations are completely native-like they are *convergent*. If, on the other hand, not all the properties of the target language are observed in the near-native grammar then L2 representations are *incomplete* or *indeterminate*.

Finally, *divergent* representations are those non-native representations which are consistently different from native properties and influenced by the L1 grammar.¹

3. Focus and word order in Spanish

In Spanish, unlike English, word order is quite flexible and any constituent may appear in different positions in the sentence. Such flexibility can be accounted for by focus-related operations which are motivated by prosodic conditions (Zubizarreta 1998). Specifically, assuming that focused elements must receive stress, which is assigned by a stress assignment rule (i.e. the Nuclear Stress Rule (NSR)) to the most embedded constituent (Chomsky and Halle, 1968; Cinque 1993) the focus is expected to appear in sentence-final position even if canonical word order is to be altered. This is illustrated in examples (2b) and (3b) where the subject must appear postverbally and in final position because it is in focus:

- (2) a. What happened?
 b. [_F La estudiante visitó al profesor] SVO
 the student visited to-the professor
 c. Who visited the professor?
 d. Visitó al profesor [_F la **estudiante**] VOS
- (3) a. [_F Juan ha llegado] SV
 Juan has arrived
 b. Ha llegado [_F **Juan**] VS

However, this focus-stress alignment rule only applies if the focus is informational. If the focus is contrastive it can receive stress in situ and the SV(O) canonical order is preserved:

- (4) [_F La **estudiante**] visitó al profesor, (no el decano) SVO
 the student visited to-the professor, not the dean

In this sentence the focused subject cannot be associated with the main stress of the sentence, which falls in final position via the NSR. Therefore, focused elements are not always required to appear in sentence-final position in Spanish. The fact that two types of focus constructions exist which allow the subject to appear in different positions may be interpreted as a case of optionality by L2 learners of Spanish. However, the availability of these orders is constrained by pragmatic principles, (i.e. focus). Consequently, in order to properly learn the rules constraining word order in Spanish subjects must learn not only when and how to apply a focus rule in order to correctly map each structure with its context, but also that an apparent optional rule is in fact not optional.

¹ Papp (2000) notes that divergent representations are ambiguous as they allow for optional rules in the L2 to be either differentiated or rejected.

Another structure used in Spanish to mark focus is clitic left-dislocations (Cinque 1990; Zubizarreta 1998). In these structures the focused element appears in final position by virtue of dislocating the given information out of the core clause. A coindexed resumptive clitic pronoun must appear in this construction as illustrated in the following example:

- (5) a. Who has brought the cake?
- b. El pastel, lo ha traído Sara
 the cake, it has brought Sarah
 “Sarah has brought the cake”

Clitic left-dislocations, unlike other focus-related operations, always require the subject to appear postverbally. Consequently, example (6) with a preverbal subject is ungrammatical:

- (6) *El pastel, Sara lo ha traído
 the cake, Sarah it has brought

Therefore clitic left-dislocations, for which there is unambiguous evidence in the input, are relevant in our study because if lack of systematisation in the input is the source of optionality, learners should find learning these forms less problematic than learning structures where alternative word orders exist as in example (1).

4. Experiments

An experiment involving 21 native Spanish speakers living in Spain and 28 English speakers learning Spanish in a UK university was carried out to investigate whether advanced L2 learners of Spanish have acquired the pragmatic restrictions of focus and are able to change word order when required. An advantage over previous research is that both cases of information focus (which always forces movement) and contrastive focus (which is not subject to word order alterations) were tested. Proficiency levels were determined by a cloze test and subjects, all in the final year of a language degree, were divided into three different groups: advanced (scores between 50 and 35), intermediate (scores between 34 and 24) and low (scores between 23 and 0). The data was collected using two different tasks: a Contextualised Production Task (CPT) and Acceptability Judgement Task (AJT). Both tests included 20 questions (including 6 distractors) that required answers with SVO, VOS, VS, SV constructions, sentences with clitic left dislocations (O#Cl-V-S) where the subject is always focused and in final position, and sentences with in situ contrastive focus. The AJT included 2 questions with SVO/VOS orders, 2 questions with CLLDS, 2 questions with contrastive focus and 8 questions with SV/VS contrasts (including four unergative verbs, half of which had narrow focus on the subject, and four unaccusative verbs, half of which had narrow focus on the subject as well). In the CPT, subjects were presented with a context and were asked to provide an appropriate answer using the information provided. Subjects were expected to use verb-subject inversion in cases of information focus, but focus in situ in cases of contrastive focus. All inverted structures were cases of information

focus in the tests. Similarly, the AJT presented two possible answers to a question based on information provided by a brief context. The only difference between the pair of sentences provided was that there was a variation in the ordering of the elements which in turn reflected the different informational status of each of the structures. Since both sentences are grammatically correct in Spanish, even though only one of them is felicitous in each particular context, subjects were asked to rate their acceptability of each of the sentences and did not have to provide absolute grammatical judgements. Next is an example of one of the questions used in this task:

- (7) Last night there was a party in Marta's flat with many foreign students, but you couldn't go. When you see Marta today you ask her "Who danced at the party?"

What would Marta say?

Bailaron las chicas italianas -2 -1 0 +1 +2
 danced the girls Italian
 "The Italian girls danced"

Las chicas italianas bailaron -2 -1 0 +1 +2

Taking into consideration the properties of word order variation in Spanish the hypotheses considered in our study are the following:

1. If learners are not able to restructure the conflicting information they get from the input they will not produce/accept sentences with non-SVO word orders in the right context (their grammars will be divergent).
2. If learners are able to restructure the conflicting information they get from the input they will produce/accept sentences with non-SVO word orders (their grammars will be convergent) beyond transfer effects.

5. Results and discussion

5.1 Contextualised Production Task

The results of the Contextualised Production Task show that the non-native group behaves quite homogeneously, as there are no significant differences in their percentage of use of inversions. Overall, this group prefers to use a non-inverted structure in all the questions they provided, which may imply that L2 advanced learners' use of all the different word orders allowed in Spanish is rare. However, one important finding is that they seem to distinguish between contrastive and non-contrastive focus by using a cleft in cases where focus was contrastive as in the following example:

- (8) Es Juan quien compró el periódico
 "It is Juan who bought the newspaper"

Overall, 42% of all the instances with contrastive focus were of the type “It is X who”. Clefts were used instead of inversion (VS) 65% of the time in those cases where the subject was narrowly focused. Another relevant result is that even though subjects show proficient use of clitic-left dislocations in those instances where these structures are appropriate they consistently fail to invert the subject producing *Cl-S-V instead of Cl-V-S. In fact only one instance of a clitic left dislocation was produced with the correct order by one of the most advanced subjects. This result was found across the three proficiency groups.

5.2 Acceptability Judgement Task

The data collected by the Acceptability Judgement Task shows that the non-native group prefers sentences with SVO orders over VOS orders even in contexts where the subject is narrowly focused and should appear postverbally. Interestingly, the control group did not accept VOS as much as expected. These results are illustrated in the following graph which shows the acceptability of SVO and VOS orders by natives and all of the non-native speakers as a group in those structures in which the subject is focussed:

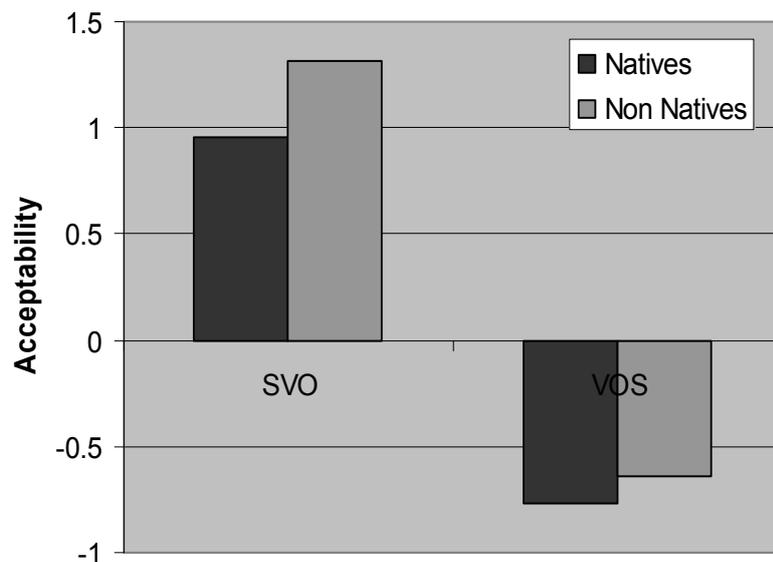


Fig. 1: Acceptability of SVO and VOS orders by native and non-native speakers.

The following figure illustrates the acceptability of the same two structures by each of the proficiency groups (advanced, intermediate and low). As expected, the subjects with lower proficiencies reject the inverted structure and give the higher acceptability scores to the structure which is allowed in their L1. The graph also shows that the advanced group is the non-native group that rejects the VOS the least but, unlike the other two non-native groups, does not give the higher scores to the

SVO structures. In this respect the advanced group is the group which does not seem to favour one of the options over the other, unlike the other two groups which clearly prefer the non-inverted option over the inverted one. This seems to imply a higher level of indeterminacy in the responses of the more advanced group:

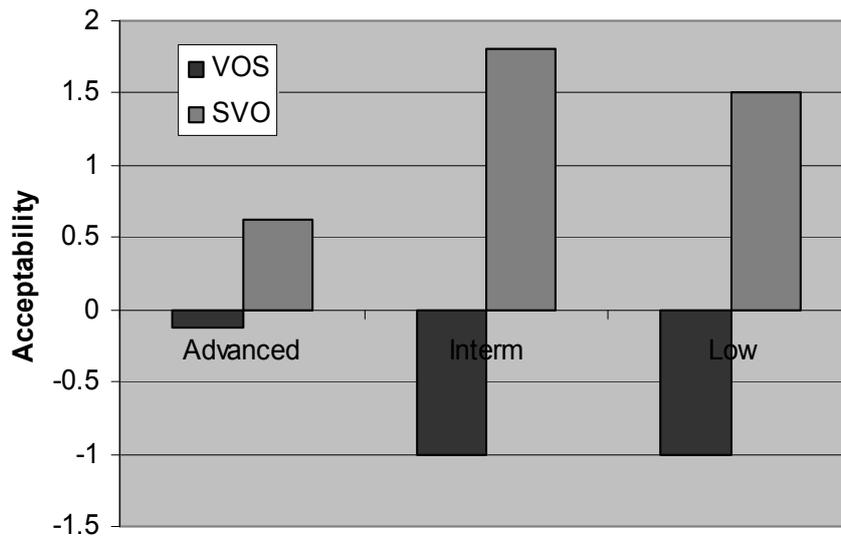


Fig. 2: Acceptability of SVO and SVO structures by three different proficiency groups and native controls.

In the non-native data, differences between the inverted and non inverted options with intransitive verbs (e.g. VS and SV) are only significant in half of the questions. This supports previous research finding that acquiring word order is problematic even at a very advanced stage of acquisition (Leonini 2003; Hopp 2005). However, the analysis of the data in Figure 4 by proficiency groups reveals that the advanced group behaves nativelike (i.e. their responses are statistically not significant) in their acceptability of SV and VS structures in which the focus is on the subject. The advanced group systematically accepts VS in the right contexts and rejects the non-inverted option appropriately, whereas the intermediate and low groups accept both options regardless of the context:

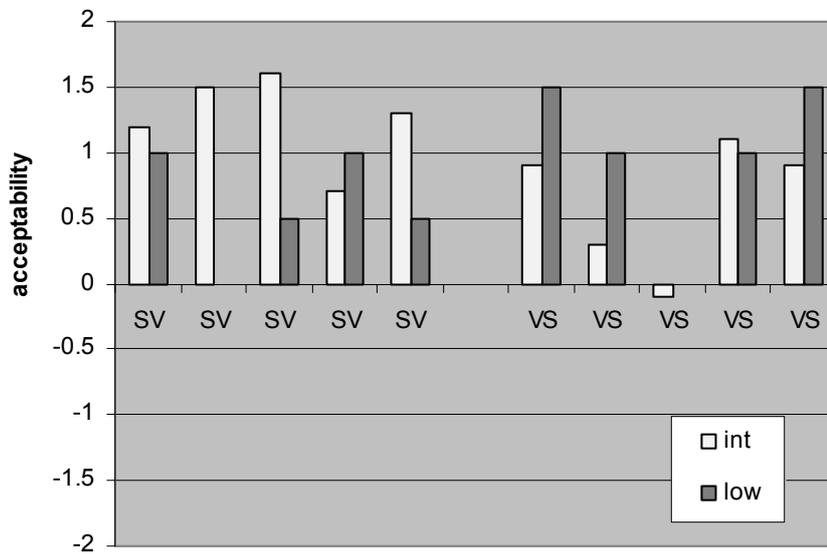


Fig. 3: Acceptability of SV and VS structures by question and proficiency group (low and intermediate).

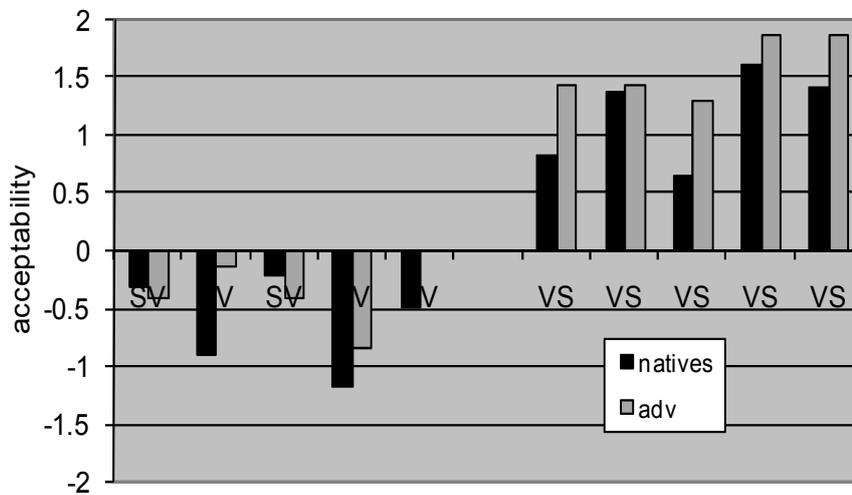


Fig. 4: Acceptability of SV and VS structures by question and proficiency group (advanced and native).

In those structures where clitic left dislocations were tested only the advanced group prefer the option with inversion (i.e. Cl-V-S vs *Cl-S-V) and reject the non-inverted structures, whereas the other two groups accept both (see Figure 5). These

results show differences between the comprehension and production of these forms by the advanced learners:

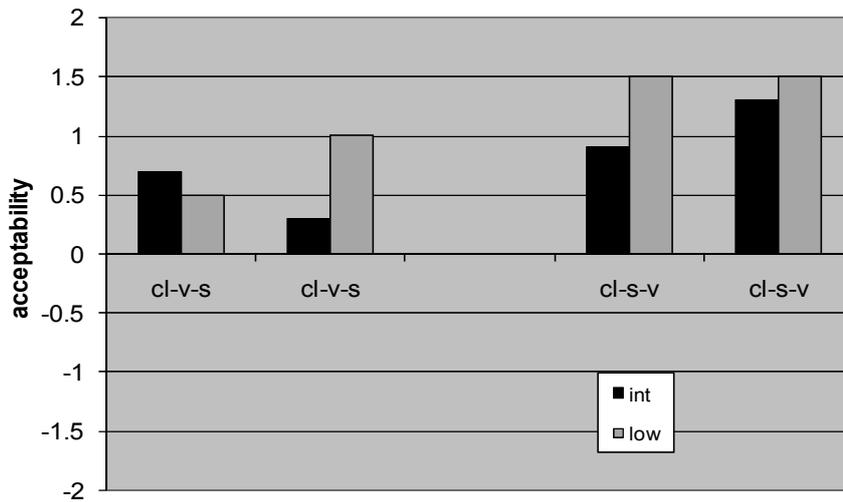


Fig. 5: Acceptability of structures with clitic left dislocations by question and proficiency group (intermediate and low).

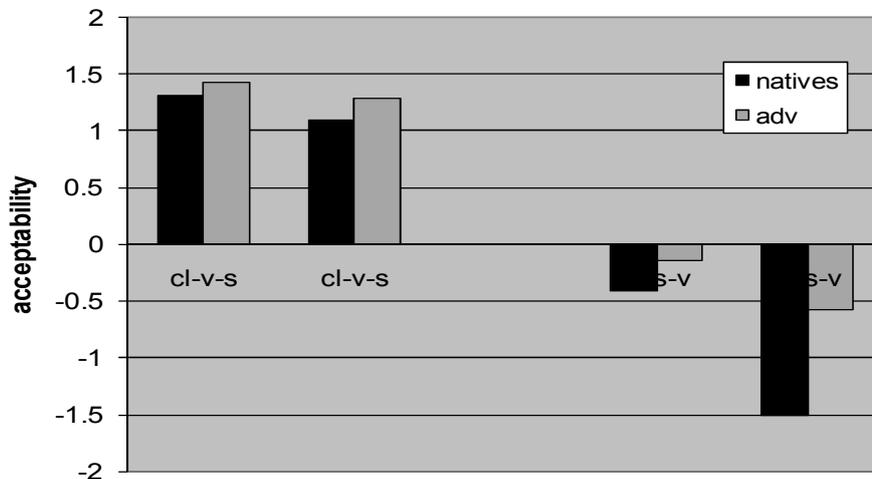


Fig. 6: Acceptability of structures with clitic left dislocations by question and proficiency group (advanced and native).

An interesting result is that unexpectedly the native group does not always accept or produce the inverted option, and this is particularly evident with VOS sentences with a transitive verb and an object, which by being part of the given

information, remains in the sentence. This may be due to the fact that this structure is a marked configuration in Spanish, i.e. a more natural answer to a question where VOS is felicitous would not include all the old information. For instance, even though examples (9a) and (9b) are grammatically correct and appropriate answers to a question such as ‘Who broke the glass?’, sentence (9a) is preferred over (9b) which contains the focus and the old information:

- (9) a. Rompió el vaso [_F **Juan**]
 broke the glass Juan
 b. [_F **Juan**]

This difference may be the reason why speakers react unexpectedly and offer a negative rating to sentences like (9a) with VOS orders.

The results clearly show that the acquisition of focus and its effects on word order is problematic even at advanced levels of acquisition. However, the most advanced learners show patterns of acceptability that show that they are moving away from the constraints of their first language and have assimilated some of the new rules of the target language as they accept inverted forms over the non-inverted in most of the questions. This shows that this grammatical area is not fossilised and restructuring of the interlanguage to accommodate the new rules is taking place.

The results also show that subjects across the three proficiency groups are sensitive to the properties of the two focus types (contrastive and not contrastive) as the use of clefts is used as a strategy to mark that the focus is contrastive. The most interesting finding shows mixed behaviour amongst the most advanced group in several of the answers; for instance they produce sentences with dislocated topics with obligatory clitics (a construction which does not exist in their L1) but without the verb-subject inversion, which is observed in the native data. This suggests that even though subjects have not achieved a stable L2 grammar they are able to form constructions that are neither represented in their L1 or the target grammar. In this respect, the non-native representations can be described as being both incomplete and divergent, in the sense of Sorace (1993).

6. Conclusions

The analysis of the data has shown that the non-native speakers do not use inversion like the natives but seem sensitive to semantic restrictions related to focus types and allow subject inversion in some of the contexts controlled in our experiments.

Only the advanced group shows native-like behaviour in most of the structures tested and in their behaviour towards preferring the inverted option. Some significant differences between the learner groups are observed. In particular, the intermediate and low groups accept both options, whereas the advanced group consistently accepted only the inverted option, which is not available in the L1. This shows that optionality declines correctly in the advanced group. Interestingly, even at a stage where subjects allow for an incorrect option, learners with the lowest proficiency allow an option that is not transferred from their L1 (in the case of clitic

left dislocations) and use it productively to show differences in the information encoded in a sentence.

The results show that learner grammars may allow two options in their representation of word order, but this can be interpreted as a phase where they are restructuring knowledge, as they learn the mapping between each structure and the contexts in which they can be used.

Finally, our predictions with regard to the effects that unambiguous input has on the lingering of optional forms in advanced grammars has been attested by the data. Word order variation is a problematic area for L2 learners of Spanish who need to learn the mapping of each of the forms available with the contexts they can be used in, but as the advanced group shows, nativelike proficiency in this area can be attained.

References

- Chomsky, N. and M. Halle (1968) *The Sound Pattern of English*. Harper Row, New York.
- Chomsky, N. (1995) *The Minimalist Program*. Cambridge, MA: MIT Press.
- Cinque, G. (1990) *Types of A' dependences*. Cambridge, MA: MIT Press.
- Cinque, G. (1993) 'A null theory of phrase and compound stress'. *Linguistic Inquiry* 24, 239-298.
- De Miguel, E. (1993) 'Construcciones ergativas e inversión en la lengua y la inter-lengua españolas', in: J.M. Liceras (ed.) *La Lingüística y el Análisis de los Sistemas no Nativos*. Ottawa: Dovehouse, 178-195.
- Domínguez, L. (2004) *Mapping focus: the syntax and prosody of focus in Spanish*. Unpublished Doctoral Dissertation. Boston University.
- Eubank, L. (1994) 'Optionality and the initial state in L2 development', in: T. Hoekstra and B. Schwartz (eds.) *Language Acquisition Studies in Generative Grammar*. Amsterdam: John Benjamins.
- Hertel, T. (2003) 'Lexical and discourse factors in the second language acquisition of Spanish word order'. *Second Language Research* 19, 4, 273-304.
- Hopp, H. (2005) 'Constraining second language word order optionality: Scrambling in advanced English-German and Japanese-German interlanguage'. *Second Language Research* 21, 34-71
- Lardiere, D. (1998) 'Dissociating syntax from morphology in a divergent L2 end-state grammar'. *Second Language Research* 14, 359-375
- Lardiere, D. (2006) *Ultimate Attainment in Second Language Acquisition: A Case Study*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Leonini, C. (2003) 'V2 in Adult L2 German: analyzing the interlanguage grammars', in: M. van Koppen, J. Sio and M. de Vos (Eds.) *Proceeding of ConSole XI*, Universiteit Leiden.
- Lozano, C. (2006) 'Focus and split intransitivity: The acquisition of word order alternations in non-native Spanish'. *Second Language Research* 22, 1-43.
- Ocampo, F. (1990) 'The pragmatics of word order in constructions with a verb and a subject'. *Hispanic Linguistics* 4, 1 87-127.
- Papp, S. (2000) 'Stable and developmental optionality in native and non-native Hungarian grammars'. *Second. Language Research* 16(2), 173-200.
- Parodi, T. and I.-M. Tsimpli, (2005) 'Real and apparent optionality in L2 pronouns: finiteness and pronouns in null operator structures'. *Second Language Research* 21, 3, 250-285.

- Poeppl, D. and K. Wexler (1993) 'The full competence hypothesis of clause structure in early German'. *Language* 69, 1-33.
- Prévost, P. and L. White. (2000) 'Missing surface inflection or impairment in second language acquisition? Evidence from tense and agreement'. *Second Language Research* 16, 103-133.
- Robertson, D. (2000) 'Variability in the use of the English article system by Chinese learners of English'. *Second Language Research* 16, 135-172.
- Sorace, A. (1993) 'Incomplete and divergent representations of unaccusativity in nonnative grammars of Italian'. *Second Language Research* 9, 22-48
- Sorace, A. (1999) 'Initial states, end-states, and residual optionality in L2 acquisition', in: A. Greenhill, H. Littlefield and C. Tano (eds.) *Proceedings of the 23rd Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press, 666-674.
- Sorace, A. (2000) 'Syntactic optionality in non-native grammars'. *Second Language Research* 16, 93-102.
- Wexler, K. (1994) 'Optional infinitives, head movement and the economy of derivations in child grammar, in: D. Lightfoot and N. Hornstein (Eds.) *Verb movement*. Cambridge, MA: Cambridge University Press, 305-350.
- Wexler, K. (1998) 'Very early parameter setting and the unique checking constraint'. *Lingua* 106, 23-79.
- White, L. (1991) 'The verb-movement parameter in second language acquisition'. *Language Acquisition* 1, 337-360
- White, L. (1992) 'Long and short verb movement in second language acquisition'. *Canadian Journal of Linguistics* 37, 273-286.
- Zubizarreta, M. L. (1998) *Prosody, Focus, and Word Order*. Cambridge, MA: MIT Press.

On the Acquisition of Clitic *se* and Aspect in Spanish

Linda Escobar and Vicenç Torrens
Universidad Nacional de Educación a Distancia, Madrid

0. Abstract

The literature on acquisition of aspect has focussed on the perfectivity-imperfectivity distinction. In this paper we deal with this topic regarding the acquisition of aspectual clitic *se* which is crucially related to perfectivity structures. The data analysed are available on the Childes database. We show that very young children make omission errors of clitic *se*. We support the view that the acquisition of telicity is difficult in languages where telicity is marked by the combination of the properties of the verb and its object (Van Hout, 2003, and Hogdson 2003).

1. Introduction

Much research on children's acquisition of clitics has focused on the development of all the appropriate morpho-syntactic levels as a locus of learning and variation among different child languages. In this paper we attempt to examine the acquisition of aspectual clitic *se* in Spanish, because it presents an interesting linguistic challenge for children, given that it affects the way the aspectual distinctions are encoded in the morpho-syntactic levels of the Spanish clause structure.

In section 2, we present previous studies on the acquisition of clitics. In section 3, we discuss Spanish telicity and provide an analysis for clitic *se*. We show how research on language acquisition may contribute to shed some light on the debate on the acquisition of aspect. In section 4, we include the present study based on longitudinal studies on the production of aspectual *se*. Section 5 is intended as the discussion and the main conclusions of the paper.

2. Previous studies on the acquisition of clitics:

The acquisition of clitics in Romance languages has been studied in the last twenty years (Guasti 1993-1994, Schaeffer 2000, Wexler et al. 2004, Tedeschi 2006) and is still subject of current debate since it derives many implications with respect to the acquisition of structural agreement in the Romance clause.

In all the studies cited before, omissions of accusative clitics are attested in languages with past participle agreement. These omissions are characterized by optionality. The way optionality is accounted for is however different. According to Schaeffer (2000), the optional use of an accusative clitic by young Italian children doesn't argue against the "Full Clause Hypothesis" which claims that functional categories are present from the beginning in child grammar. The underlying problem of the optional use of accusative clitics is explained by claiming that children lack a

pragmatic principle which would allow them to interpret discourse relations the way adults do.

Tedeschi (2006, this volume) argues that optional clitic omission is not caused by pragmatic factors, but by a competition between discourse and syntactic requirements, due to economy constraints which would affect children's performance. Interestingly, her findings show that, apart from optional omission of clitics, much lower than in Shaeffer's (2000) study though, clitics sometimes agree with the subject (instead of agreeing with the object), indicating that "children could not always link the clitic to its antecedent, or that they inverted Agent and Patient in their answers" (Tedeschi 2006).

Lyczkowski (1999) demonstrated, on the basis of naturalistic data, that Spanish-speaking children rarely omit object clitics. The same result has been achieved experimentally by Wexler et al. (2004).

In our study below, we want to examine acquisition of clitic *se* to test whether this element is problematic for children, provided the fact that it is not an ordinary object clitic but an element highly related to aspectual perfectivity (cf. Rigau 1994, Sanz 1995, Zagona 1996), which seems to be somehow problematic for children according to Hodgson's (2003) recent production and comprehension studies.

3. Clitic *se* and Perfectivity

Perfect *have* in Spanish is classified as perfective. However, this does not mean it is associated with the achievement of a goal or that it can be seen as telic or resultative, as the strangeness of (1) and (2) indicates:

(1) Él ha caído
"He has fallen"

(2) Él ha ido
"He has gone"

Sentences (1) and (2), if uttered at all, are interpreted as an activity that was done without reaching a particular goal. Hence, the tense is past but there is present relevance. As for the notion of perfectivity or telicity, i.e. reaching a goal, it is not connected to *have* in (1) or (2), but is present in a separate functional category, as argued by Sanz (1996). We assume that *se* is inserted in this functional category in order to provide the sentences in (1) and (2) with the notion of telicity:

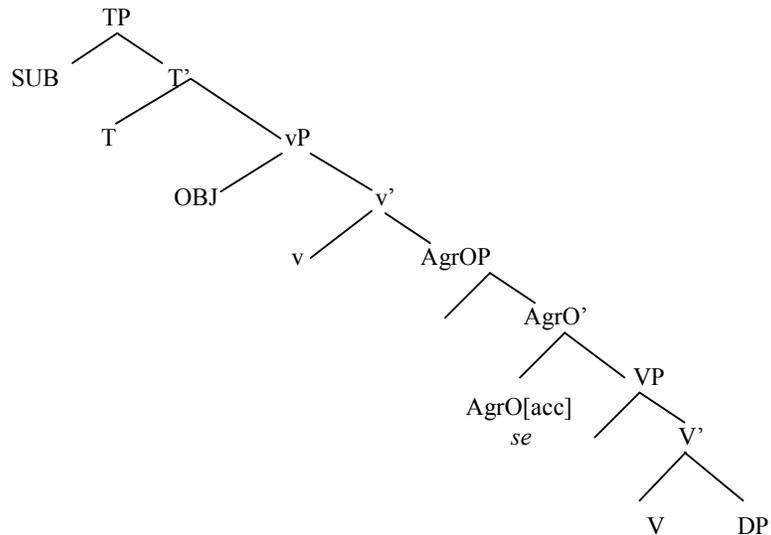
(3) Él se ha caído
He SE has fallen
"He has fallen down"

(4) Él se ha ido
He SE has gone
"He is gone"

Following Chomsky (1995), we will assume that the tense features of e.g. *ha* and *ido/caido* are Interpretable. In addition, the functional category hosting *se* must also have interpretable features. Note that *se* can be interpreted at either the semantic (i.e. LF) or phonetic interface. We will assume that *se* occupies a special functional category that, apart from containing aspect, is responsible for Case checking.

Just like a tensed verb (in most languages) makes nominative Case checking possible, a perfective verb would enable objective Case checking. Following Tenny (1987), Sanz (1996) links perfectivity and transitivity by assuming the feature [measure] to occupy a head that is checked by an object NP. Hence, 'I ate' is not perfective but 'I ate the apple' is. She assumes the presence of a Tr(ansitive)P as in e.g. Jelinek (1997), but an AgrO head might contain the same feature.

(5)



By inserting clitic *se* in AgrO, we assume that this functional head has perfective features. In addition, this functional category is responsible for the inherent Case checking of the object.

Torrego (1988) argues that inherent Case is semantic, interpretable Case. Following this view, we may assume that clitic *se* is responsible for the inherent Case of the affected object, because it is needed for interpretation. Furthermore, we may also follow Chomsky (1995) and argue that affected objects marked by *se* also check the structural accusative case by further moving from Spec AgrO to the specifier of *v*.¹

¹ As Sanz (2000) points out, the term inherent Case, when applied to direct objects, usually refers to partitive Case. However, the objects marked by *se* are those that turn an activity into an accomplishment and therefore are considered specific. In Sanz's (2000) analysis, they are

It seems to be the case that there are different types of *se*, but the one we are discussing here can only be analysed as having aspectual properties. Crucially, it co-occurs with affected objects in transitive structures, as exemplified in (6), and with subjects of unaccusative verbs, as exemplified in (7)

- (6) a. El niño se comió la manzana
 the boy SE ate the apple
 “The boy ate up the apple”
 b. La niña se leyó el libro hasta el final del todo
 The girl SE read the book up to the end
 “The girl read the book up to the very end”
- (7) a. El niño se cayó
 the boy SE fell
 “The boy fell down”
 b. El barco se hundió
 The ship SE sank
 “The ship sank”

We assume that the object marked by *se* is affected by the action expressed by the verb. According to the examples in (6) *the apple* and *the book* denote the affected object, because the apple gradually disappears in the eating-process, and the number of pages one must read in order to finish the book gradually diminishes in the reading-process.

In what follows, we will assume that telic *se* checks person agreement with the subject of unaccusative verbs like *caerse* in the example in (7a), which explains why a singular but not a plural form of the subject can co-occur with the verb also in singular, as exemplified in (8):

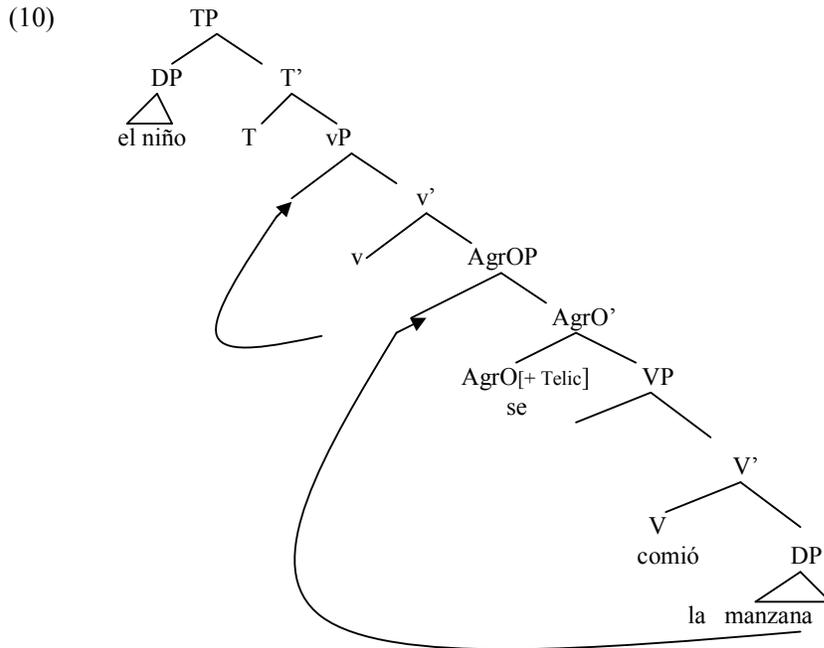
- (8) Se cayó él / *Se cayó ellos
 SE fell(sing) he/ SE fell(sing) they
 “he fell” / “they fell”

If the verb is in the plural, note that the example turns to be grammatical again:

- (9) Se cayeron ellos
 SE fell(plur) they
 “They fell”

closer to the types of objects that Tenny (1987, 1994) considers measurers of an action, and therefore she assumes that the semantic Case that these objects receive is related to the feature [measure].

As for the case of the affected object, we assume that it is generated as an internal argument of V, but raises further into Spec AgrO to check the telic feature against *se* and then moves into Spec v to check accusative case:



Sanz (2000) provides a thorough analysis of the clitic *se* in transitive constructions like the following:

- (11) Pedro *se* leyó un libro
 Pedro *se* read a book
 "Pedro read a book (completely)"
- (12) Pedro (**se*) leyó un libro durante tres horas
 "Pedro read a book for three hours"
- (13) a. Pedro #(se) leyó un libro en una hora
 Pedro *se* read a book in an hour
 "Pedro read a whole book in an hour"
 b. Pedro #(se) leyó el libro entero ayer
 Pedro *se* read the book whole yesterday
 "Pedro read the whole book yesterday"

It is claimed that the presence of the clitic means that the sentence is unambiguously telic, and that if there are elements that point at the sentence being telic (like the adverb *todo* “all” with consumption verbs, or the adverbial *en una hora* “in an hour”, the clitic appears as a marker. She provides the following examples:

- (14)a. El niño ??(se) lo comió todo
 the child *se* it ate all
 “The child ate it all”
- b. ??(Se) comió el pastel entero
se ate the cake whole
 “He ate the whole cake”
- c. ??(Se) comió todos los pasteles
se ate all the cakes
 “He ate all the cakes”
- d. ?(Se) bebió una cerveza en un minuto
se drank a beer in a minute
 “He drank a beer in a minute”
- e. ?(Se) comió el pastel en una hora
se ate the cake in an hour
 “He ate a cake in an hour”

Hence, the telicity of the sentence is ensured as long as *se* is present.

In sum, we have argued that telic *se* should be analysed as AgrO. According to Chomsky’s (1995) proposal, Agreement should be extended for both subjects and objects. In the analysis above, we have in fact shown that telic *se* marks the inherent case of affected objects, whereas structural case is checked against Spec, *v*. Chomsky (1995) distinguishes interpretable features, i.e. features relevant for semantic interpretation, from non-interpretable ones, which are irrelevant for interpretation. Non-interpretable features must be checked and deleted in the course of the derivation, while interpretable features need not enter into checking relations. In this system, the structural accusative case of the affected object must be checked against *v* in the course of the derivation. In contrast, the [telic] feature is interpretable and can stay along the derivation.

4. The present study

4.1 Method

In this paper, data from three Spanish children are examined. We intend to investigate the acquisition of telic *se* by two-year-olds. Since there is variation in contexts of “perfectivity”, (past simple versus present perfect) children have to find out which strategy is used in Spanish, and, therefore, in the course of development, it is expected that they will make mistakes both in the production and comprehension of telic *se*. Moreover, the compositional analysis of aspectual perfectivity in Van Hout (2003) predicts difficulty with the acquisition of this clitic.

We want to test whether children have any difficulty with *se* in telic contexts. So far, there are no other initial results which can confirm the problematic status of aspectual *se* in Spanish

Our first contribution to this study is based on the longitudinal acquisition studies of past tenses (perfectum and imperfectum) with complex structure verbs with *se* and other simple structure verbs without *se* in the current Childes corpora.

The data analysed are available on the CHILDES database and include the whole period of clitic *se* development (until target production is achieved) for three children: Irene, Magin and María. The ages range between 1;6 and 2;11 for the three children, sampled every month and with intervals every three months.

4.2 Quantitative analysis

We examined children's productions of Present and Past tenses. We wanted to test whether children have a preference for clitic *se* in perfective contexts as is the case in adult grammar. The data showed that children omitted clitic *se*, particularly in contexts where it is especially required: in perfective contexts with verbs inflected for the past simple tense and with an affected object.

First we looked at the correct use of clitic *se* and we obtained the following.

	Correct	Errors	Omissions
1;6-1;8	1	3	3
1;9-1;11	12	5	19
2;0-2;2	44	4	1
2;3-2;5	64	5	0
2;6-2;8	61	0	0
2;9-2;11	64	0	0
Total	246	17	23
% correct	246/286 (86%) correct		57/382 (14%) incorrect

Table 1. Correctness of constructions with aspectual SE Present tense

Simple past		
	Correct	Omission
1;9-1;11	1	0
2;0-2;2	3	0
2;3-2;5	4	8
2;6-2;8	0	0
2;9-2;11	10	0
Total	18	8
% correct	18/26 (69.2%) correct	
	8/26 (30.8%) incorrect	

Table 2. Correctness of constructions with past tense with aspectual SE

Imperfectum		
	Correct	Omission
1;9-1;11	0	0
2;0-2;2	1	0
2;3-2;5	3	1
2;6-2;8	0	0
2;9-2;11	41	1
total	45	2
% correct	45/47 (95.7%) correct	2/47 (4.3%) incorrect

Table 3. Correctness of constructions with past tense with aspectual SE

Present perfect		
	Correct	Omission
1;9-1;11	8	2
2;0-2;2	20	4
2;3-2;5	12	1
2;6-2;8	8	0
2;9-2;11	18	1
Total	58	8
% correct	58/66 (87.9%) correct	8/66 (12.1%) incorrect

Table 4. Correctness of constructions with past tense with aspectual SE

Simple past		
	Correct	Omission
1;9-1;11	1	0
2;0-2;2	0	0
2;3-2;5	1	8
2;6-2;8	0	0
2;9-2;11	20	1
Total	22	9
% correct	22/31 (71%) correct	9/31 (29%) incorrect

Table 5. Correctness of constructions with past tense without aspectual SE

Imperfectum		
	Correct	Omission
1;9-1;11	0	0
2;0-2;2	1	0
2;3-2;5	2	0
2;6-2;8	0	0
2;9-2;11	3	0
Total	6	0
% correct	6/6 (100%) correct	0/6 (0%) incorrect

Table 6. Correctness of constructions with past tense without aspectual SE

Present perfect		
	Correct	Omission
1;9-1;11	4	0
2;0-2;2	12	2
2;3-2;5	19	0
2;6-2;8	7	0
2;9-2;11	5	0
Total	47	2
% correct	47/49 (95.9%) correct	2/49 (4.1%) incorrect

Table 7. Correctness of constructions with past tense without aspectual SE

The first general finding was that all verbal tenses with or without *se* give a significant relationship with age ($P < 0.001$). In addition, we found that children commit the same amount of errors in constructions with and without clitic *se*. However, the number of productions of verbs with *se* varies depending on the verbal tense they use. Examples of errors are shown at the qualitative analysis section. We found that children produce significantly more errors in sentences with a simple past than sentences with an imperfectum ($P < 0.0016$). Note that as we argued before in the introduction, unlike the simple past, the present perfect in Spanish may be ambiguous between a perfective or an imperfective reading.

In Spanish, there is a large number of pronominal verbs that require a benefactive clitic, such as *mancharse* (stain). Unlike adults, children tend to use clitic *se* in contexts of imperfectivity. In fact, there is no significant difference between children's productions of *se* with the present perfect or with the imperfectum.

Our next step was to find out whether children omit clitic *se* in the same contexts. As it is shown in tables 1-7, the data show that children omit clitic *se* in contexts where adult grammar requires it, namely with perfectivity. Compare the omission errors of *se* with the past simple with those with the imperfectum, in the tables above.

One way to test children's development of perfectivity was to look at the correct use of clitic *se* along with other temporal adverbs that imply perfectivity. We found that children significantly don't commit errors in this respect. Also, we found that children significantly didn't commit errors with clitic *se* on sentences with affected objects.

4.3 Qualitative analysis

Examples of errors that children made are examined next. In the first place, a number of omission errors in the Present are exemplified in the next examples:

- (15) a. No *(te) *escapes* (Irene, 2)
 don't hide
 b. ... y el *pequeñin* *(se) *queda* (Irene, 2;14)
 ... and the kid stayed

Both verbs *escaparse* “hide” and *quedarse* “stay” are pronominal verbs and therefore require the use of a benefactive clitic. The number of omissions of this type of clitic were very low (14%). In contrast, we observed a large number of errors with respect to telic verbs like *caerse* “fall”, *ponerse* “turn”, *mancharse* “stain”, *romperse* “break”, and *encontrarse* “find”. Some examples follow:

- (16) Simple past
- a. *cayó* # *ota ve* *cayo* # (Irene, 1;8)
 fell(3rdsing) again fell(3rdsing)
 “he fell again”
- b. *lo* *cojo* *lo* *tito* *y* *Tina* *cayó* (Irene 1;10)
 AccCL (I) take AccCL (I) through and Tina fell
 “I take it and Tina fell”
- c. *puso* *malito* (Irene, 2;14)
 turned(3rdsing) sick
 “He got sick”
- d. *cayó* *la* *abuela* (Magín 1;10)
 fell the granny
 “The granny fell”
- e. *cayó* *y* *rompió* *la* *quisma* (Irene, 1;11)
 fell(3rd sing) and broke(3rd sing) the head
 “The fell and broke his head”
- f. *taba* *Juanito* *y* *Aba* *y* *encontraron* *una* *mariposa* (Irene, 2)
 was Juanito and Aba and found a butterfly
 “Juanito and Aba were there and found a butterfly”
- (17) Present Perfect
- a. *ha caído* *mami* (Magín 1;11)
 has fallen mom
 “Mom has fallen”
- b. *ha manchado* *la* *rodilla* (Magín 1;11)
 has stained the knee
 “He has stained the knee”

The data show that two-year-olds commit a large number of omission errors with Past tenses, more than with Present tenses. We want to relate this contrast to the fact that children have difficulty with perfectivity.

5. Discussion

Much acquisition research on children's development of aspect has focused on the imperfective-perfective distinction. To the extent that developing understanding of aspect begins with the initial comprehension of the semantics, we expect that the mapping of the semantics onto the appropriate morpho-syntactic levels will be a locus of learning for children and of variation among different child languages.

According to Hodgson's (2003) findings based on a comprehension/production study, Spanish children have difficulty in assessing perfective meaning from perfective morphology. The data presented in this study suggests that the arguments of the verb, particularly the direct object, play a significant role in learning the entailment of completion carried by perfective morphology. The presence of the agent may play a small role in acquiring aspectual meaning, however the data shows that it was not as strong of an indicator as was the presence of the full change of state of the direct object.

Clitic *se* has been analysed as heading AgrO, and as such it can be analysed as a direct object, which also plays a significant role on the acquisition of "perfectivity".

Van Hout (2003) has proposed a crosslinguistic difference with respect to acquisition of telicity. In this sense, she makes a distinction between languages that mark the entailment of completion on the verb itself, as is the case of Slavic languages (Russian, Polish), in which perfective aspect is marked as a prefix on the verb, and languages which mark telicity by combining the properties of the verb and its object, as in the case of Germanic languages and Finnish. Spanish is also a language where the entailment of completion is obtained from the properties of the verb and its arguments, in particular its direct object, as argued by Hodgson (2003). This author also points out that Van Hout's distinction between what she calls predicate telicity and compositional telicity is relevant for acquisition. The acquisition of languages like Russian and Polish, which have predicate telicity in which the entailment of completion is calculated by the verb + affix, is predicted to be easier than the acquisition of compositionality telicity in languages like English and Dutch, in which the completion of the event is computed based on the joint properties of the verb + object.

Rigau (1994) analyses *se* as a benefactive clitic which will always appear in aspectual contexts of perfectivity, since it requires the presence of an internal argument. According to Sanz (1996), clitic *se* requires "telicity" or the presence of an *affected* internal argument in the sense of Tenny (1994). According to our longitudinal studies of the acquisition of aspectual *se*, the main error that children make is to omit aspectual *se* in cases where it is required, namely with telic verbs. We argue that they omit this because they usually opt for less complicated structures represented by verb+affix combinations, which typically represent atelic configurations. The data presented in this study show that children do not make any meaningful errors when aspect is part of the verb morphology. i.e., with the imperfectum.

De Miguel and Fernandez Lagunilla (2003) claim that Perfectivity is not a sufficient condition to allow for the presence of aspectual *se*. Rather, its presence relies on the complexity of certain events. In any case, we can argue that the compositional analysis of clitic *se* + verb makes the right prediction of difficulty in languages like Spanish, where aspectuality is not only morphologically marked. Furthermore we have claimed that the presence of *se* is highly related to the presence of an Interpretable feature that is responsible for inherent case assignment. It seems to be the case that normal developing children do not have any difficulty with agreement checking relations where non-interpretable features are involved. In contrast, it may be the case that the acquisition of interpretable features takes longer. However, more research is required. In particular, we hope to obtain experimental data soon in order to confirm or reject our hypothesis.

References

- Chomsky, N. (1995) *The Minimalist Program*. Cambridge, MA: The MIT Press.
- Guasti, M.T. (1993/94) 'Verb syntax in Italian child grammar: finite and nonfinite verbs'. *Language Acquisition* 3, 1–40.
- Hodgson, M. (2003) 'The Acquisition of Spanish Perfective Aspect: a study on children's production and comprehension'. *ZAS Papers in Linguistics* 29, 105 – 117.
- Jelinek, E. (1997). 'Topic and focus in Navajo inverse', in: B. Agbayani and S.-W. Tang, (eds.) *The proceedings of the Fifteenth West Coast Conference on Formal Linguistics*. Stanford, CA: CSLI, 241-255.
- Lyczkowski, D. (1999) *Adquierelelo: On the Acquisition of Pronominal Object Clitics in Spanish*. Master thesis, Harvard University.
- Miguel, E. de and M. Fernández Lagunilla (2000) 'El operador aspectual *se*.' *Revista Española de Lingüística* 30, 13-43.
- Rigau, G. (1994) 'Les propietats dels verbs pronominals'. *Els Marges* 50, 29-39.
- Sanz, M. (1996) *Telicity, Objects and the Mapping onto Predicate Types. A cross-linguistic study of the role of syntax in processing*. Doctoral dissertation, University of Rochester, N.Y.
- Sanz, M. (2000) *Events and Predication. A New Approach to Syntactic Processing in English and Spanish*. Current Issues in Linguistic Theory 207. Amsterdam: John Benjamins.
- Schaeffer, J. (2000) *The Acquisition of Direct Object Scrambling and Clitic Placement*. Amsterdam: John Benjamins.
- Tedeschi, R. (2006) 'The Acquisition of object clitics: Data from an elicited production task'. *Annali Online di Ferrera – Lettera*, vol 2, 31-42.
- Tedeschi, R. (this volume) 'Referring expressions in early Italian. A study on the use of lexical objects, pronouns and null objects in Italian pre-school children'.
- Tenny, C. L. (1987) *Grammaticalizing Aspect and Affectedness*. Cambridge, MA: The MIT Press.
- Tenny, C.L. (1994) *Aspectual Roles and the Syntax-Semantics Interface*. Dordrecht: Kluwer.
- Torrego, E. (1998) *The Dependencies of Objects*. Cambridge: MIT Press.
- Van Hout, A. (2003) 'Acquiring telicity crosslinguistically. On the acquisition of telicity entailments associated with transitivity', in: M. Bowerman and P. Brown (eds.)

Crosslinguistic perspectives on argument structure: Implications for learnability.
Mahwah: Lawrence Erlbaum.

Wexler, K., A. Gavarró and V. Torrens (2004) 'Feature checking and object clitic omission in child Catalan', in: R. Bok-Bennema, B. Hollebrandse, B. Kampers-Mahne and P. Sleeman (eds.) *Romance Languages and Linguistic Theory 2002*. Amsterdam: John Benjamins.

Zagona, K. (1996) 'Compositionality of Aspect: Evidence from Spanish Aspectual *Se*', in: C. Parodi et al. (eds.), *Aspects of Romance Linguistics*. Washington, DC: Georgetown University Press, 475-488.

Germanic and Romance Discourse Devices in Acquisition*

Jacqueline van Kampen and Manuela Pinto
Utrecht University

1. Introduction

In a discourse fragment, say a story, we see a set of intended referents (for example: *a girl, an attic, a bed, a little bear*). The members of that set appear and reappear in changing configurations when the story unfolds. That is due to the fact that the head of each new predicate selects referents for a configuration according to its sub-categorization/theta frame.

Language acquisition begins with learning such predicate frames by means of situation-bound clauses, since such clauses are naturally supported by gesture-sustainable referents ('physically given' referents, cf. Ariel 2001). The reference tracking devices are at first mainly 1st and 2nd person pronouns, demonstratives, bare nouns as quasi names and, as we will show, 3rd person pronouns/clitics, but only in as far as they are (overtly or implicitly) accompanied by a gesture that brings in focus a referent in the situation. As soon as a minimal amount of predicate subcategorization frames has been acquired and stacked up in the lexicon, a completely new development sets in (Van Kampen 2006). Child language starts adding the devices that perform reference tracking in (linguistic) discourse. There is a rise in the use of articles and 3rd person pronouns/clitics. Due to this development, the language and its user become more situation-free and discourse-bound (Van Kampen 2002, 2004).

The reference tracking devices are learned from the adult input. They indicate whether an argument is newly introduced or has already been referred to earlier. This at least is the contribution of the West-European article and pronoun system. The Dutch, French and Italian story fragments in (1) show how dense the reference tracking devices can be.

- (1) a. Dutch
[De kleine beer]_i ging de trap op [naar de zolder]_k. **Daar**_k zag *hij*_i [een meisje]_m. *Hij*_i was stomverbaasd. **Die**_m had *hij*_i nog nooit gezien. *Ze*_m lag in *zijn*_i bedje. *Ze*_m sliep.
- b. French
[Le petit ours]_i grimpa l'escalier jusqu'[au grenier]_k. **Là**_k, *il*_i vit [une jeune fille]_m. *Il*_i était stupéfait. **Elle**_m/[**cette fille**]_m, *il*_i ne **l**_m'avait jamais vu. **Elle**_m s'était couchée sur le petit lit. **Elle**_m dormait.

* The research for this paper received financial support from the Uil OTS and NWO (first author, grant 355-70-009). We would like to thank Sergio Baauw for useful comments.

c. Italian

[L'orsetto]_i sali [in soffitta]_k. ***Li***_k, *pro*_i vide [una ragazzina]_m. *Pro*_i fu sorpreso. ***Lei***_m/[***la ragazzina***]_m, *pro*_i non *I*_m'aveva mai vista prima. *Pro*_m era stesa nel *suo*_i lettino. *Pro*_m dormiva.

“The little bear went upstairs to the attic. There he saw a girl. He was flabbergasted. He had never seen her. She was lying in his bed. She was asleep”

The reference tracking anaphoric pronouns in (1) are indicated with subscripts under the italics. In addition to that system there is a superimposed discourse device marked by bold face. These are the specific anaphors that have a topic-shift function. (Van Kampen 2004) They indicate that the new clause offers one of its arguments as a new point of orientation, different from the orientation point of the preceding sentence. The choice of the antecedent is not free. It has to be the argument marked as prominent in the preceding clause. The ‘focus’ of the preceding clause is turned into the topic of the new sentence.¹ These anaphors are in principle sentence-initial and topic-shifting.

Germanic V2nd languages (Dutch/German/Swedish) use a demonstrative variant to indicate the topic-shift. They appear as such in Dutch (Comrie 2000; Van Kampen 1997:92ff, 2004), in German (Bosch, Katz and Umbach 2007; Diessel 1999; Zifonun et al. 1997), in Swedish (Mörnsjö 2002). These ‘anaphoric demonstratives’, which we will call *d*-pronouns, derive from the article paradigm (German *der, die, das* etc.; Swedish *den, det*) or from the demonstrative paradigm (Dutch *die, dat*; German *denen*).² The use of the *d*-pronoun is a stylistically smooth option. Romance languages, by contrast, are more restrictive. In case of shifting a topic, they may use a full, strong, personal pronoun in adjunct position, but they often use a full DP in adjunct position, see (1)b,c.

The use of the *d*-pronouns is not open to Romance languages. This difference seems directly related to the V2nd type of Germanic versus the SVO pattern of Romance. We will demonstrate that in more detail. First we discuss the properties of these devices in Germanic Dutch and subsequently we will have a short look at Romance French and Italian. It will turn out that for the binary <±topic-shift> different languages make a choice from the same saliency hierarchy. Finally, we will look at the acquisition steps for <±topic-shift> devices in V2nd Dutch and in non-V2nd Romance French (see for French also Rozendaal this volume).

2. Properties of the *d*-pronoun

Germanic V2nd languages allow an aboutness position in the sentence-initial “Vorfeld”. Within generative grammar the “Vorfeld” appears as the C-domain with

¹ We follow here Reinhart’s (1981) characterization of the sentence topic as ‘what the sentence is about’.

² We will represent the *d*-pronoun by DEM in the glosses.

the finite verb in C⁰ position and an aboutness constituent in Spec,C. The Spec,C, we will argue, is cut out for anaphoric *d*-pronouns as <+topic-shift> device.

The topic-shifting *d*-pronouns are true discourse anaphoric pronouns. Unlike demonstratives, they are restricted to the sentence-initial scope position in Spec,C. The *d*-pronouns in (1)a have sentential scope, just like *wh*-pronouns and relatives have. They have therefore been indicated as A-bar anaphors in Van Kampen (1997:92ff).³ As discourse anaphors they remind of the (non-preposed) demonstrative pronoun, but the function is different and restricted to the preceding focus. This is reflected in the fact that in Dutch the free distal/proximate opposition of demonstratives is not present in the A-bar *d*-pronoun. The A-bar *d*-pronouns have the unmarked distal variant only, see the paradigm in (2)b and the examples in (3).⁴

- (2) a. demonstratives
 structural oblique
deze<-neuter> *dit*<+neuter> *hier* <+ proximate>
die<-neuter> *dat*<+neuter> *daar* <-proximate>
- b. *d*-pronouns
 structural oblique
die<-neuter> *dat*<+neuter> *daar*

- (3) a. Toen zag zij het huis van de beren.
 Then saw she the house of the bears
 Dat/*dit wilde ze van binnen zien
 DEM wanted she from inside see
 “Then she saw the house of the bears. She wanted to see it inside.”
- b. Zij ging eerst op de grote stoel zitten
 She went first on the big chair sit
 Maar die/*deze vond ze te hard
 But DEM found she too hard
 “First she sat down on the big chair. But she found it too hard.”

³ A-bar anaphors appear in A-bar position and they are related to an argument position. An A-bar position is a derived position in the syntactic tree where only non-arguments (*wh*-words/topics) can occur. An A-position is a position where only arguments (object/subject) can occur, namely theta positions and specifiers construed with agreement (Rizzi 1999).

⁴ Comrie (2000) includes the proximal *deze* as an example of a pragmatically conditioned ‘demonstrative’, our *d*-pronoun. Comrie relies on Huizinga’s (1936) *Erasmus* as the database. The use of *deze* is restricted to <+human> antecedents and to written Dutch, see (i) (ANS 997:29).

(i) Toen sprak de minister van Justitie. *Deze* hield staande dat
 Then spoke the secretary of Justice. This (one) made the contention that

The examples from Huizinga’s *Erasmus* show that the use of *deze* is not restricted to the sentence-initial position in written Dutch. It is somewhat like *the* latter in formal English and *ce dernier* in formal French. We will leave this construction aside. See Van Kampen (1997).

The form and function of the *d*-pronoun remind of the relative pronouns with an antecedent. The (High) German relative pronouns mainly use the same paradigm as the *d*-pronoun (*der, die, das, denen*). The Dutch relative pronouns are partly from the *d*-pronoun paradigm (*die, dat*), like German, and partly from the *w*-pronoun paradigm (*wat wie*), like English, see (4).⁵ For the reasons of this mixture, see Van Kampen (2007).

(4) relative pronouns in Dutch

	<i>d</i> -set <±neuter> referent	<i>w</i> -set <±animate> referent
structural	<i>die</i> <-neuter> <i>dat</i> <+neuter>	<i>wat</i> <-animate> <i>wie</i> <+animate>
oblique	[<i>daar</i>] ... (op)	

We elaborate this paradigm issue, because the relevance of the notion A-bar anaphor has not sufficiently been seen, perhaps due to the partial correspondences of the morphological paradigms. Various quantitatively oriented studies (Bosch Katz and Umbach 2007; Kaiser and Trueswell 2004) have posed the question what the referent of the ‘demonstrative anaphor’ would be. They subsequently found that it is an anaphor with a strong tendency to refer to the non-subject of the preceding sentence. In our view this should be reinterpreted by adding the A/A-bar distinction. The topic-shifting *d*-pronoun in (5)a appears in Spec,C A-bar position (Van Kampen 2004). In (5) *die* can be both subject and object, but only (5)a is grammatical. The *d*-pronoun cannot appear in sentence-internal A-position in (5)b.⁶

(5) [De beer]_i heeft [de lucht]_k in huis opgesnoven

‘The bear sniffed up the air in the house’

- a. [_{Spec,C} *Die*_k [_C vond]] *hij*_i verdacht ruiken
 DEM found he fishy smell
 ‘He thought that it had a fishy smell’
- b. [_{Spec,C} *Hij*_i [_C vond]]**die*_k verdacht ruiken
 He found *DEM fishy smell
 ‘He thought that it had a fishy smell’

The pronoun *die* may appear in sentence-internal position, but only when it is in the company of a focusing adverb like *nog* (‘still’), *ook* (‘also’) or *niet* (‘not’). In that case, *die* gets a marked contrastive interpretation and stress, as in (6). We assume that it is a sentence-internal A-bar position for contrastively marked

⁵ Note that the Latin use of ‘relative root clause connection’ (sentence-initial *w*-paradigm anaphor) in main clauses is probably a matter of topic-shift (Kühner and Stegmann 1992).

⁶ Dutch strongly prefers the use of a general demonstrative *die* for pronominalized topics. In many cases *die* replaces the <+neuter> *dat*, especially when the antecedent is <+animate>, as in *het*<+neuter> *meisje. Die...* (‘the girl. DEM ..). Gender-evading preferences are also at work in the Dutch relative system.

‘Association with Focus’. We follow here Von Stechow (1991), cf. Van Kampen (1997: 134f).

- (6) Is het meisje hier nog geweest?
 Is the girl here still been?
 Nee, ik heb *dié* nog niet gezien, wel de beer
 No, I have DEM yet not seen, but the bear
 “Has the girl been here? No, I haven’t seen her yet, but the bear I have”

The use of the A-bar *d*-pronoun contrasts with the use of the 3rd person pronoun in Dutch. The latter maintains the topic of the preceding sentence. A 3rd person pronoun is used in A(rgument)- position and indicates that there is no topic-shift (<-topic-shift>). See the examples in (7)a of the object *het meisje* (‘the girl’) maintained as the object *haar* (‘her’) and (7)b of the subject *de kleine beer* (‘the little bear’) maintained as the subject *hij*.

- (7) [De kleine beer]_i zag [het meisje]_k in zijn bedje
 “The little bear saw the girl in his bed”
 a. Hij_i vond *haar*_k/**die*_k er lief uitzien
 He found her/**DEM* there nice out-see
 “He thought that she looked rather nice”
 b. dat meisje vond *hij*_i/**die*_i er lief uitzien
 that girl found he/**DEM* there nice out-see
 “He thought that that girl looked rather nice”

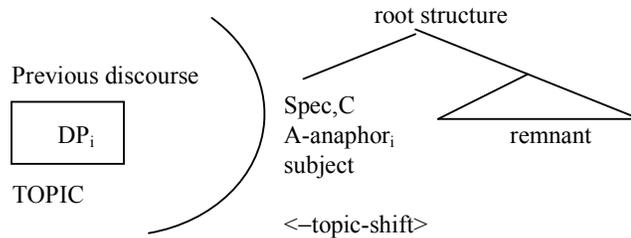
In the last sentence (7)b, the object has been topicalized, which induces subject-verb inversion in V2nd Dutch. Of course, the subject also may occupy the sentence-initial Spec,C position, see (8).

- (8) *Hij*/**die* (de kleine beer) vond het meisje lief
 He/**DEM* (the little bear) found the girl nice
 “He thought that the girl was nice”

In (8) the subject is a 3rd person pronoun and there is no topic-shift. We take here the position of Holmberg (1986) and Rizzi (1991) that the Spec,C position in V2nd languages is an A-bar position for topic-hood, but that it may be reinterpreted as a Spec,I A-position for the canonical subject. That is to say, the sentence-initial pronouns in (5)a and (8) both occupy the Spec,C position, but only in (5)a it is an A-bar position for the *d*-pronoun.⁷ The personal pronoun *hij* in (8) is in a canonical subject position, as represented by the tree structure in (9).

⁷ German seems to fit into this structural picture too, as evidenced by a corpus study of German newspaper texts in Bosch and Umbach (2007). In this corpus, 3rd person subject pronouns appear equally (roughly 50-50%) in sentence-initial and in sentence-internal position, whereas 3rd person object pronouns uniquely (almost 100%) occupy a sentence-

(9)



Objects in sentence-initial position are A-bar topics. In unstressed contexts, object pronouns in Spec,C will invariably appear as a *d*-pronoun, see (10).

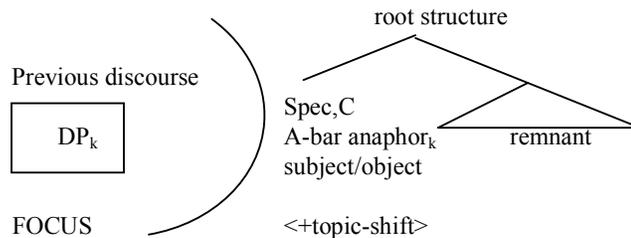
(10) [De grote beer]_i wees naar [zijn zoontje]_k

“The big bear pointed at his son”

- a. *Die*/**h(e)m*_i (zijn zoontje) vond *hij*_k (de grote beer) wat klein
DEM/**him* (his son) found he (the big bear) what small
“He found him a little bit small”

The anaphor **h(e)m* in (10) is an unstressed personal pronoun. For that reason, it cannot enter the Spec,C. The structural conditions for the <+topic-shift> *d*-pronoun in Spec,C are given in the tree in (11).

(11)



The formal characteristics of the *d*-pronoun are now argued to be as in (12).

(12) Topic *d*-pronouns <+D, +C>

- are an argument located in Spec,C of V2nd languages.
- indicate a topic-shift.
- follow the paradigm of some A-bar anaphor (distal demonstrative in Dutch, definite article in German, Swedish).
- take as an antecedent a major constituent focused in the preceding clause.

internal position. The *d*-pronouns (subjects and objects) appear for 93% in the sentence-initial Spec,C position. One wonders whether the remaining 7% fits a contrastive interpretation.

Up till now we have discussed the characteristics of the A-bar reference tracking device, points (12)a-c. In the next section we will have a look at the discourse properties of the antecedent, point (12)d.

3. Properties of the antecedent

Whereas personal pronouns may refer to any antecedent DP, the Germanic A-bar *d*-pronouns require that their antecedent be a major constituent and be marked as prominent in the preceding clause. The *d*-pronoun in (13)a refers back to an argument that had a prominence with focus quality.

- (13) De grote beer zijn zontje zag in zijn bedje **een meisje** liggen
 The big bear his son saw in his bed a girl lying
 “The big bear’s son saw a girl lying in his bed”
 a. *Die* keek erg verbaasd (een meisje / *de grote beer / *zijn zontje)
 DEM looked very surprised (*the big bear / *his son / a girl)
 “She looked very surprised”

The constituent *een meisje* carries the sentential stress (Cinque 1993; Evers 2003). For that reason, *een meisje* can be picked up as the shifted topic in the next sentence.

If, by contrast, the object phrase *het meisje* moves to the left as in (14), or if it were pronominalized by *haar* (‘her’) as in (15), it loses the focus and sentential stress. Therefore, it is no longer referred to by the *d*-pronoun.

- (14) De beer zijn zontje heeft het meisje nog **op de zolder** gefotografeerd
 The bear’s son has the girl yet in the attic photographed
 “The bear’s son has taken a picture of the girl in the attic”
 a. *Die* was erg klein (de zolder / *de beer / *zijn zontje / *het meisje)
 DEM was very small (the attic / *the bear / *his son / *the girl)
- (15) Het meisje holde de trap op. De kleine beer riep **haar** nog na
 The little girl ran the stairs on. The little bear called her still after
 “The little girl ran up the stairs. The father of the little bear called after her”
 a. **Die* luisterde niet
 *DEM listened not
 “She didn’t listen”

The major constituent property of the A-bar anaphor is demonstrated in (16).

- (16) Heb jij **het vriendinnetje van de kleine beer** naar boven zien gaan?
 Have you the girlfriend of the little bear upstairs see go?
 “Did you see the little bear’s girlfriend go upstairs?”
 a. Ja, *die/!zij* is naar bed gegaan (the girlfriend)
 Yes, DEM/!she is to bed gone
 “Yes, she went to bed”

- b. Nee, *hij/*die* heeft haar een bord pap gegeven
 No, he/*DEM has her a plate porridge given (the little bear)
 “No, he has given her a plate of porridge”

Usually, the argument in focus is not the subject, and hence the subject is usually not the antecedent of the *d*-pronoun, but that is not relevant. The *d*-pronoun may in principle refer back to a subject, if the subject has sufficient prominence, see (17).

- (17) Het meisje met de gouden haren is ook gefotografeerd
 The girl with the golden locks is also photographed
 “The girl with the golden locks was taken a picture of”
 a. *Zij/ze* (= het meisje met de gouden haren) is ooit model geweest
 She (= the girl with the golden locks) is ever model been
 “She used to be a model”
 b. *Die* (=het meisje met de gouden haren) is ooit model geweest
 DEM (= the girl with the golden locks) is ever model been
 “She used to be a model”

The constituent *het meisje met de gouden haren* is running subject and referred to by the personal nominative pronoun *zij/ze* (<- topic-shift>) in (17)a. Yet, it can also be referred to by the *d*-pronoun demonstrative *die* in (17)b. The construction has the flavor “as opposed to others”. A contrastive effect for topic-shift is not uncommon, but we like to stress here that it is not essential for the A-bar *d*-pronoun in Spec,C. It may be noticed, though, that a contrastive effect of the sentence-internal *die*, see for example (6), is inevitable and obligatory, as is its stress.

4. A-bar anaphors for <+topic-shift> in French and Italian

The distinction between the two kinds of free anaphoric pronouns (A-/A-bar) also holds for French and Italian.⁸ Since French and Italian do not belong to the V2nd type, and by consequence do not have a general rule for moving a constituent

⁸ English does not have a specific pronominal device for topic-shift. English may use the demonstrative *that* in sentence-initial position, but only to refer to a preceding state of affairs, rather than to a preceding antecedent taken up as a topic, see (i)a (cf. Mikkelsen 2005; among others). In the latter case, English may use a stressed personal pronoun, as in (i)b.

- (i) a. I like to wear blue suede shoes
 That (‘wearing blue suede shoes’) gives me the idea of being Elvis
 b. I only like Maxima. Shè is a star

In (i)b the object *Maxima* in focus is taken up as the topic of the new sentence by the stressed pronoun *shè*. Of course, stressing a pronoun may also result in contrastive interpretations (cf. Comrie 2000; Bosch and Umbach 2007). This option (stressed 3rd person pronoun) is available in Dutch/German as well. This makes the stressed pronoun in English an unreliable candidate for a comparison with the *d*-pronoun.

to Spec,C, these languages employ different devices for <+topic-shift>. We will first consider the discourse devices for <±topic-shift> in French and subsequently in Italian. It will turn out that different languages make different choices from a general saliency hierarchy (cf. Ariel 1990; Gundel et al.1993) to express <±topic-shift>.

French prefers a left-dislocation construction to induce <+topic-shift> (Gívon 1983; Ashby 1988: 206). A left-dislocated constituent in an A-bar position is doubled by a clitic with which it shares case, number and gender features, consider (18). The A-bar constituent that occupies the left-dislocated position can be a lexical DP or a pronoun. In French, A-bar anaphors are full strong pronouns.

- (18)a. Le petit ours, [IP il grimpaît l'escalier]
 The little bear he-cl climbed the stairs
 "The little bear went upstairs"
- b. Lui, [IP il grimpaît l'escalier]
 He, he-cl climbed the stairs
 "He went upstairs"

This type of dislocation is called Clitic Left Dislocation (CLLD). It is often assumed that the dislocated constituent is base-generated in left-dislocated position and licensed by rules of predication (Chomsky 1977; Cinque 1990; see also Avram and Coene this volume). The subject or object clitic is locally bound to the dislocated element as a kind of 'shadow pronoun'.⁹

The <±topic-shift>.device in French is illustrated in (19) and (20) below. When there is no topic-shift, the 3rd person clitic appears in argument position, see (19).

- (19)[Le petit ours]_i a voulu suivre [le grand ours]_k
 "The little bear wanted to follow the big bear"
- a. *Il*_i/**lui*_i, *il*_i était curieux
 He/*he, he was curious
 "he was curious"

When there is a topic-shift, the strong 3rd pronoun is used in dislocated A-bar position. The dislocated pronoun is doubled by a clitic.¹⁰ See for arguments that the strong pronoun in French is dislocated Lambrecht (1981) among others.

⁹ The term 'shadow pronoun' was used by Perlmutter (1972) and is due to the Arabian grammatical tradition.

¹⁰ In the examples (18)-(20) both subject and object arguments represent a masculine person. The masculine (unstressed) clitic *il* has a strong pronominal variant *lui*. The feminine strong pronominal variant is a stressed *elle*, as opposed to the unstressed clitic *elle*. Now topic-shift of <+feminine> would give *elle*, *elle*. Avoidance of the adjacency of two identical elements leads to: stressed *elle* for <+topic-shift>, versus unstressed *elle* for <-topic-shift>. Of course, stressing a pronoun may also result in contrastive interpretations. The problem reminds of the use of a stressed pronoun in English. A stressed pronoun may be used for topic-shift if the language does not have both weak and strong pronouns.

- (20) [Le petit ours]_i a voulu suivre [le grand ours]_m
 “The little bear wanted to follow the big bear”
 a. [lui, il]_m/*il_m grognit un peu
 He,he/*he grumbled a bit
 “he grumbled a bit”

It may be added that the non-V2nd languages (French, Italian, English) may use their strong 3rd person pronouns for <+topic-shift>, but they need not to. They may as well repeat the full DP *le grand ours, il* (‘the big bear, he’).

For Italian, Grimshaw (1995) and Grimshaw and Samek-Lodovici (1998) have shown that *pro*-drop is restricted to ‘topic-connected’ arguments. Carminati (2002) scrutinized the quantitative effects of Grimshaw’s statement. She subsequently found that the anaphor *pro* shows a strong tendency to refer to the subject of the preceding sentence. See also Serratrice (this volume). Within the present context we would like to stress again that the subject-tendency is not grammatically determined and hence in principle irrelevant, cf. (17).

The <±topic-shift> device in Italian is illustrated in (21) and (22) below. A difference with French is that Italian applies *pro*-drop if the pronominal subject is not contrastive, but apart from that the devices for <±topic-shift> seem to be identical. The <-topic-shift> A-anaphors are null (*pro/Agr*) when subject, and they are a clitic when object. Both appear within A-structure/IP. Note that it is generally assumed that clitics are in an A-bar position, since they bind an (empty) argument position. The distribution A/A-bar as we use it here is meant differently. It intends to separate A-bar anaphors in sentence-initial position with sentential scope from all A-anaphors within IP, including the clitic.

- (21) [L’orsetto]_i vide [la ragazzina]_m solo la sera
 The little bear saw the girl only in the evening
 “The little bear saw the girl only in the evening”
 a. *Pro*_i/*[lui, pro]_i non l_m’aveva mai vista prima
 (He) not her-clhad ever seen before
 “He had never seen her before”

When there is a topic-shift, the A-bar anaphoric strong 3rd pronoun is used in dislocated position. The dislocated pronoun is doubled by a *pro* when subject or a clitic when object. It stands to reason that the dislocated element can also be a noun.

- (22) [L’orsetto]_i salì in soffitta per salutare [la ragazzina]_m.
 The little bear went-up in attic for greet the girl
 “The little bear went upstairs to the attic to say hallo to the girl”
 a. [Lei, pro_m]/*pro_m stava ancora dormendo
 “She was still sleeping”
 “She was still sleeping”

The structural conditions for <+topic-shift> in Romance French and Italian (24) are then argued to be parallel to the structural conditions for <+topic-shift> in V2nd Germanic Dutch (23).

- (23) Structural conditions for <+topic-shift> anaphors (Germanic V2nd)
- a. The A-bar anaphor refers to the focus prominence of the preceding clause.
 - b. The A-bar anaphor is a *d*-pronoun that is in principle restricted to the A-bar Spec,C position.
- (24) Structural conditions for <+topic-shift> anaphors (Romance non-V2nd)
- a. The A-bar anaphor refers to the focus prominence of the preceding clause.
 - b. The A-bar anaphor is a strong personal pronoun that is in principle restricted to the dislocated A-bar position.

The Romance/Germanic difference is that the Germanic A-bar anaphor binds an empty place (a trace), whereas the Romance A-bar anaphor binds a clitic or *pro*/Agr, see (25).

(25)

Topic-shift	Dutch/German	French/Italian
Anaphor	Spec,C A-bar <i>d</i> -pronoun	clause adjoined A-bar personal pronoun
Argument	empty position in A-structure	clitic or <i>pro</i> /Agr in A-structure
Antecedent	± subject non-topic	± subject non-topic

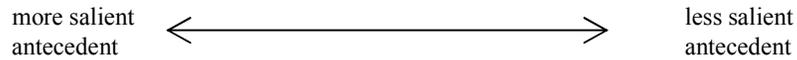
The V2nd type of language supports a more grammaticalized construction to express the topic-shift discourse function

5. Saliency hierarchy for anaphoric pronouns

Ariel (1990), Gívon (1983), Gundel et al (1993), among others, have set up accessibility hierarchies. They propose that the form of anaphoric expressions signals the relative accessibility of the antecedent. There is a reversed correlation between the two. Antecedents that are already very accessible need no more than a simple anaphoric expression. These anaphoric expressions are ranked high on their hierarchy scale. Antecedents that are less accessible need a more specific anaphoric expression. These anaphoric expressions are ranked low on their hierarchy scale. Consider the accessibility hierarchy for pronominal elements taken from Ariel (2001: 29) in (26).¹¹

¹¹ Note that Ariel talks about ‘referent’ where we prefer ‘antecedent’. Ariel’s accessibility of the antecedent is determined by saliency factors such as topichood, recency and stereotype-ness of the antecedent. Ariel makes a distinction between physical givenness and linguistic givenness of an antecedent, roughly our situation-bound versus discourse-bound anaphors.

(26) null > clitic > reduced pronoun > full pronoun > demonstrative



A more structural view on anaphoric pronouns is possible for the set of discourse devices discussed here. There is an A/A-bar opposition for anaphoric pronouns. It stands for <±topic-shift> and it is based on a single opposition along the saliency hierarchy for anaphoric pronouns, see (27). Only the 3rd person masculine pronoun in Dutch, French and Italian is given in the list. The black/grey opposition in (27) indicates which saliency difference has to be selected to express the <±topic-shift> function.

(27) Hierarchy of anaphoric pronominal devices: less salient → most salient

	null <i>pro</i>	clitic pronoun	weak pronoun	strong pronoun	<i>d</i> - pronoun
Dutch			<i>ie</i> (subject) <i>'m</i> (object)	<i>hij</i> (subject) <i>hem</i> (object)	<i>die</i> (subject) <i>die</i> (object)
French		<i>il</i> (subject) <i>le</i> (object)		<i>lui, il</i> (subject) <i>lui, le</i> (object)	
Italian	pro/agr (subject)	<i>lo</i> (object)		<i>lui</i> (subject) <i>lui, lo</i> (object)	

For Dutch, both the strong pronouns (subject *hij, zij, het* ‘he, she, it’ and object *hem, haar, het* ‘him, her, it’) as well as their weak variants (*ie, ze, 't* and *'m, d'r, 't*) fall in the group of <-topic-shift> pronouns, see the example in (17)a repeated here as (28)a.

(28) Het meisje met de gouden haren is ook gefotografeerd
 The girl with the golden locks is also photographed
 “The girl with the golden locks was taken a picture of”
 a. *Zij/ze* (= het meisje met de gouden haren) is ooit model geweest
 She (= the girl with the golden locks) is ever model been
 “She used to be a model”

The grammatically defined <±topic-shift> opposition selected from a general hierarchy scale of pronominal devices is supported by experiments reported in Kaiser and Trueswell (2004). They tested the effects of the Dutch full (feminine singular) pronoun *zij* and the weak (feminine singular) pronoun *ze* in sentence-initial position. Their experiments show that both are equally used for <-topic-shift> to

She includes ‘physically given’ antecedents in her analysis. The present analysis is directed at ‘linguistically given’ antecedents only, as the ones that realize the core property of human language “situation-free, c.q. discourse-bound” (Chomsky 1968).

maintain the topic, they say ‘subject’, of the preceding sentence. Note how in the light of (27) above, they focused an opposition in the grey area for Dutch. The main <±topic-shift> opposition for V2nd Dutch is the personal pronoun versus the *d*-pronoun¹²

6. The acquisition of the referential system

It was argued above that 3rd person pronouns and topic *d*-pronouns are referential elements that may be used anaphorically to refer to a previously mentioned antecedent. Following Postal (1966) we will classify them as D^o elements, like articles.¹³ These characteristics are listed in (29).

- (29) 3rd person pronouns and *d*-pronouns
- a. are referential signs D^o (determiners)
 - b. may have a DP discourse antecedent

Postal’s point of view is confirmed by the graphs for the acquisition of articles and 3rd person pronouns as we will show in (34). The simultaneous acquisition of articles and anaphoric pronouns demonstrates that the real acquisition step is the introduction of a referential system added to argument structure. This view is in line with Williams (1994) who argued for non-acquisitional reasons that there is a close relation between the grammatical theta/case marking of arguments and anaphoric signs for referentiality.

In previous work (Van Kampen 2002, 2004, 2006) it was argued that children start with situation-bound anaphoric reference that is still discourse-free and without reference to previously mentioned antecedents (see also Lyons 1979; Atkinson 1979; Hickmann 1982; among others). Articles and discourse anaphors are lacking in early child language. We make a rough division between two phases of child language. A situation-bound system before D^o-marking, and a situation-free system after D^o-marking. The acquisition of D^o-marking realizes within half a year the introduction of articles, 3rd person clitics and pronouns, and *pro*-drop, at least for the languages considered here.

¹² The weak reduced subject pronouns *ie* and *t* cannot be used sentence-initially in Dutch. Different test sentences, with subject inversion, are therefore needed to test the generalization of the present claim. This holds as well for the object pronouns.

¹³ The long-debated question whether pronouns are only D^os or also DPs, has fortunately evaporated due to a more reduced labeling convention (Chomsky 2000). The D^o/DP labels are use now for exposition only.

(30)

<i>Phase 1: situation-bound</i>	<i>Phase 2: discourse-bound</i>
No topic-maintenance device - all anaphors are gesture-sustained - no {articles, agr, clitics, pronouns, <i>pro</i> }	Topic-maintenance device - anaphors need not be gesture-sustained - {articles, agr, clitics, pronouns, <i>pro</i> }

Language acquisition is a gradual process and D⁰-marking is the crucial turning point here. It may take roughly half a year and a million and a half of short sentences (say 30 weeks of 5000 small sentences a day).

6.1 The acquisition of the referential system in Dutch

We will first have a look at the acquisition of Dutch. For Dutch, we counted the use of referential markings in the speech of Sarah (Van Kampen corpus in CHILDES, MacWhinney 2006).

The most important acquisition steps within the first phase is the marking of illocution by a finite verb in the first or second position (the V2nd position).¹⁴ The first phase is characterized by an abundant use of deictic situation-bound 1st and 2nd person pronouns and demonstratives in both non-finite and finite marked sentences. This seems reasonable. The 1st-2nd person pronouns express a <±speaker> opposition and the demonstratives a <±proximate> opposition, all situation-bound oppositions.

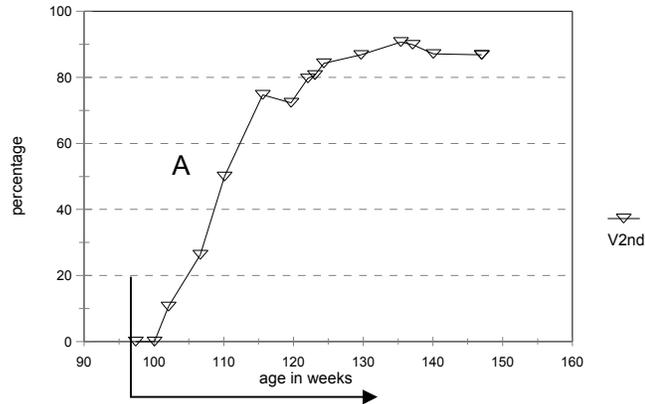
Examples of a non-finite clause with a demonstrative are given in (31).

- (31) a. *deze hebben* (Sarah week 86 / 1;7.21)
that (one) have
“I wanna have that one”
- b. *die niet lachen* (Sarah week 107 / 2;0.17)
that (one) not laugh
“as far as that one is concerned, he is not laughing”

The primary selection of the demonstrative was earlier observed by Haegeman (1996) for the Dutch child Hein. The graph in (32) represents the acquisition of the finite verb in first/second position (graph from Evers and Van Kampen 2001). We take it that the child has acquired systematic marking when she realizes > 80% of the adult norm.

¹⁴ For proposals that V_{fin}-to-C is connected to the illocutionary force of a clause, see Evers (1981), Wechsler (1991), Van Kampen (1997), Gärtner (2002), among others.

(32) Sarah (Van Kampen corpus, CHILDES)



Situation-bound demonstratives (present from the very beginning on)
Graph A: the rise of finite verbs in V2nd position

In the recordings till week 120 (7 recordings between 1;10.13-2;3.16), Sarah did not use any anaphor to mark a reference to the linguistic discourse. There was hardly any use of 3rd person pronouns (A-anaphors), and there was no use of A-bar *d*-pronouns referring to a linguistic discourse antecedent. The referent of the demonstrative was always present in the immediate speech situation. In the sentence with a finite predicate, we counted 50 examples of contrastive demonstratives. Presentationals were excluded from the count. All 50 examples were related to a referent in the situation. Examples of such demonstratives (referring to a referent in the situation) are given in (33).

(33) Anaphoric pronouns: gesture-sustained

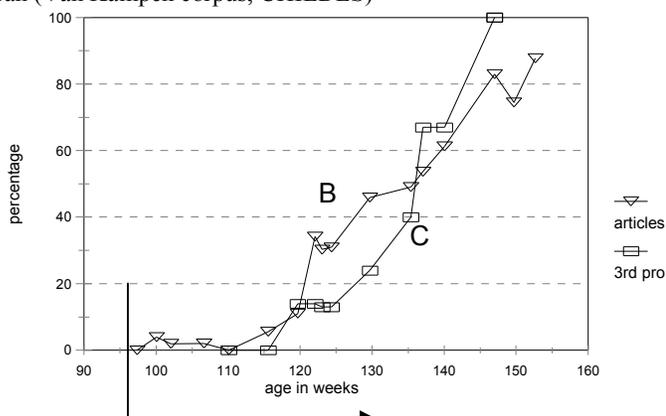
- a. (playing Memory; one card doesn't match) (week 107 / 2;0.17)
 Sarah: *die* kan niet mee(r).
 "that cannot anymore"
- b. (looking at a picture) (week 116 / 2;1.10)
 Sarah: oehoe, uilen op het dak.
 "oehoe, owls on the roof"
 mother: ja, twee uilen op het dak.
 "yes, two owls on the roof"
 Sarah: *deze* hoefe niet op (h)et dak.
 "these need not (go) on the roof"

Early child language uses the <+topic-shift> form *die* abundantly. Because there is no linguistic context yet, each sentence in child language names its own topic, as if it were a first mention.

It is only in the second phase, after week 120, that Sarah starts using articles before nouns, graph B, with some regularity. The graph for 3rd person pronouns

(graph C) shows the growing reliability of the child on discourse anaphors. The parallel acquisition graphs (same period, same speed) in (34) indicate that indeed the acquisition of discourse anaphors and articles are closely related.¹⁵ It constitutes a striking support for the claim that D⁰-marking is a matter of argument identification, rather than some noun-extension, as argued for in Williams (1994).

(34) Sarah (Van Kampen corpus, CHILDES)



Situation-bound demonstratives (present from the very beginning on)

Graph B: articles before nouns

Graph C: A-anaphors (3rd p. pronouns *hij/zij/het* 'he/she/it' and 'm/d'r/'t)

It may be deduced from the graphs in (34) that A-anaphors are acquired simultaneously with articles, and that the use of demonstratives runs ahead of both. The rise in the use of articles and the parallel rise in the use of 3rd person pronouns take place between week 120 and week 145. A qualitative study of the Sarah files shows that the 3rd person pronouns are indeed used as A-anaphors to indicate the maintenance of a topic that was linguistically introduced in the preceding sentence. At the same time, the <-proximate> demonstratives *die* and *dat* used previously for situation-bound reference, are now also applied as A-bar *d*-pronouns that indicate a topic-shift w.r.t. the preceding sentence.¹⁶

¹⁵ From Van Kampen (2004). Each point in graph C represents the ratio of 3rd person pronouns w.r.t. nouns ($DP<+pronoun>/DP<+/-pronoun>$) in the speech of Sarah, measured as a percentage of the ratio $DP<+pronoun>/DP<+/-pronoun>$ in the speech of the mother within the same file. Graph B represents Sarah's systematic use of articles (and other D⁰ elements) before nouns, the ratio $<+D[-NP]>/<+/-D>[-NP]$. In Dutch, the use of a D⁰ is obligatory with singular count nouns and definite plural nouns and only the +/- oppositions in these contexts were counted.

¹⁶ Psycholinguistic experiments as well as conventional recordings have the disadvantage to invite situation-bound utterances. The crucial point here is that child language after the

Examples of discourse-bound 3rd person pronouns for <-topic-shift> and *d*-pronouns for <+topic-shift> are given in (35) and (36).

(35) (talking about a bird in a picture-book) (week 125/2;4.27)

mother: ja, hij heeft de schaar, de vogel.

“yes, he has the scissors, the bird”

Sarah: schaar ["] vogel ["]. teen! *hij* heb een teen, he.

“scissors ["] bird ["]. toe! he has a toe, isn't it.”

(36) (shifting the attention to a picture at a jigsaw puzzle) (week 133/2;6.18)

mother: dan past die (=stukje) misschien daar?

“then that (piece) fits there?”

Sarah: *die* is voor pappa, die hondje

“that is for daddy, that doggie”

The <+topic-shift> *d*-pronouns in Dutch appear in the position before the finite verb (Spec,C). The finite verb in the second position had been learned before.

Romance French and Italian use, respectively, weak 3rd person pronouns and *pro*-drop for <-topic-shift> and strong person pronouns in adjunct positions for <+topic-shift>. One may wonder how the acquisition of this system relates to the acquisition of articles and verbal agreement. In the next section we will look at the acquisition steps in non-V2nd Romance French.

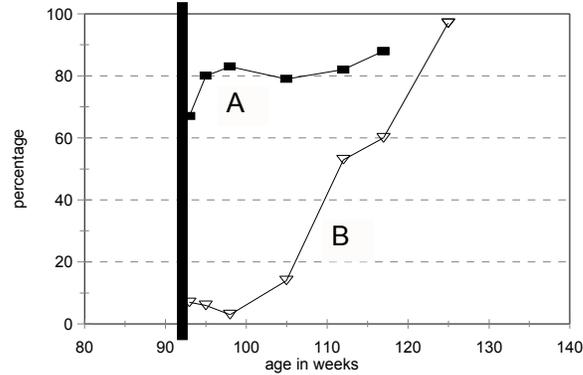
6.2 The acquisition of the referential system in French

Dutch children figure out the predicational I⁰ with its V2nd rule, before they acquire D⁰-marking. This was shown for Sarah by the graphs in (32) and (34). The acquisition of the verbal system {theta-frames and <±finite> paradigms} functions as a prerequisite for the acquisition of the referential D⁰-system.

French children have a more easy access to the I⁰-system and their acquisition of the D⁰-system seems to take place several months earlier than in Dutch. The most important point, though, is that in French as well as in Dutch the I⁰/C⁰-marking of illocutions crucially precedes the referential D⁰-marking, which is a matter of (linguistic) discourse structure (Van Kampen 2004). For French, we counted the use of referential markings in the speech of Grégoire (Champaud corpus in CHILDES, MacWhinney 2006). The French articles are used systematically after week 120 (graphs from Van Kampen 2004).

acquisition of D⁰-marking may and also do from time to time refer to a linguistic antecedent not present in the situation.

(37) Grégoire (Champaud corpus, CHILDES)



Graph A: the rise of finite verbs

Graph B: the rise of determiners

(The black vertical marks the starting point of Champaud corpus)

French children acquire articles earlier than Dutch children. Between week 93 and week 120 there is a rise in the graph for determiners in (37). In that period (7 recordings between 1;9.18-2;3), Grégoire did not use any referential marking to indicate an anaphor in the linguistic situation. A gesture-sustainable antecedent was always present. In the sentences with a finite verb, we counted 58 examples of dislocated nouns doubled with a clitic to indicate a referent in the situation.¹⁷ Grégoire also used (gesture-sustained) contrastive demonstratives related to a referent in the situation. Again, presentationals were excluded from the count. There were only a few 3rd person single clitics (A-anaphors), and there was no use of A-bar anaphors (dislocated pronoun doubled by a clitic) referring to a linguistic discourse antecedent. That is clearly a different acquisition step.

Examples of the use of dislocated nouns doubled by a clitic referring to a referent in the situation are given in (38). The dislocations sometimes are to the left as in (38)a, but most of the time they are to the right, as in (38)b (Van Kampen 2002, 2004; Van der Linden and Sleeman 2007). The preference of right-dislocations seems an effect of the presence of a situation-bound gesture-sustainable referent.

(38) Anaphors: gesture-sustained

a. (looking at a picture in a book)

(Grégoire 1;9.28/week 95)

crocodile, il mange
“crocodile, he eats”

¹⁷ See for the early use of dislocated nouns in French child language De Cat (2002). De Cat argues that this shows an early discourse competence in children. This is quite the opposite of what we claim.

- b. (holding a car) (Grégoire 1;11.22/week 103)
elle roule, *la* voiture
 “she goes, the car”
- c. *celle-là*, *elle* est petite (Grégoire 2;1.25/week 112)
 “that one, she is small”
- d. c’est é canard et *ça* roule (Grégoire 2;3/week 117)
 “it’s a duck and it goes”

Since there is a situational context only, each sentence in the language of the child names its own topic. The same type of evidence comes from elicited narratives with picture sequences in a study by Hickmann and Hendriks (1999). They report that in this context, French children up to the age of seven use dislocated nouns doubled by a clitic (*il... le chien*, ‘he ... the dog’/*le chien, il* ‘the dog, he’) for the first mentions of a new discourse topic. It shows that even older children may heavily rely on the situational context when pictures are involved. In the adult language, a newly introduced discourse topic can, in general, not be referred to by a definite description (lexical definite DP or pronoun/clitic).

As stated in the previous section, the rise of articles indicates the growing use of nouns as referential arguments in linguistic discourse. The rise of articles is narrowly related to the rise of anaphors. These are in Dutch the rise of 3rd person pronouns and in French the appearance of clitics. See the table in (39).

(39) French Grégoire: anaphoric subject clitics for topic-maintenance

age in weeks	a. determiners	b. dislocated noun + clitic (in % w.r.t single clitic)		c. single subject clitic		d. single object clitic <i>le/la</i>
				<i>il</i>	<i>elle</i>	
93	7%	8	89%	0	1	0
94	6%	7	78%	0	2	0
98	3%	7	78%	0	2	0
105	14%	19	95%	0	0	0
112	53%	3	---	0	0	0
117	60%	8	61%	2	4	0
125	97%	11	37%	19	0	9
127-129	100%	51	35%	66	28	10

The gray area in (39) indicates that the acquisition point of articles (>80%) (39)a is simultaneous with a sudden rise of single (non-doubled) subject clitics (39)c and object clitics (39)d in the speech of French Grégoire. This sudden rise of single clitics can be characterized as the acquisition of discourse structure reflected by topic-maintenance. Unlike the pronouns in Dutch, French clitics do not appear simultaneously, but right after the determiners. This is probably, because clitics imply the acquisition of a different argument placement in addition to the argument pronominalization.

There are a few instances of a single clitic in the speech of Grégoire before week 120, for example the one in (40).

- (40) playing the child's hand sticks to the investigators ear) (1;10.20/week 94)
 investigator: tu cognes ? "you bump against?"
 Grégoire: elle colle "it (=the hand) sticks"

In such instances, a gesture accompanies the clitic (Van Kampen 2002). It is the gesture that brings in focus the intended referent, not the unstressed clitic. The gesture directs the hearer's attention towards an object present in the utterance situation (Kleiber 1994: chapter 5). See also Tedeschi (this volume) for child Italian.

Examples of the discourse-bound devices for <±topic-shift> that appear after week 120 are given in (41).

- (41) Anaphors: discourse-bound
 Topic-shift versus topic-maintenance
 (inventing a story) (2;5.27/week 129)
 Grégoire: maman, elle m'a protégé pour écraser la jeep.
 "mummy, she has protected me from (being) crashed by the jeep"
 Grégoire: *la jeep, elle a écrasé ma maman.*
 "the jeep, she has crashed my mummy"
 investigator: mais qu'est ce qu'elle faisait cette jeep au bord de la mer?
 "but what did that jeep do at the see?"
 Grégoire: *elle a roulé sur la mer.*
 "she has gone on the see"

As in the example above, later child French as well as adult French show a preference for left-dislocations (Givon 1983; Ashby 1988: 206). This shift in preference, from right-dislocated topics in early child French to left-dislocated topics in later child French, reflects a growing reliance on linguistic discourse reference by means of sentential topics. Discourse reference tracking by a topic in Spec,C or in sentence adjunct position must get scope over the new sentence. This may explain its appearance at the left periphery of the sentence.

7. Conclusion

Beside the identification of arguments within the sentence by order restrictions and case marking, there is an identification of arguments within discourse. West-European languages use articles to distinguish arguments as <±previously mentioned>. Besides articles there are personal pronouns that are to be indexed, c.q. identified, with an antecedent. Superimposed on that system there are additional devices to indicate whether a clause has the same or a different element as its 'topic'. Topic is an argument the sentence is 'about'. It is not necessarily the subject. If a sentence takes a topic different of the topic of the preceding sentence, there are devices to mark the sentence as <+topic-shift>. These <+topic-shift> devices vary with the type of language. The Germanic V2nd languages use a demonstrative pronoun in the Spec,C position (an A-bar pronoun). This *d*-pronoun refers to an argument in the preceding sentence that had a focus-kind of prominence. Neither the

topic referred to by the A-bar *d*-pronoun, nor the preceding argument with focus prominence needs to have a <+subject> or a <-subject> status, see section 3. The same independence from sentence-internal functions such as <±subject> holds for Romance. The Romance languages mark the <+topic-shift> by a dislocated argument supported by a sentence-internal clitic, see section 4. The dislocated argument may have the status of a strong personal pronoun.

The acquisition of <±topic-shift> devices takes place more or less simultaneously with the acquisition of other pronominal devices. All these devices make the language situation-free and discourse-bound. The switch from the situation-bound early child language to the later situation-free child language does not take place before the discourse units, the successive sentences, have acquired an internal coherence due to argument theta-frames of the denotational verb and the opposition between <±finite> verb. There is a crucial acquisition order. Sentence-internal I⁰-marking for situation-bound early child language precedes discourse-oriented D⁰-marking for the later situation-free child language. This fundamental acquisition order has already pointed out in Van Kampen (2002, 2004)

There is a common point in the acquisition of the <±topic-shift> devices. Both Germanic V2nd and Romance child language start with sentences marked by situation-bound device for <+topic-shift>. In the beginning, each utterance in the language of the child stands on its own and establishes its own topic. Later on, the child's speech enters the linguistic discourse of an actual or presupposed continuing discourse.

A second common point of the <±topic-shift> devices in Germanic and Romance is the close connection with D⁰-marking, the acquisition of articles and their like. The West-European D⁰-marking of arguments is clearly a matter of discourse orientation as it requires the distinction between <+definite>, i.e. previously mentioned, and <-definite>, i.e. newly introduced. Whereas simple naming by proper names and gesture-sustained deixis can be used in situation-bound language use, D⁰-marking is different. It requires an explicit or implicit discourse structure. Its base-point is reference as identification within a linguistic context. This is more than some plausibility view. The longitudinal acquisition curve for articles coincides with the acquisition curve for 3rd person pronouns in Germanic V2nd (Dutch). The coincidence of these two curves supports the claim that the basic acquisition procedure that we see here is the acquisition of discourse reference. The indexing system, so to speak, that is used by logicians and linguists alike. As argued in Van Kampen (2002, 2004) and Avram and Coene (this volume), the 1st and 2nd person pronouns/clitics have to be kept out of the acquisition graphs as they are situation-bound (speaker, hearer). The longitudinal graphs show that we are on the right track. Of course, graphs of more children acquiring different Germanic V2nd languages are needed.

The Romance personal pronouns appear in principles as clitics. They have a specific distribution of their own, different from the corresponding fully spelled out arguments. The basic factor in their learnability must have been the presence of the argument structure associated with the denotational verb. Note that table (39) shows no difference between the acquisition of subject and object clitics (pace Hamann et

al. 1996). The simultaneous acquisition of subject and object clitics supports the idea that the underlying condition of this acquisition step is the presence of the argument theta-frame of the verb.

Psycholinguistic and quantitative studies have often observed that people tend to use an unstressed 3rd person pronoun (Germanic) or clitic/*pro* (Romance) to refer to the subject of the preceding clause. We have argued that these observations are correct, but irrelevant. It is not to deny of course that other types of languages employ a grammaticalized relation between subject status and topic-hood, for example Sesotho (a Bantu language). The subject in Sesotho must be at the same time the definite presumed old information topic. The Sesotho child is up to the challenge and introduces all the subject-changing transformations required by discourse (Demuth 1989). There, as well as here, it turns out that the real acquisition step is not the transformational variance in distributions, but reference-tracking in discourse.

References

- Ariel, M. (1990) *Accessing Noun-Phrase Antecedents*. London/New York: Routledge.
- Ariel, M. (2001) 'Accessibility Theory: an overview', in T. Sanders, J. Schilperoord and W. Spooren (eds.) *Text Representation: Linguistic and Psycholinguistic Aspects*. Amsterdam/ Philadelphia: John Benjamins, 29-89.
- Ashby, W.J. (1988) 'The syntax, pragmatics, and sociolinguistics of left- and right-dislocations in French'. *Lingua* 75, 203-229.
- Atkinson, M. (1979) 'Prerequisites for reference', in: E. Ochs and B.B. Schieffelin (eds.), *Developmental Pragmatics*. London: Academic Press, 229-249.
- Avram, L. and M. Coene (this volume) 'Object clitics as last resort. Implications for language acquisition'.
- Bosch, P., G. Katz and C. Umbach (2007) 'The non-subject bias of German demonstrative pronouns', in: M. Schwarz-Friesel, M. Consten and M. Knees (eds.) *Anaphors in Texts*.
- Carminati, M.N. (2002) *The Processing of Italian Subject Pronouns*. PhD Dissertation, University of Massachusetts, Amherst.
- Cat, C. de (2002) *French Dislocation*. PhD dissertation, York University.
- Chomsky, N. (1968) *Language and Mind*. New York: Harcourt, Brace and World.
- Chomsky, N. (1977) 'On wh-movement', in: P. Culicover, T. Wasow and A. Akmajian (eds.) *Formal Syntax*. New York: Academic Press, 71-133.
- Chomsky, N. (2000) 'Minimalist inquiries: the framework', in: R. Martin, D. Michaels, J. Uriagereka (eds.) *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik*. Cambridge: MIT Press, 89-155.
- Cinque, G. ((1990) *Types of A-bar-Dependencies*. Cambridge Mass: MIT Press.
- Cinque, G. (1993) 'A null theory of phrase and compound stress'. *Linguistic Inquiry* 24, 239-297.
- Comrie, B. (2000) 'Pragmatic binding: Demonstratives as anaphors in Dutch', in: M.L. Juge and J.L. Moxley (eds.) *Proceedings of the 23rd Annual Meeting of the Berkeley Linguistics Society*. Berkeley: B.L.S., 50-61.
- Diessel, H. (1999) *Demonstratives. Form, function, and grammaticalization*. Amsterdam/ Philadelphia: John Benjamins.
- Evers, A. (1981) 'Two functional principles for "Move V"'. *Groninger Arbeiten zur Germanistischen Linguistik* 19, 96-110.

- Evers, A. (2003) 'Verbal clusters and cluster creepers', in: P.A.M. Seuren and G. Kempen (eds.) *Verb Constructions in German and Dutch*. Amsterdam/ Philadelphia: John Benjamins, 43-91.
- Evers, A. and J. van Kampen (2001) 'E-language, I-language and the order of parameter setting'. *UiL OTS Working Papers* 00105-S-S.
- Gärtner, H.-M. (2002). On the force of V2 declaratives. *Theoretical Linguistics* 28. 33–42.
- Givón, T. (1983) 'Topic continuity in discourse: the functional domain of switch reference', in: J. Haiman and P. Munro (eds.) *Typological Studies in Language, Vol. 2: Switch Reference and Universal Grammar*. Amsterdam/Philadelphia: John Benjamins, 150-280.
- Grimshaw, J. (1995) 'Optionality and optimality'. Paper presented at the *Workshop on Optionality*, Utrecht University, September 1-2.
- Grimshaw, J. and V. Samek-Lodovici (1998) 'Optimal subjects and subject universals', in: P. Barbosa, D. Fox, P. Hagstrom, M. McGinnis and D. Pesetsky (eds.) *Is the Best Good Enough? Optimality and competition in syntax*. Cambridge Mass: MIT Press and MIT WPL, 193–219.
- Gundel, J.K., N. Hedberg, and R. Zacharski (1993) 'Cognitive Status and the form of referring expressions in discourse'. *Language* 69, 274-307.
- Hamann, C., L. Rizzi and U. H. Frauenfelder (1996) 'On the acquisition of subject and object clitics in French', in: H. Clahsen (ed.) *Generative Perspectives on Language Acquisition* Amsterdam/ Philadelphia: John Benjamins, 309-334.
- Haegeman, L. (1996) 'Root Infinitives, Clitics and Truncated Structures', in: H. Clahsen (ed.) *Generative Perspectives on Language Acquisition*. Amsterdam/Philadelphia: John Benjamins, 271-307.
- Hickmann, M. (1982) *The Development of Narrative Skills: Pragmatic and meta-pragmatic aspects of discourse cohesion*. PhD. dissertation University of Chicago.
- Hickmann, M. and H. Hendriks (1999) 'Cohesion and anaphora in children's narratives'. *Journal of Child Language* 26, 419-452.
- Holmberg, A. (1986) *Word Order and Syntactic Features in the Scandinavian Languages and English*, PhD dissertation, University of Stockholm.
- Kaiser, E. and J. Trueswell (2004) 'The referential properties of Dutch pronouns and demonstratives: Is salience enough?' *Proceedings of Sinn und Bedeutung* 8, Arbeitspapier Nr. 1777, FB Sprachwissenschaft, Universität Konstanz.
- Kampen, J. van (1997) *First Steps in Wh-movement*. Delft: Eburon.
- Kampen, J. van (2002) 'Learnability order and discourse function in the acquisition of the French pronominal system'. Paper presented at *Going Romance*, Groningen, November 29.
- Kampen, J. van (2004) 'Learnability order in the French pronominal system', in: R. Bok-Bennema, B. Hollebrandse, B. Kampers-Manhe and P. Sleeman (eds.) *Selected Papers from Going Romance 2002*. Amsterdam/Philadelphia: John Benjamins, 163-183.
- Kampen, J. van (2006a) 'Early operators and late topic-drop/pro-drop', in: V. Torrens and L. Escobar *The Acquisition of Syntax in Romance Languages*. Amsterdam/Philadelphia: John Benjamins, 203-223.
- Kampen, J. van (2006b) 'Lexical growth and grammatical competence: potential risks for children with a hearing deficiency'. *Language Acquisition and Development: Proceedings of GALA2005* Cambridge: Cambridge Scholars Press/CSP, 297-304
- Kampen, J. van (2007) 'Relative agreement in Dutch', in: M. van Koppen and B. Los (eds.) *Linguistics in the Netherlands 2007*. Amsterdam: Benjamins, 112-124.

- Kühner, R and C. Stegmann (1912) *Ausführliche Grammatik der lateinischen Sprache II* Hannover: Hahn.
- Lambrecht, K. (1981) *Topic, Antitopic and Verb Agreement in Non-standard French*. Amsterdam/Philadelphia: John Benjamins.
- Linden, E. van der and P. Sleeman (2007) 'Clitic dislocation', in: M. van Koppen and B. Los (eds.) *Linguistics in the Netherlands 2007*. Amsterdam: Benjamins, 173-186.
- Lyons, J. (1979) 'Deixis and anaphora', in: T. Myers (ed.) *The Development of Conversation and Discourse*. Edinburgh: University Press, 88-103.
- MacWhinney, B. (2006) *The CHILDES Project: Tools for Analyzing Talk*, 3rd edition. Mahwah New Jersey: Lawrence Erlbaum Associates.
- Mikkelsen, L. (2005) 'Subject choice in copular clauses', ms. UC Berkeley.
- Mörnsjö, M. (2002) *V1 Declaratives in Spoken Swedish*. PhD dissertation, Lund University.
- Postal, P. (1966) 'On So-called Pronouns in English', in: F. P. Dinneen (ed.) *Monograph Series on Language and Linguistics 19*. Georgetown: Georgetown University Press, 177-206.
- Reinhart, T. (1981) 'Pragmatics and linguistics: An analysis of sentence topics'. *Philosophica* 27, 53-94.
- Rizzi, L. (1991) 'Proper head government and the definition of A-positions'. *GLOW Newsletter* 26. 46-47.
- Rozendaal, M. (this volume) 'The acquisition of the morphosyntax-pragmatic interface in French L1: Evidence from reference with articles and pronouns'.
- Serratrice, L. (this volume) 'Null and overt subjects at the syntax-discourse interface: Evidence from monolingual and bilingual acquisition'.
- Tedeschi, R. (this volume) 'Referring expressions in early Italian. A study on the use of lexical objects, pronouns and null objects in Italian pre-school children'.
- Von Stechow, A. von (1991) 'Focusing and backgrounding operators', in: W. Abraham (ed.) *Discourse Particles* Amsterdam/Philadelphia: John Benjamins, 24-26.
- Wechsler, S. (1991) 'Verb second and illocutionary force', in: K. Leffel and D. Bouchard (eds.) *Views on Phrase Structure* Dordrecht: Kluwer, 177-191.
- Williams, E. (1994) *Thematic Structure in Syntax*. Cambridge Mass.: MIT Press.
- Zifonun, G., L. Hoffmann, and B. Strecker (1997) *Grammatik der Deutschen Sprache*. Berlin: De Gruyter.

Modality, Non-finite Forms, and the Manifestation of the RI Stage in Null and Non-Null Subject Languages

Juana M. Licerias¹, Susana Perales² and Aurora Bel³

¹University of Ottawa, ²University of Cantabria and ³Universitat Pompeu Fabra

1. Introduction

One of the aspects of child language that has attracted the attention of numerous researchers concerns the production of non-finite forms in contexts where a finite form would be required, i.e., what have come to be known as root infinitives (RIs). Depending on the language, these non-finite forms may be actual infinitives (as in French, German or Dutch) or bare forms with no tense or agreement morphology (as in English).

Also, evidence for the existence of a RI stage has not been found in all languages. In fact, there seems to be a notable difference among languages depending on whether they allow null subjects or not. Thus, non-null subject languages are languages where a robust RI stage has been attested, whereas null subject languages do not appear to exhibit such a noticeable RI period (Guasti 1994; Rizzi 1993/1994).

In this paper, we analyze the presence and interpretation of RIs in null subject languages and propose an explanation of this phenomenon based on the combination of two features: person [+/-P] and infinitive [+/-R]. We will further argue that the way these two features are implemented in the grammar of a given language can be used to predict the duration of the RI across languages.

2. The RI phenomenon across languages: Cross-linguistic differences

It is well known that there are differences in the amount of RIs found in child language data depending on whether a language allows null subjects or not. Thus, children acquiring null subject languages like Spanish, Italian or Catalan are said not to go through an RI stage (Guasti 1994; Rizzi 1993/1994), whereas children acquiring non-null subject languages like Dutch, German or French usually exhibit a robust RI stage.

One of the first attempts to deal with the appearance of RIs in child language is Radford's (1990) maturation hypothesis, which states that children do not have access to functional categories at the beginning of the acquisition process and so they start out with simply a verbal phrase (VP), as in (1). The remaining functional categories are supposed to become available in the course of maturation. The problem with this proposal lies in that evidence of absence of functional categories is very difficult to find in languages with rich verbal morphology, which raises the question of why functional categories are available earlier in some languages than in others.

(1) [VP ...V...]]]]]

A different account is found in Rizzi (1993/1994), who proposes that there is a principle of Universal Grammar (UG) that is responsible for the fact that all sentences include a complementizer phrase (CP), even if in some sentences this CP is phonetically empty. Nonetheless, this principle is subject to maturation in the case of the child. Thus, children can truncate the structure of the clause at any node below the CP. When the child truncates the structure of the clause below the TP node, a non-finite form appears (see (2)), whereas if the structure is truncated above the TP level, a finite tensed form comes in. In the case of the infinitives of null subject languages, truncation below TP is not possible because infinitives need to rise to AgrP to check agreement features, which explains the absence of these forms in the data.

(2) [CP [AgrP [NegP [TP  [VP ...V...]]]]]]

Wexler (1994) advances an alternative explanation. For this author, both Agr(eement) and T(ense) have an uninterpretable D feature that needs to be checked via DP subject rising to the specifier of TP. However, Wexler proposes that children abide by the *Unique Checking Constraint*, which says that only T's or Agr's feature can be checked. This being so, children occasionally omit T from the representation in order to comply with the UCC. When this happens, a RI comes along. The reason why children learning null subject languages are not constrained by the UCC is that in these languages Agr licenses null subjects and therefore it does not have an uninterpretable D feature that requires checking, as in (3).

(3) [CP [AgrP [NegP [TP [VP ...V...]]]]]]

Despite differences in the technical details, Rizzi's and Wexler's proposals share some basic underlying assumptions. Firstly, both claim that crosslinguistic differences in the number of RIs produced by children are motivated by the different parametric properties of infinitives in the respective languages. Secondly, they also share the premise that children are sensitive to these parametric properties from very early on. Finally, Rizzi's and Wexler's proposals share the belief that all functional categories are accessible to the child from the beginning, contra Radford (1990).

Although numerous authors have adhered to the claim that the existence or not of a RI stage is contingent on the (un)availability of null subjects in a given language, there have also been authors who have advocated for the existence of a RI stage in null subject languages. The proposals in this direction can be separated into two groups. On the one hand we have those who claim that there is a true RI stage similar to that of non-null subject languages, but that children learning null subject languages recover faster (Bel 1998, 2001; Licerias, Bel and Perales 2006) and, on the other hand, those who claim that there are analogous RIs forms in null subject languages: bare forms (Hernández-Pina 1984; Tsimpli 1992; Radford and Ploennig-

Pacheco 1995; Davidson and Goldrick 2003; Clahsen, Aveledo and Roca 2003; Buesa 2007; Pratt and Grinstead 2007) or imperatives (Salustri and Hyams 2003).

2.1 RIs in null subject languages

If we look at the production rates of RI forms in the data files from children learning languages like Spanish and Italian and compare the raw numbers with the total obtained in the data from children learning other languages like German or French, we will see that the totals are radically different, with the latter children obtaining far higher scores than the former. However, there are authors who show that if we count RIs file by file, we will find that at some points the number of RIs is also high in children who are learning null subject languages. For instance, Bel (1998, 2001) counted the number of RIs in the production data of Jùlia, a Catalan-speaking child, from the age of 1;10 to 2;5. She found that at 1;10 the percentage of RIs amounted to 21%, whereas at 2;0 there was a significant drop to 3.2%.

Age	% RIs
1;10	21
2;0	3.2
2;1	3.9
2;2	9.8
2;5	5.6

Table 1: RIs in child Catalan (Bel 1998, 2001)

A similar result was found by Licerias, Valenzuela and Díaz (1999), who classified the data from María (a Spanish-speaking child) into two stages: one from the age of 1;0 to 1;9, in which a score of 32.2% was obtained, and another stage from 2;5 to 2;7 in which a low percentage of 2.6% was recorded.

Magín		María	
Age	% RIs	Age	% RIs
1;0 – 1;9	9.8	1;0 – 1;9	32.2
2;5 – 2;7	0.3	2;5 – 2;7	2.6

Table 2: RIs in child Spanish (Licerias, Valenzuela and Díaz 1999)

Bel (2001) also counted the number of RIs for María in a file by file fashion. She found that, up to the age of 1;10, the percentage of RIs was robust enough to be taken into consideration, as shown in the following table:

Age	% RIs
1;7	20.2
1;8	11.4
1;9	8.6
1;10	8.3
1;11	2
2;0	5.4
2;1	2.2

Table 3: RIs in child Spanish (Bel 2001)

These results suggest that looking at the total percentage of RIs produced by a single child may be misleading, because we may be missing stages in which the production of RIs in null subject languages is worth noting. It must be borne in mind that the RI stage in non-null subject languages is taken to last till 2;7 or even 3;0 years old.

A similar result is obtained in Basque, another null subject language. Ezeizabarrena (2002) provided the distribution of RIs in four different stages. Looking at her data, it becomes clear that these children continue to produce RIs after the age of two, but no further than the age of three.

Mikel		Jurgi		Oitz	
Age	% RIs	Age	% RIs	Age	% RIs
1;7 – 1;11	31.6	1;11 – 2;7	20.5	1;6 – 2;2	38.6
2;0 – 2;3	15.2	2;8 – 3	10	2;3	22.5
2;4 – 2;9	2.6	3;1 – 3;3	3.1	2;4 – 2;6	13.1
2;10 – 4	3.2	3;4 – 4;1	1.9	2;7 – 3	0.5

Table 4: RIs in child Basque (Ezeizabarrena 2002, 2003)

The data from Mikel and Oitz (first and second rows of data) show that RIs reach their maximum percentages around the second year of age. From then on, the percentages decrease, as shown in table 4 (see third row of data for Mikel and Oitz, and second row for Jurgi).

Schaeffer and Ben Shalom (2004) report similar results for Hebrew. They challenge Rhee and Wexler's (1995) study and reanalyze the data to show that Hebrew-speaking children also go through a RI stage, albeit shorter than that of children learning other non-null subject languages.

Age	% RIs
1;0 – 1;11 (12 children)	27
2;0 – 3;3 (24 children)	5

Table 5: RIs in child Hebrew (Schaeffer and Ben Shalom 2004)

As can be observed, these authors divide the children’s production data into two stages: one before age two, in which the percentage of RIs is rather high, and another one from age two onwards, in which the percentage of RIs is very low. These authors conclude that Hebrew-speaking children do go through a RI stage, but they recover faster than children learning non-null subject languages.

For Russian, we have looked at the data gathered by Bar-Shalom and Snyder (1997), represented in table 6, and Gagarina (2002), displayed in tables 7 and 8 (adapted from Bar-Shalom and Snyder 1997)¹.

Varya		Tanya	
Age	% RIs	Age	% RIs
1;6	17.2	2;5.11	28
1;7	24.3	2;5.24	10
1;9	18.4	2;6.1	10.7
1;10	5.3	2;6.8	8
2;0	3.7	2;7.6	2.6
2;4	2.1	2;8.26	4

Table 6: RIs in child Russian (adapted from Bar-Shalom and Snyder 1997)

The data from Varya show that RIs are quite common up to the age of 1;10, when a significant decrease is observed. In Tanya’s data, RIs are found even at age 2;6, which suggests a longer RI period than that found in Spanish, but similar to the one found in Basque. Interestingly, Basque and Russian verbs share the fact that they do not have a distinct infinitival marker, and that their infinitives encode aspect. We will come back to this later. In the data from Gagarina (2002) a similar trend can be observed; the production of RIs is high, and in some cases it lasts beyond the second year of age, as shown in tables 7 and 8 (adapted from Gagarina 2002).

Liza			
Age	% RIs	Age	% RIs
1;7	0	2;1	9.8
1;8	30.8	2;2	9.2
1;9	27.9	2;3	6.9
1;10	16.3	2;4	6.3
1;11	17.4	2;5	2.6
2;0	9.4		

Table 7: RIs in child Russian

¹ There has been some debate about whether Russian should be considered a null or a non-null subject language. Bar-Shalom and Snyder (1997) claim that since children produce a fair amount of RIs in this language, it should be included among the non-null subject languages. Nonetheless, we argue that null subject languages also exhibit RI effects, and therefore we oppose the direct relation established by these authors between availability of null subjects and production of RIs.

Vanja			
Age	% RIs	Age	% RIs
2;1	26.4	2;4	9.9
2;2	16.3	2;5	3.7
2;3	14.3		

Table 8: RIs in child Russian

The data presented above challenge the generalization that null subject languages do not exhibit RI effects. In fact, the languages reviewed in this section also obtain high percentages of RIs during a certain period, a fact that is obscured if we collapse the data into one single percentage.

2.2 RI analogues in null subject languages

Apart from the authors who claim for the existence of a RI period in which the production of RIs is similar to that of children learning non-null subject languages, there are authors who have proposed that a similar phenomenon is manifested in null subject languages. Hoekstra and Hyams (1998), after analyzing acquisition data from different languages, proposed that the relation between tense and discourse is encoded by different elements across languages: Number morphology in Dutch and English, Person morphology in Spanish and Italian, and Tense morphology in languages like Japanese. In the case of null subject languages, the underspecification of the feature Person brings about the *Avoid Plural Phenomenon*, which alludes to the inexistence of plural verbal forms during this stage of development.

Other authors (Hernández-Pina 1984; Tsimpli 1992; Radford and Ploennig-Pacheco 1995; Davidson and Goldrick 2003; Clahsen, Avelledo and Roca 2003; Buesa 2007; Pratt and Grinstead 2007) have adhered to the proposal that a bare form (the third person singular in the present tense) is the analogous RI form in null subject languages. They claim this is so because these forms are frequent in child language data, and because there is a high number of agreement errors involving this form. They also argue that this is a basic morphological form, with no suffixes, just the stem of the verb followed by the thematic vowel or ‘word marker’ (Harris 1991). Pratt and Grinstead cite examples like the ones in (4), which are considered as agreement errors because the child is using a third person verbal form to refer to herself.

- (4) a. No puede (Eduardo 2;5.29)
not can-3rd sg
‘He can’t’
[responds to the investigator’s question of whether he can put two pieces of a puzzle together]
- b. No quiere (Graciela 2;3.4)
not wants-3rd sg
‘He doesn’t want’ [responds to mother asking her if she wants a band-aid]

- c. Sí puede nadar (Carlos 2;9.15)
 yes can-3rd sg swim
 “Yes, he can swim”
 [responds to investigator asking if he can swim]

More recently, Salustri and Hyams (2003, 2006) have claimed that the imperative should be taken as the RI analogue for null subject languages because (i) they share with RIs “the mapping of *irrealis* mood onto a tenseless clausal structure” (Salustri and Hyams 2003: 693); (ii) they occur more often in child data than in adult data; (iii) they occur more often in null subject languages than in non-null subject languages, and (iv) they are restricted to eventive predicates. Crucially, the imperative form in these languages is homophonous with the third person singular form advocated for by Tsimpli (1992), Pratt and Grinstead (2007) and Buesa (2007), among others.

Bel (1998, 2001) and Bel, Liceras and Perales (2006) argue against considering that the third person singular is an RI analogue. They claim that for that to be the case, we should find that the third person singular form has different tense values. In the data, third person singular forms always have a temporal present value, whereas true RIs have both present and non-present value. Moreover, if the third person singular is used as a default form, we should find that it is used in contexts that do not refer to a third person subject, which is contrary to fact. In fact, utterances like the ones in (4) may be considered as ‘reference errors’ and not ‘agreement errors’ because the child is using a third person to refer to herself, but agreement is right. This is shown by the tendency of the child to refer to herself in the third person singular, as the following examples from Catalan show (Bel and Rosado 2006):

- (5) a. la Júlia vol la nina (Júlia 2;1)
 the Júlia wants the doll
 “Júlia wants the doll”
 b. a Júlia baixa a pantaló (Júlia 2;2)
 the Júlia puts down the trousers
 “Júlia puts down the trousers”

As these examples show, the child (Júlia) uses a third person singular form to refer to herself instead of using a first person but, strictly speaking, subject-verb agreement is essentially correct.

Bel (2001) and Bel, Liceras and Perales (2006) also argue against adopting the imperative as a RI analogue first, because RIs and imperatives do not share the same interpretive properties (RIs may refer to first and third person singular, whereas imperatives always refer to the second person). Secondly, RIs and imperatives do not have the same temporal reference, namely, RIs may refer to various tenses (past, present, future) whereas imperatives always refer to speech time. Furthermore, the fact that imperatives are restricted to eventives is expected since stative imperatives are not available (*know the lesson!). Finally, as Bel notes, arguing that imperatives are RI analogues based on a single percentage of appearance is misleading because

the data show that children continue to produce imperatives well beyond the RI stage.

3. On the semantic interpretation of RIs

It is well known that RIs and inflected forms do not share the same semantic or interpretive properties. In fact, numerous studies have concluded that as far as semantics is concerned, finite and non-finite forms do not alternate freely (Hoekstra and Hyams 1998; Hyams 2001). Specifically, it is the case that finite forms always have a temporal or *realis* interpretation (past or ongoing), whereas non-finite forms (or RIs) have a modal or *irrealis* value (deontic, volitional or future interpretation). Moreover, there are differences in the aspectual value of verbs in that finite forms can be either eventive or stative predicates, whereas RIs are restricted to eventives. These temporal and aspectual observations led Hoekstra and Hyams to propose the following generalizations:

- (6) *Modal Reference Effect (MRE)*: with overwhelming frequency RIs have modal interpretations
- (7) *Eventivity Constraint (EC)*: RIs are restricted to eventive verbs

Hyams (2001) goes one step further and claims that these two generalizations are determined by the child's attempt to establish a more primitive opposition: *realis* mood (actual occurrence, whether past or ongoing of an event) versus *irrealis* mood (desire, necessity of futurity of some event). Thus, according to this author, "the alternation between finite and non-finite forms falls out of the child's attempt to set up a system of semantic oppositions and map them onto morphological forms" (Hyams 2001: 47). According to this hypothesis, the RI stage is universal across languages, what differs is the specific forms children choose to mark the *realis* versus *irrealis* opposition. Table 9 summarizes the forms used by children learning different languages, as proposed by this author and colleagues.

	[+/-NS]	[+/-R]	<i>Realis</i>	<i>Irrealis</i>
Dutch, German	-	+	Inflected	Infinitives
English	-	-	Bare and Inflected	Semi-auxiliaries
Italian, Spanish	+	+	Inflected	Imperatives
Greek	+	-	Inflected	Bare subjunctive/perfective

Table 9: Crosslinguistic expression of the *realis/irrealis* opposition

In languages like Dutch and German, which do not allow null subjects[+/-NS] and have an infinitival form, the latter would be chosen by the child to convey an *irrealis* interpretation. According to Hyams (2001), the choice of the infinitive to express *irrealis* mood is motivated by the intrinsic modal value of infinitival markers. In English, a language without null subjects and without a distinct infinitival marker, both bare and inflected forms would be carriers of *realis* interpretation and semi-auxiliaries (*hafta*, *wanna*, *gonna*) would be used by the child to express the *irrealis* (Ud Deen 1997; Hyams 2001). In Italian and Spanish, null subject languages with distinct infinitival forms, inflected forms would express *realis* meaning and imperatives (bare verbal forms) would express *irrealis* meaning (Salustri and Hyams 2006). Finally, in Greek, a language without infinitives that allows null subjects, a bare form (subjunctive or perfective) would be in charge of the expression of *irrealis* interpretations (Hyams 2001, 2005).

Irrespective of the actual forms that are used in each language, this proposal paves the way to consider that the RI stage is indeed a universal stage of development across children and across languages, and therefore children will vary as to what forms they use to establish that primitive semantic opposition. In the next sub-section we will be looking at the semantic properties of RIs in null subject languages.

3.1 Interpretive properties of RIs in null subject languages

According to the studies that have isolated true RIs in children learning null subject languages, those non-finite forms are used to convey both *realis* and *irrealis* meanings. The following are examples from the languages discussed in 2.1:

- (8) *Catalan*
- | | | | | |
|----|--------------|-------------------|-----------------------|--------------|
| a. | Sortir | | <i>Irrealis</i> value | (Júlia 1;10) |
| | Come out-INF | | | |
| b. | Aixó | recollir, mama | <i>Realis</i> value | (Júlia 2;1) |
| | This | pick up-INF mummy | | |
- (9) *Spanish*
- | | | | | |
|----|--------------|--------------|-----------------------|-------------|
| a. | Sentar | | <i>Irrealis</i> value | (María 1;8) |
| | Sit down-INF | | | |
| b. | El otro | buscar | <i>Realis</i> value | (María 1;8) |
| | The other | look for-INF | | |
- (10) *Basque*
- | | | | | |
|----|----------|--------------|-----------------------|-------------|
| a. | Hori | amatau | <i>Irrealis</i> value | (Mikel 2;1) |
| | That | turn off-INF | | |
| b. | Hemendik | pasa | <i>Realis</i> value | (Jurgi 2;7) |
| | Here | pass-INF | | |

(11) *Hebrew*

- | | | | | | |
|----|----------|---------|------------|-----------------------|-----------|
| a. | Lashevet | al | ha-shulxan | <i>Irrealis</i> value | (AD 2;1) |
| | Sit-INF | on | the table | | |
| b. | Malon | Lauf | | <i>Realis</i> value | (AM1 2;1) |
| | Balloon | fly-INF | | | |

(12) *Russian*

- | | | | | | |
|----|---------|--------------|--|-----------------------|---------------|
| a. | Rubasku | snimat | | <i>Irrealis</i> value | (Sasha J 2;4) |
| | Shirt | take off-INF | | | |
| b. | Tetya | pet | | <i>Realis</i> value | (Sasha P 1;9) |
| | Woman | sing-INF | | | |

These examples show that RIs in null subject languages do not conform to the *realis/irrealis* opposition predicted by Hyams, since they allow both interpretations. A possible explanation for this lies in the fact that *irrealis* forms in Romance languages are also morphologically salient (subjunctive) and therefore we expect that they are incorporated soon into the child's grammar. On the contrary, for languages that do not have a specific morphology for *irrealis* mood, the child needs to resort to the use of infinitives until modal verbs are integrated. This would suggest that this semantic opposition, rather than a primitive, would be the reflex of how children attempt to sort out the morphological intricacies of their language.

4. The role of the features Person [P] and Infinitive [R]

4.1 The [P] feature and the duration of the RI stage

Hoekstra and Hyams (1998) proposed that it was the underspecification of the Number feature that was responsible for the appearance of RIs in non-null subject languages. They also proposed that the underspecification of Person (the corresponding feature in null subject languages) was responsible for the *Avoid Plural Phenomenon*. After analyzing data from null subject languages, we claim that the Person feature is not underspecified in these languages and that the presence of this feature does not trigger the appearance of RIs in the child production data, but explains that in some languages the production of RIs is restricted to a shorter period of time. Schaeffer and Ben-Shalom (2004) also advocate that it is the Person feature that is responsible for the quick abandonment of the RI stage in null subject languages. They claim that Person (and also Tense) are “the most transparent bridges between syntax and pragmatics, which facilitates the acquisition of obligatory finiteness” (Schaeffer and Ben-Shalom 2004: 93). Nonetheless, and although we agree with these authors that the Person feature plays a crucial role in the abandonment of the RI stage, we propose a more syntactic explanation in that it is due to the fact that (i) its morphological realization is both salient and pervasive (a different ending for each person) and (ii) its pronominal (interpretable) nature makes children analyze it as one of the vocabulary entries in the numeration. In other words, the input provides a strong morphological trigger that is phonologically

salient and pervasive and the fact that it has referential (pronominal) properties also contributes to its triggering power. Furthermore, as we will argue below, this triggering power of the Person feature is further strengthened by the infinitival feature [R].

4.2 The [R] feature and the *realis/irrealis* opposition

Hoekstra and Hyams (1998) claimed that the infinitival morpheme was responsible for the modal interpretation of RIs in child language. However, later on Hyams (2005: 12) acknowledged, “the relation between infinitival morphology and modality hypothesized by Hoekstra and Hyams (1998) cannot be maintained”. In our proposal, however, the infinitival morphology is not what determines the semantic interpretation of RIs, but it is a facilitator for the abandonment of the RI stage. Thus, children learning languages that have a distinct infinitival morpheme will acquire the distinction finite/non-finite faster than children learning languages in which either the infinitival morpheme is phonologically identical to other verbal forms (e.g. Dutch, French), or it is simply non-existent (e.g. English). This is due to the fact that a distinct, unambiguous and consistent infinitival marker that realizes the [R] feature will constitute a clear trigger for the abandonment of the RI period (Liceras, Bel and Perales 2006)².

4.3 A typological distribution of [P] and [R]

Building on the above considerations about the role of the features Person and Infinitive, we have proposed a typological distribution of how the RI stage will manifest in a given language based on the combination of these two features (see also Bel, Liceras and Perales 2006; Liceras, Bel and Perales 2006; Perales, Liceras and Bel 2005). The idea is that depending on how these two features are morpho-phonologically realized in a given language, the duration of the RI stage will vary accordingly. The factors that must be taken into account are: (i) whether the person feature is realized either on both lexical and auxiliary verbs [ALV] or only on auxiliaries [AV]; and (ii) if the infinitival form in a given language has a specific morphological realization or if it is non-distinct (n.d.) from other forms of the verb. A classification of languages according to these features is shown in table 10:

² One reviewer points out that even though the Dutch infinitive is identical to the (finite) plural form, they behave differently in the syntax (infinitives always in final position) and that this should help Dutch-speaking children to abandon the RI stage earlier. Our proposal places the emphasis on a specific morphological feature (Person), which applies within the word domain. In the absence of such morphological trigger (e.g. Dutch) the child needs to resort to word order to acquire the finite/non-finite distinction. This implies that the child needs to be able to parse longer strings (sentences) in order to figure out how such distinction is implemented, which would explain why it takes longer for Dutch children to abandon the RI stage.

[P]	[R]	RI stage	Languages
+ [ALV]	+	° Short	Catalan, Italian, Spanish RIs with <i>realis</i> and <i>irrealis</i> value
+	+	° Short	Hebrew RIs with <i>realis</i> and <i>irrealis</i> value
+ [ALV]	-	° Short	Greek <i>Irrealis</i> (bare subjunctive) / <i>Realis</i> (finite forms)
+ [AV]	+ n.d.	°° Longer	Basque RIs with <i>realis</i> and <i>irrealis</i> values
+ [ALV]	+ n.d.	°° Longer	Russian RIs with <i>realis</i> and <i>irrealis</i> value
-	+ n.d.	°°° Very long	Dutch, German, French High percentage of <i>irrealis</i> RIs
-	-	°°°° Longest	English RIs with <i>realis</i> and <i>irrealis</i> value

Table 10: A typological distribution of the features [P(erson)] and [R(infinitive)]

As can be seen, the combination of these two features permits us to classify languages according to the duration of the RI stage. In the first place, we find a group of languages for which a short RI stage has been documented (Catalan, Italian, Spanish, Hebrew and Greek). It is common to these languages that the feature Person is implemented in their verbal morphology and, with the exception of Greek, these languages also have a distinct infinitival marker. Thus, we argue that both the morpho-phonological saliency and the semantic (pronominal) value of the feature Person are responsible for the rapid acquisition of finiteness in these languages. This rapidness is also increased by the presence of a unique infinitival morpheme that makes the opposition finite/non-finite even more visible.

Secondly, Basque and Russian have a Person feature but, in the case of Basque, it is only morpho-phonologically perceptible in auxiliary verbs. Also, these two languages do not have an infinitival marker, which obscures somehow the distinction finite/non-finite, thus making it a bit more difficult to acquire such distinction, which results in a longer RI stage.

Finally, we have the languages for which the longest RI stages have been documented (Dutch, German, French and English). In these languages, verbal forms are not marked for Person, and moreover, the infinitive marker is shared with other verbal forms (Dutch, German and French) or lack an infinitive marker altogether (English).

In the following table we have represented the ages of children learning different languages in which the percentage of RIs falls under 15%. As can be observed, the production of RIs in Spanish, Catalan and Hebrew starts to decline earlier than in Basque and Russian.

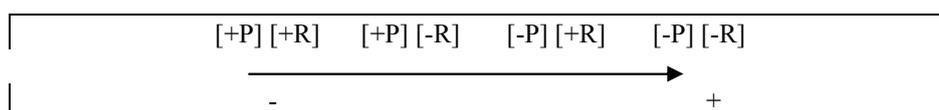
	+ 15%	- 15%
Spanish	1;7	1;8
Catalan	1;10	2;0
Hebrew	1;0 – 1;11	2;0
Basque	1;7 – 2;3	2;4
	1;11 – 2;7	2;8
	1;6 – 2;3	2;4
Russian	1;6 – 1;9	1;10
	2;5	2;6
	1;8 – 1;11	2;0
	2;1 – 2;2	2;3

Table 11: RIs by age

We have taken a percentage of +/- 15% as a tentative landmark for the existence of a RI stage or not. We can see in this table that for children learning Spanish, Catalan and Hebrew the production of RIs decreases around the age of two, whereas for Basque and Russian learners this period extends a bit beyond this age. We have explained this subtle difference by resorting to the morpho-phonological transparency of two features: [P] and [R], which would be responsible for the shorter or longer duration of the RI stage. Further research of languages with the same and other feature combinations should be carried out in order to define more precisely the percentage around which it can be considered that the RI period is over. Also, more languages need to be incorporated to the picture in order to get a clearer understanding of the contribution of each one of the features in the duration and manifestation of the RI stage.

5. The RI stage length continuum

To recap, after analyzing acquisition data from children learning null subject languages, we have confirmed that these children do go through a RI stage, albeit shorter than that of children learning non-null subject languages. Also, RIs in null subject languages do not respect the modal reference effect, as they are also used to convey temporal meanings. Thus, we have proposed that the RI stage is manifested differently across languages depending on two features: [P] and [R]. The way in which these two features are manifested morpho-phonologically in a given language will give rise to a longer/shorter RI stage.



This amounts to saying that the RI stage characterizes a universal stage of development across children and across languages. This stage is no more than the

reflex of the child's attempt to sort out how finiteness is encoded in his/her language. In languages where morphology makes the finite/non-finite distinction transparent (languages with a person feature and with a distinct infinitival marker), the child is able to abandon the RI stage earlier than in languages in which there is no such transparent distinction. As for the specific semantic properties of non-finite forms, we have shown that RIs have *irrealis* as well as *realis* value although some languages (Dutch, German ...) display a different distribution, at least in spontaneous production data. Nonetheless, in order to gain insight into this issue, more experimental evidence would be needed to understand how children convey and comprehend *irrealis* mood during the RI stage.

References

- Bar-Shalom, E. and W. Snyder (1997) 'Optional infinitives and their implications for the pro-drop debate', in: M. Lindseth and S. Franks (eds.) *Formal Approaches to Slavic Linguistics: The Indiana Meeting 1996*. Ann Arbor: Michigan Slavic Publications.
- Bel, A. (1998) *Teoria Lingüística i Adquisició del Llenguatge. Anàlisi Comparada dels Trets Morfològics em Català i en Castellà*. Ph.D. Dissertation, Universitat Autònoma de Barcelona.
- Bel, A. (2001) 'The projection of aspect: A key in the acquisition of finiteness?', in: M. Almgren et al. (eds.) *Research on Child Language Acquisition. Proceedings of the 8th Conference of the International Association for the Study of Child Language*. Somerville, MA: Cascadilla Press, 1297-1313.
- Bel, A., J. Licerias and S. Perales (2006) 'Modalidad e infinitivos independientes en la adquisición de lenguas'. Paper presented at the XXIV International Conference of the Spanish Association for Applied Linguistics, Madrid, April 1.
- Bel, A. and E. Rosado. (2006) 'On some asymmetries in the acquisition of number marking in Catalan and Spanish'. Paper presented at GLOW Workshop on the Acquisition of the Syntax and Semantics of Number, Barcelona, 6-8 April.
- Buesa, C. (2007) 'Early Root Infinitives in a null subject language: A longitudinal case study of a Spanish child', in: A. Belikova, L. Meroni and M. Umeda (eds.) *Proceedings of the 2nd Conference on Generative Approaches to Language Acquisition in North America (GALANA)*. Somerville, MA: Cascadilla Proceedings Project, 39-50.
- Clahsen, H., F. Avelado and I. Roca (2003) 'The development of regular and irregular inflection in Spanish child language'. *Journal of Child Language* 29(3), 591-622.
- Davidson, L. and M. Goldrick (2003) 'Tense, agreement and defaults in child Catalan: An optimality theory analysis', in: S. Montrul and P. Ordóñez (eds.) *Linguistic Theory and Language Development in Hispanic Languages*. Cambridge, MA: Cascadilla Press, 193-211.
- Ezeizabarrena, M.J. (2002) 'Root infinitives in two pro-drop languages', in: A.T. Pérez-Leroux and J.M. Licerias (eds.) *The Acquisition of Spanish Morphosyntax: The L1/L2 Connection*. Dordrecht: Kluwer, 35-65.
- Ezeizabarrena, M.J. (2003) 'Null Subjects and optional infinitives in Basque', in: N. Müller (ed.), *(In)vulnerable Domains in Multilingualism*. Amsterdam: John Benjamins, 83-106.
- Gagarina, N. (2002) 'Thoughts on optional infinitives (in Russian)', in: D. Hole, P. Law and N.Zhang (eds.) *Webfest for Horst-Dieter Gasde*. <http://www.zas.gwz-berlin.de/mitarb/homepage/webfest>

- Guasti, M.T. (1994) 'Verb syntax in Italian child grammar: Finite and non-finite verbs'. *Language Acquisition* 3(1), 1-40.
- Harris, J. (1991) 'The exponence of gender in Spanish'. *Linguistic Inquiry* 22(1), 27-62.
- Hernández-Pina, R. (1984) *Teorías Psicolingüísticas y su Aplicación a la Adquisición del Español como Lengua Maternal*. Madrid: Siglo XXI.
- Hoekstra, T. and N. Hyams (1998) 'Aspects of root infinitives'. *Lingua*, 106, 81-112.
- Hyams, N. (2001) 'Now you hear it, now you don't: the nature of optionality in child grammars', in: A.H-J. Do et al. (eds.) *BUCLD Proceedings*. Somerville, MA: Cascadilla Press, 34-58.
- Hyams, N. (2005) 'Aspect matters', in: K.U. Deen, J. Nomura, B. Schulz and B.D. Schwartz (eds.) *Proceedings to the Inaugural Conference on Generative Approaches to Language Acquisition – North America (GALANA)*. Cambridge MA: UCONN/MIT Working Papers in Linguistics.
- Liceras, J., E. Valenzuela and L. Díaz (1999) 'L1/L2 Spanish grammars and the pragmatic deficit hypothesis'. *Second Language Research* 15(2), 161-190.
- Liceras, J., A. Bel and S. Perales (2006) 'Living with optionality: Root infinitives, bare forms and inflected forms in child null subject languages', in: N. Sagarra and A.J. Toribio (eds.) *Selected Proceedings of the 9th Hispanic Linguistics Symposium*. Somerville, MA: Cascadilla Press, 203-216
- Perales, S., J. Liceras and A. Bel (2005) 'Some remarks on the realization of the *realis/irrealis* opposition in child language: Towards a universal characterization of the root infinitive stage'. *Ottawa Working Papers in Linguistics* 34, 87-98.
- Pratt, A. and J. Grinstead (2007) 'Optional infinitives in child Spanish', in: A. Belikova, L. Meroni and M. Umeda (eds.) *Proceedings of the 2nd Conference on Generative Approaches to Language Acquisition in North America (GALANA)*. Somerville, MA: Cascadilla Proceedings Project, 351-362.
- Radford, A. (1990) *Syntactic Theory and the Acquisition of English syntax: The nature of early child grammars of English*. Oxford: Basil Blackwell.
- Radford, A. and I. Ploennig-Pacheco (1995) 'The morphosyntax of subjects and verbs in child Spanish: A case study'. *Essex Reports in Linguistics* 5, 23-67.
- Rhee, J. and K. Wexler (1995) 'Optional infinitives in Hebrew'. *Massachusetts Institute of Technology Working Papers in Linguistics* 26, 383-402.
- Rizzi, L. (1993/1994) 'Some notes on linguistic theory and language development: The case of root infinitives'. *Language Acquisition* 3, 371-393.
- Salustri, M. and N. Hyams (2003) 'Is there an analogue to the RI stage in the null subject languages?', in: B. Beachley, A. Brown and F. Conlin (eds.) *Proceedings of the BU Conference on Language Development* 27. Cambridge, MA: Cascadilla Press, 692-703.
- Salustri, M. and N. Hyams (2006) 'Looking for the universal core of the RI stage', in: V. Torrens and L. Escobar (eds.) *The Acquisition of Syntax in Romance Languages*. Amsterdam/Philadelphia: John Benjamins, 159-182.
- Schaeffer, J. and D. Ben-Shalom (2004) 'On root infinitives in child Hebrew'. *Language Acquisition* 12(1), 81-96.
- Tsimpli, I.M. (1992) *Functional Categories and Maturarion. The Prefunctional Stage of Language Acquisition*, PhD Dissertation, University College London.
- Ud Deen, K. (1997) 'The interpretation of root infinitives in English: is eventivity a factor?', ms. UCLA.
- Wexler, K. (1994) 'Optional infinitives, head movement and the economy of derivations', in: D. Lightfoot and N. Hornstein (eds.) *Verb Movement*. Cambridge: Cambridge University Press, 305-362.

Acquiring a Null Object Language When children sound target-like, but really aren't

Ruth E. Vasconcellos Lopes
Universidade Estadual de Campinas/CNPq¹

1. Introduction

The aim of this paper is to examine the acquisition of null direct objects in two Brazilian monolingual children.

As is well known, Brazilian Portuguese (BP, henceforth) exhibits null objects in any syntactic context, as opposed to other languages that allow the phenomenon (1), but one striking aspect of BP null object is that it occurs more freely when the antecedent has a [-animate] feature (2):²

(1) Comprei o casaco depois que experimentei Ø.
Buy_1ps_past the coat after that try+on_1ps_past Ø.
“I bought the coat, after I tried (it) on”

(2) O Emilio perdeu [a carteira] e não consegue achar Ø/?ela
The Emilio lost_3sg the wallet and not can_3sg find_inf Ø/?pron_fem
“Emilio has lost his wallet and can't find (it)”

On the other hand, BP allows for strong and/or weak 3rd person pronouns – which we will refer to as ‘lexical pronouns’ from now on – in object position; nevertheless, the distribution between a null object and a lexical pronoun is not free in the language, but is restricted to the semantic features of the antecedent, namely, animacy and specificity. Default null objects refer to [-animate] antecedents as in (2) above, whereas a [+animate] antecedent requires a null object only if [-specific] (3a), and a lexical pronoun if [+specific] (3b):

(3) a. Policial insulta [presos] antes de torturar Ø/?eles
Policeman insults prisoners before of torture_inf Ø/?them
“Policemen insult prisoners before torturing (them)”
b. O policial insultou [o preso] antes de torturar *Ø/ele
The policeman insulted_3sg the prisoner before of torture_inf *Ø/him
“The policeman insulted the prisoner before torturing him”

¹ This paper is part of a research project funded by the Conselho Nacional de Desenvolvimento Científico e Tecnológico, CNPq/Brazil, grant # 303617/2004-7. I'd like to thank the II Romance Turn audience for their comments. The usual disclaimer applies.

² For a comprehensive review on the null object in BP, see Cyrino and Reich (2002) and references therein.

European (EP) and Brazilian Portuguese also license VP ellipses, differently from other Romance languages. In both varieties it may occur either with an auxiliary or a main verb, as long as the verb which licenses the elliptical constituent also appears in the antecedent VP:

- (4) Ela tem lido livros às crianças e ele também tem lido [-]³
 She has read books to+the children and he too has read
 “She has read books to the children and he has, too.”

VP-ellipses are also found in short answers to Yes/No questions (see Martins 1994; among others)

- (5) — Você entregou o livro pra Maria?
 You gave the book to Mary?
 “Did you give Mary the book?”
 — Entreguei.
 Give_past_1ps
 “Yes, I did.”

Cyrino and Matos (2002) show that “*in Brazilian Portuguese both null objects and VP ellipses may be licensed by functional heads distinct from T*” (p. 191) – a point to which we will return. Based on that, our hypothesis is that null anaphoric direct objects in BP have to be locally licensed by an ASP(ect) category, in order for the specificity feature of the null object to be checked. The assumption underlying this view takes the null object in BP to be a nominal ellipsis.⁴

Our hypothesis makes the straightforward prediction that convergence into the adult grammar, in what regards the anaphoric null objects, will be dependent upon the acquisition of ‘aspect’, translatable into the grammatical features of (im)perfectivity in ASP.

The remainder of this paper is organized in the following way. Based on our hypothesis, we will further explore some predictions for language acquisition in section two. We will also consider well established results for the acquisition of direct objects in non-null-object languages in comparison to BP. In section three we present our results and discuss them. We close the paper in section four with some final remarks.

³ Example from Cyrino and Matos (2002).

⁴ See Lopes and Cyrino (2005) for such an analysis, based on Cyrino (1997), which assumes that the null object is the result of reconstruction of the antecedent at LF, which could elided in PF. Some researchers have proposed that the null object in BP is an empty pronoun, *pro*, but the recurrent problem with these proposals is that there is no agreement on the requirements of identification and licensing of the empty category (cf. Galves 2001; Kato 1993; Bianchi and Figueiredo 1994; Barra Ferreira 2000; among others). Besides, these proposals are unable to capture the animacy constraints we have discussed, invariably offering a stipulation on this aspect of the null object.

2. Predictions for language acquisition

It is quite well established in the literature that children acquiring languages with object clitics go through a clitic omission stage (Avram 2001; Hamann 2003; Jakubowicz et al. 1998; Tsakali and Wexler 2004; among many others)⁵, which is not attested in children acquiring languages with a strong and/or weak pronoun paradigm for object, such as English (Huang 1999; Hyams and Wexler 1993; Wang et al. 1992; among others). It has also been shown that children acquiring Chinese, a null object language, produce null objects from the onset and are quantitatively close to the production of adults (see Wang et al. 1992; Yip and Matthews 2000).

The natural question is: What happens in a language that has lost 3rd person clitics, developing a null object option plus a strong and/or weak pronominal one, depending on the semantic features of the antecedent, while still preserving 1st and 2nd person clitics?

This picture presented on the first paragraph allows us to make the following prediction: Since children acquiring BP will not have to deal with the complexities involved in deriving clitics, they should exhibit the expected adult-like pattern for objects. In other words, the null option will be produced from the onset, as in Chinese, and the lexical pronouns will also be an option, just as they are in English child grammars.

However, if our hypothesis is on the right track, anaphoric null objects are not to be expected until the relevant features in ASP are acquired. Therefore, if null objects are produced from the start, they should be special instances of this phenomenon and they might be licensed otherwise. In order to check that, we will examine the syntactic contexts in which the null objects appear in Brazilian children's grammar correlating them to lexical and grammatical aspect.

Finally, we will also assume that VP-ellipses are licensed in the language through ASP, in which case we should not find adult-like short answers to Yes/No questions until such category is fully operative in the child's grammar. We will follow Cyrino and Matos (2002) according to whom such ellipses are licensed by the T(ense) head in EP but not in BP, as pointed out before. Thus, different patterns of short answers should be found among Portuguese and Brazilian children.

3. Acquisition Data

3.1 Participants and Procedures

Spontaneous speech production from two children (AC. and R.), aged 1;8 to 3;7 and 1;9 to 2;8, respectively, was examined and quantified. Both children, monolinguals exposed to standard BP, are daughters to monolingual parents with

⁵ We acknowledge that this point is not uncontroversial, since there might be some microparametric variation among clitic object languages. We will ignore the point here since it bears no relevance to our discussion here.

graduate-level education or beyond.⁶ The children were taped at home, always in the presence of one of the parents, in playful naturalistic situations. Sections lasted from 30 to 60 minutes. There are no observable dialectal differences in both varieties of Brazilian Portuguese with regard to the null object.

Only 3rd person object contexts were taken into account. Categorically null objects, such as in propositional ellipses were disregarded in order not to bias our results artificially:

- (6) A(dult): E o que acontece na história do Príncipe do Egito?
 “And [what happens in the story of the Egyptian prince]?”
 C(hild): Já esqueci Ø. (AC, 3;7)
 Already forgot-1stsg Ø
 “I’ve already forgotten it”

Short answers to Yes/No questions were not considered for two reasons: (i) in order to sort out contexts of null objects and cases of VP ellipses and (ii) there were very few instances of such questions addressed to the children during taping. (7) illustrates one of the few examples found. We will return to this point, though.

- (7) (adult and child are pretending to host a tea-party)
 A: A senhora aceita um suco?
 The madam accepts [a juice]?
 “Would you, madam, like a glass of juice?”
 C: Aceito Ø. (AC, 2;1)
 Accept_1ps_present
 “Yes, I do”

We have also analyzed Mood, Tense and Aspect (lexical and grammatical) usage by one of the children.

3.2 Results and Discussion

We will start with the general results for both children in Table 1 below.

Null		Lexical pronouns		DPs		Total	
N	%	N	%	N	%	N	%
275	29.2	93	9.8	575	61	943	100

Table 1: General results for both children

Although both children use null objects, they are still quantitatively far from the target grammar, where null objects reach around 60% and lexical pronouns, 15%,

⁶ I would like to thank CEAAL/PUCRS (Centro de Aquisição e Aprendizagem da Linguagem) for allowing me access to their database (AC.). Data from R. belong to CEDAE/UNICAMP.

according to Duarte (1986). It is interesting to observe the high percentages of DPs in object position, a point to which we will return briefly.

Table 2 considers only the null and pronominal realizations of the object. When DPs are excluded, and the option, thus, is between a lexical pronoun or a null object, it becomes clear the child's preference for the latter.

Child	Null		Lexical pronoun		Total	
	N	%	N	%	N	%
R.	134	75.2	44	24.8	178	100
AC	141	74.2	49	25.8	190	100
Both	275	74.7	93	25.3	368	100

Table 2: Mean results for null and pronominal realizations of the object

But as we will discuss below this does not mean that the child's null is always the same one. We will examine the behavior of null and pronominal elements during development in Table 3.

Age	AC		R		Total
	% null	% pronoun	% null	% pronoun	N
1;8-1;9	100	0	100	0	9
1;10	100	0	75	25	17
2;1	100	0	69.7	30.3	95
2;3	85	15	84.7	15.3	85
2;8	73	27	64	36	52
3;0	64	36			78
3;7	81	19			32
Total	74.2	25.8	73.5	24.7	368

Table 3: Percentage of null and pronominal objects for each child over time

Table 3 clearly shows an increasing pattern of the use of pronouns over time, while a decrease on the use of nulls is obviously observed. The next natural question is: Are we dealing with one and the same null category or does its status change over time? The results show a very interesting behavior in both children, albeit taking place in different age groups. Both of them start out with a production of 100% of null objects, but obviously such figure decreases when pronouns kick in. For R. that happens when she is 1;10 and for AC, when she is 2;3.

Qualitatively examining the data, what we see is that the initial null objects are instances of deictic-like elements in imperative contexts, especially in the initial files, but when the productive distinction between perfective and imperfective sentences appear, then the null objects become anaphoric. Coincidentally, that is also when pronouns start to be produced in object position. We will return to that when discussing the results for 'aspect' as well. Let's compare (8) – a deictic use of null – to (9), an anaphoric null.

- (8) a. Garda (= guarda) aqui. (R., 1;9)
 Keep \emptyset here
 “Keep it here”
 (The child utters the sentence while holding her pacifier, obviously referring to it)
- b. Messi (= mexe). (AC, 1;8)
 Stir \emptyset
 “Stir it up”
 (The child is holding a glass of juice and asks her mother to stir it for her)
- (9) Não vou guardar. (AC, 3;7)
 not will_{1sg} keep \emptyset
 “I won’t put them away”
 (referring to her toys. The child wants to watch a movie on TV, so she comes to her mother in order for her to turn the TV on. But the mother knows that the child was playing in her room and that there are toys all over the place. Her mother tells her to put the toys away before watching the movie. The child walks away, while muttering the sentence in (9).)

The initial use of deictic null objects occurs with imperative sentences and is restricted to inanimate 3rd person antecedents. It is a root phenomenon which we will assume is derived under MoodP, being neither T nor Asp available in the Lexical Array, since they have a root deictic nature:⁷

(10)[MoodP [vP/VP]]

We have analyzed the use of imperative sentences in AC and found that, in fact, there is a high percentage of such structures in the initial files examined, dropping drastically and tending to a steady pattern from 2;3 on, as can be observed in Figure 1.

⁷ According to Salustri and Hyams (2003), the Mood head checks the ‘irrealis’ feature of the verb. We will assume their analysis for the imperative sentences, but not the proposal that they are an analogue to root infinitive sentences in null subject languages, which is beyond our point here. Nevertheless, BP should be an interesting empirical checking space for such a proposal.

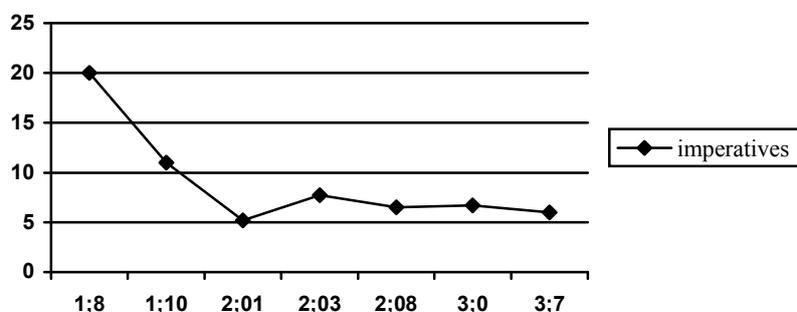


Fig. 1: Percentage of imperative sentences through different ages in AC

It is obviously not the case that the child goes through an ‘imperative-only’ stage. In fact, the first file examined contained a high percentage of verbs in the indicative mood (in the 1;8 file, 68.4% of linking verbs occur in the present tense – in templates such as X is Y – and 11.6% in the past tense, more specifically in the ‘perfect preterite’ – a perfective form).

However, the first past forms found in the data suggest a telic interpretation of the sentence⁸ and children acquiring BP seem to expand the deictic use of null objects to such contexts, as can be seen in (11):

- (11) Telô Ø ! (= tirou) (R. 1;9)
 Took Ø
 “Someone took (it)”

Therefore, we propose that imperative sentences can only license deictic null objects, since the only functional category available is MoodP. Yet, this correlation is not bidirectional; in other words, if the only functional category available is MoodP, then only deictic null objects will be licensed but under TP and AspP both deictic and anaphoric null objects can be found – the case in (11). It is very important to notice that it does not seem to be the case that children have problems in establishing anaphoric relations. We claim that using only deictic null objects up to a certain point in the development is not an extra-linguistic cognitive constraint, since they use full DPs anaphorically, which explains the high percentages of DPs in object position found in Table 1. Rather, it is a grammatical constraint: Their grammars are still incapable of licensing anaphoric null objects.

Let’s examine now AC’s grammar development with respect to the lexical (*Aktionsart*) and grammatical aspect.

⁸ See Antinucci and Miller 1976; Bickerton 1984; Brown 1973; De Lemos 1981; Shirai and Anderson 1995; Slobin 1985; Wagner 2006; Weist et.al. 1991; among many others on what has been doomed as the aspect-first-hypothesis.

Imperfective forms, such as the progressive forms (12) or ‘imperfective preterite’ ones (13) become productive from the age of 2;3 onwards:

- (12) a. *Ãh hã... tá querendo.* (AC 3;7)
 [] is wanting Ø.
 “Yes, (someone) wants it
- b. *[...] mas não tão comendo.* (AC 3;7)
 ... but they are not eating Ø.
 “But they are not having them”
- (13) *Quando eu era bem pequenininha.* (AC 3;7)
 When I be_imp_past very small.
 “When I used to be a baby”

Up to this stage, most sentences are morphologically marked as perfective forms occurring with achievement verbs. At that state, activity and accomplishment verbs are found. The achievement ones drop from 75%, in file 1;8, to 17.3% in file 2;3.

Thus, we are lead to assume that the ASP head is projected since very early on, but initially it contains an unspecified default [+ perfective] feature, inoperative to license null anaphoric objects and to check the antecedent features of the nominal ellipsis, especially the specificity one. We also claim that the T head is operative very early, given the fact that it accommodates present/past morphological distinctions on the verb. Although children do not interpret the temporal distinctions as such these distinctions seem to be part of the syntactic structure anyhow.

In order to unify these phenomena, we will backtrack to Cyrino and Matos’ (2005) proposal for VP ellipses in Portuguese. According to these authors, ellipsis is licensed in EP by T and in BP “*by functional heads distinct from T*” (p. 191), as we have discussed in the Introduction. More recently, Cyrino and Matos (2005) take the licensing head in BP to be the first head to merge with the (to be) elided VP. We assume here that this licensing head is ASP. Cyrino and Matos (2005) test the difference between the two varieties with the focusing adverb *também* (too):⁹

- (14) *Ela tem lido livros às crianças e ele também tem lido [-]*
 She has read books to-the children and he too has read
 “She has read books to the children and he has, too.”
- (15) a. *Ela tem lido livros às crianças e ele tem também lido [-]*
 She has read books to-the children and he has too read
- b. BP: [-] = [VP [V t] the books to the children]
- c. EP: [-] = i. ??[VP [V t] the books to the children]
 ii. ok [V t] [-]

⁹ In Cinque’s (1999) terms.

In BP even with the adverb in a lower position – lower than TP –, the strict elliptical reading is still available (15b), while in EP the only reading is a non-specific one (15c) – ‘he has been reading anything’. Cyrino and Matos (2005) propose the structure in (16) to account for (15):

(16) ... ele [TP tem [AdvP [Adv^o também] [VP aux t [AspP lido [vP -]]

According to our hypothesis, we should not expect to find low focusing adverbs or aspectual adverbs prior to the acquisition of the perfectivity features in ASP. On the contrary, they should count as independent evidence for the acquisition of such a head which, when operative, would license anaphoric null objects.

In AC’s data we found only one instance of the use of *também*, when she is 1;10:

(17) Eu também tenho Ø (AC, 1;10)
 I have too Ø
 “I also have one”
 (The child is ‘serving cups of coffee’ to herself, her mother and a teddy bear)

The adverb in (17) is above TP, considering the verb moves to T in Portuguese, and the object is still a deictic null one. On the other hand, aspectual adverbs start to be produced from 2;3 onwards, the age when the child makes the morphological distinction between perfective and imperfective forms. Examples (18) to (21) show a different range of temporal and aspectual forms, as well as the use of lexical and null objects, the latter as deictic or anaphoric instances. Her grammar seems to be going through a stage of transition:

(18) Aqui já comeu Ø. (AC, 2;3)
 Here already ate Ø.
 “Here, he ate it already”

(19) Já tem out(r)o bicho. (AC, 2;3)
 Already has another bug.
 “There is another bug already”

(20) Não comeu Ø ainda. (AC, 2;3)
 Not ate Ø yet.
 “He has not eaten (his veggies) yet.”

(21) Já tomou Ø. (AC, 2;3)
 Already took Ø.
 “She has already taken her shower.”

Finally, we would like to go back to the predictions concerning short answers to Yes/No questions, shown in (5) and repeated below as (22):

- (22)— Você entregou o livro pra Maria?
 You gave the book to Mary?
 “Did you give Mary the book?”
 — Entreguei.
 Give_past_1ps
 “Yes, I did.”

If T licenses VP ellipses in such contexts in EP and ASP licenses them in BP, one should expect to find adult-like short answers in children acquiring EP earlier than in children acquiring BP, given the assumptions made before. Although we did not investigate that in our own data, for reasons pointed out earlier, Oliveira (2000) shows that Brazilian children use different strategies in different stages of development on their short answers:

- (23) A: Qué água?
 “Do you want some *water*?”
 C: awa (= água) (R. 1;8)
Water.
- (24) A: Qué chocolate?
 “Do you want some *chocolate*?”
 C: kiate (= chocolate) (R. 1;8)
“Chocolate”
- (25) A: Tirô tudo?
 Took everything?
 “Did you take *everything* out?”
 C: Tudo (R. 1;8)
Everything
- (26) A: Cê quer pôr o microfone embaixo do gravador?
 “Do you want to place the microphone *under* the tape-recorder?”
 C: bassu (= embaixo) (R. 1;8)
Under.
- (27) A: Quebrou o balde, filha?
 Broke the bucket, daughter?
 “Is the bucket broken, dear?”
 C: É (R. 1;8)
 Be_3ps_pres
 “Yes”

(28)A: Tá machucadinho?
 Is hurt_little?
 “Is (the toy) a bit bruised?”
 C: É (I. 2;2)
 Be_3ps_pres
 “Yes”

(29)A: Você acha Aldáisa bonita?
 “Do you find Aldáisa pretty?”
 C: Acho (I. 2;4)
 Find_1ps_pres
 “Yes, I do (find her pretty)”

From (23) to (26), the child is merely repeating one of the words found in the adult’s question. In examples (27) and (28) the child seems to use the verb ‘to be’ as a *placeholder* in T. Answers with the verb ‘to be’ is also a possibility in the adult language, however they carry a clear pragmatic distinction, bearing an assertive reading which lacks in the child’s responses. Finally, the example in (29) conforms to the adult grammar.

In contrast, according to Santos (2006), the picture is quite different in EP:

(30)M: *Queres* andar no cavalinho?
 “Do you want to ride the horsie?”
 C: *Qué.*¹⁰ (I. 1;5)
 Want_3ps_pres

(31)M: *Fez* ai ai ao Tomás?
 “Did (something/someone) hurt you, Tomás?”
 C: *Fez.*¹¹ (T. 2;2)
 Hurt_3ps_pres

(32)M: *Tinham* chocolate lá dentro?
 “Did they have chocolate inside?”
 C: *Ti(nh)am.* (I. 2;3)
 Had_3pp
 “Yes, they did”

Examples (30) to (32) clearly attest that the Portuguese children are producing VP ellipses from very early on in an adult-like fashion. Our prediction, therefore, holds true.

¹⁰ There is an agreement mismatch in the verb form which is irrelevant for our purposes here.

¹¹ The question involves a light verb (make) plus a nominal onomatopoeic expression (ai + ai) = hurt. The expected answer should contain the light verb as seen in the child’s response.

In what follows we will summarize the findings so far, pointing out the strong empirical correlation seen in the data in what regards the use of anaphoric null objects once ASP is operative in the grammar.

4. Summary

It is quite intriguing that many superficially unrelated processes go on pretty much at the same age range:

- (i) There is a drastic drop in the production of imperative forms at age 2;1, from there on reaching figures that are stable until the last file examined (see Figure 1);
- (ii) At age 2;3 the presence of anaphoric null objects is attested, together with the initial production of lexical pronouns in object position (see Table 3);
- (iii) Also at age 2;3 the perfective/imperfective distinction seems to become specified in AspP, considering the production of imperfective forms and the presence of varied verb types when lexical aspect is taken into account. Especially, there is a significant drop in the use of achievement verbs with perfective forms;
- (iv) That is also the age when aspectual adverbs are first attested. Prior to that file, however, temporal adverbs are largely produced;
- (v) non adult-like patterns for short answers to Yes/No questions give way to the convergent forms.

These facts cannot be taken as a mere coincidence, but as a result of the proper selection of features in a functional head that becomes active in the children's grammar. According to our hypothesis that job is done by the ASP head which will license anaphoric null objects in the language.

Finally, a word on the use of lexical pronouns. Both children initially tend to associate them with a [+human] feature of the antecedent and expand the use of the nominal ellipsis for other non-human [+animate, +specific] antecedents (ca. 50% of all cases, lasting until the 2;8 files) – the only cases that truly don't sound adult-like. Is the delay in the production of lexical pronouns also linked to the same phenomenon, i.e., are they also licensed by ASP? Or is it a matter of lexical learning? The former should not be the case, since the non-target-like null objects with animate antecedents last longer, when anaphoric null ones are already in place. It seems plausible to consider that children initially treat lexical pronouns as strong ones, which according to Cardinaletti and Starke's (1999) taxonomy would be associated to [+human] antecedents. It may be the case that children are avoiding the movement a weak pronoun should undergo, leaving its thematic position. If that were true, then the delay in convergence in the use of lexical pronouns in object position could be a matter of lexical learning.

One has to bear in mind that deictic null objects are certainly an option in BP adult grammar, that is why children, even though not convergent into the expected grammar, sound target-like, but really aren't.

References

- Antinucci, F. and R. Miller (1976) 'How children talk about what happened'. *Journal of Child Language* 3, 167-89.
- Avram, L. (2001) 'Remarks on the optional clitic stage in child Romanian'. *Bucharest Working Papers in Linguistics* 3, 16-28.
- Barra Ferreira, Marcelo (2000) 'Argumentos nulos em português brasileiro'. M.A. Thesis, UNICAMP, Campinas.
- Bianchi, V. and M. C. Figueiredo Silva (1994) 'On some properties of agreement-object in Italian and Brazilian Portuguese', in: M. Mazzola (ed.) *Issues and theory in Romance linguistics*. Washington, DC: Georgetown University Press, 181-190.
- Bickerton, D. (1984) 'The language bioprogram hypothesis'. *Behavioral and Brain Sciences* 7, 173-88.
- Brown, R. (1973) *A First Language*. Cambridge, MA: Harvard University Press.
- Cardinaletti, A. and M. Starke (1999) 'The typology of structural deficiency. A case study of the three classes of pronouns', in: H. van Riemsdijk (ed.) *Clitics in the Languages of Europe*. Berlin: Mouton de Gruyter, 145 - 233.
- Cinque, G. (1999) *Adverbs and Functional Heads*. Oxford: Oxford University Press.
- Cyrino, S. (1997) *O Objeto Nulo no Português do Brasil - um estudo sintático-dicarônico*. Londrina: Editora da UEL.
- Cyrino, S. and G. Matos (2002) 'VP ellipsis in European and Brazilian Portuguese'. *Journal of Portuguese Linguistics* 1, 177-195.
- Cyrino, S and G. Matos (2005) 'Local licensers and recovering in VP ellipsis'. *Portuguese Journal of Linguistics*. 4, 79-112.
- Cyrino, S and U. Reich (2002) 'Uma visão integrada do objeto nulo no Português Brasileiro' *Romanistisches Jahrbuch*. Berlin: Walter de Gruyter, 360-386.
- De Lemos, C. (1981) 'Interactional processes in the child's construction of language', in: W. Deursch (ed.) *The child's construction of language*. London: Academic Press, 57-76.
- Duarte, M. E. L. (1986) 'Variação e sintaxe: clítico acusativo, pronome lexical e categoria vazia no Português do Brasil'. MA Thesis: PUC, São Paulo.
- Galves, C. (2001) *Ensaio sobre as gramáticas do português*. Campinas: Editora da Unicamp.
- Hamann, C. (2003) 'Phenomena in French and normal impaired language acquisition and their implication for hypotheses on language development'. *Probus* 15, 91-122.
- Huang, P-Y (1999) 'The development of null arguments in a Cantonese-English bilingual child'. MPhil Thesis, Chinese University of Hong Kong.
- Hyams, N. and K. Wexler (1993) 'On the grammatical basis of null subjects in child language', *Linguistic Inquiry* 24, 421-459.
- Jakubowicz, C.; L. Nash; C. Rigaut and C-L. Gérard (1998) 'Determiners and clitic pronouns in French-speaking children with SLI'. *Language Acquisition* 7, 113-160.
- Kato, M.A. (1993) 'The distribution of pronouns and null elements in object position in Brazilian Portuguese', in: W. Ashby, M. Mithun, G. Perissinoto and E. Raposo (eds.) *Linguistic Perspectives on the Romance Languages*. Amsterdam/Philadelphia: John Benjamins, 225-235.
- Lopes, R.E.V. and S. Cyrino (2005) 'Evidence for a cue-based theory of language change and language acquisition: The null object in Brazilian Portuguese', in: T. Geerts; I. van Ginneken and H. Jacobs (eds.) *Romance Languages and Linguistic Theory 2003*. Amsterdam/Philadelphia: John Benjamins, 343 - 359.
- Martins, A.M. (1994) 'Enclisis, VP deletion and the nature of Sigma'. *Probus* 6, 173-205.

- Oliveira, M. (2000) *Frases Assertivas e sua Variação nas Línguas Românicas: seu papel na aquisição*. São Paulo: Humanitas.
- Salustri, M. and N. Hyams (2003) 'Is there an analogue to the RI stage in the null subject languages?', in *Proceedings of Boston University Conference on Language Development 27*. Somerville MA: Cascadilla Press, 692-703.
- Santos, A. L. (2006) 'On the nature of early VP ellipsis'. Paper presented at *The Romance Turn II*. Utrecht, September 8.
- Shirai, Y. and R. Andersen (1995) 'The acquisition of tense-aspect morphology: a prototype account'. *Language* 71, 743-62.
- Slobin, D. (1985). 'Crosslinguistic evidence for the language-making capacity', in: D.Slobin (ed.) *The Crosslinguistic Study of Language Acquisition*. Hillsdale: Lawrence Erlbaum, 1157-1255.
- Tsakali, V. and K. Wexler (2004) 'Why children omit clitics in some languages but not in others: New evidence from Greek', in: J. van Kampen and S. Baauw (eds.) *Proceedings of GALA 2003* (LOT Occasional Series). Utrecht: LOT, 493-504.
- Wagner, L. (2006) 'Aspectual bootstrapping in language acquisition: telicity and transitivity'. *Language Learning and Development* 2, 51-76.
- Wang, Q.; D. Lillo-Martin; C. Best and A. Levitt (1992) 'Null subjects and objects in the acquisition of Chinese'. *Language Acquisition* 2, 221-254.
- Weist, R., H. Wysocka and P. Lyytinen (1991) 'A cross-linguistic perspective on the development of temporal systems'. *Journal of Child Language* 18, 67-92.
- Yip, V. and S. Matthews (2000) 'Dual input and learnability: null objects in Cantonese-English bilingual children'. *Bilingualism: Language and Cognition* 3, 193-208.

Beyond Deficiency

Reconsidering the null subject phenomenon

Katérina Palasis-Jourdan and Michèle Oliviéri
Laboratoire BCL, Université de Nice Sophia-Antipolis, CNRS

0. Abstract

This paper investigates the child null subject phenomenon in L1 acquisition within the generative framework. We consider that the traditional hypotheses, usually based on child deficiency, have their limits. We therefore propose a new analysis of this phenomenon, focusing our attention on null subjects in finite clauses. To that purpose we introduce new data recorded from 17 French children between 2;3¹ and 3;1 and a new approach borrowed from dialectology.

Our first section forwards preliminary clarifications with regard to the null subject phenomenon, the existing hypotheses, their limits and our methodology. In the second section we detail our data and demonstrate that the child null subject stage can be accounted for in French and in other non pro-drop languages, taking a step further beyond the traditionally assumed child competence or performance deficiency.

1. Preliminary clarifications

1.1 On the null subject phenomenon

The child null subject phenomenon represents a long-standing issue within the generative framework. We can briefly describe it as follows: children acquiring a non pro-drop language, such as French or English, systematically omit some of their subjects until the approximate age of 3;0. As far as French is concerned, the omission rate amounts to circa one quarter of the children's sentences. Rasetti (1996) for instance forwards a rate of 26.2% and our data show an average rate of 22.4%². This means that French children systematically produce sentences that do not conform to the correct <-pro-drop> value of the adult target language. The question of course is why.

Two types of null subjects have been identified and studied since the 1980s: null subjects with finite verbs –as shown in Table 1 hereafter–, and null subjects with non-finite verbs –also known as ‘root infinitives’ (Rizzi 1993) or ‘optional infinitives’ (Wexler 1992). A few examples of this second type of null subjects are displayed in Table 2. Both types of null subjects have been identified as ungrammatical compared to adult French.

¹ The ages are given in years;months.

² See Palasis-Jourdan (2005) for details. All the examples in this paper come from this corpus, except when specified.

Subjectless sentences	Names and Ages	Grammatical equivalents	Glosses
* est là	Raphaël, 2;7	“ <u>e</u> lle est là”	“ <u>s</u> he is there”
* veux livre	Tess 2;7	“ <u>j</u> e veux le livre”	“ <u>I</u> want the book”
* fais ça	Thibault, 2;9	“ <u>j</u> e fais ça”	“ <u>I</u> make that”
* a cheval	Alizée, 2;8	“ <u>i</u> l y a un cheval”	“ <u>t</u> here is a horse”
* sais pas	Maxence, 2;9	“ <u>j</u> e (ne) sais pas”	“ <u>I</u> don’t know”

Table 1: A few examples of null subject finite clauses in early French

Subjectless sentences	Names and Ages	Grammatical equivalents	Glosses
* faire un ballon	Thibault, 2;9	“je veux faire un ballon”	“I want to make a ball”
* commenc[e]	Thomas, 2;6	“commencé/er”	“begun/to begin”
* moi pas tombé	Raphaël, 2;7	“moi je (ne) suis pas tombé”	“I didn’t fall”
* m’asseoir là	Raphaël, 2;7	“je veux m’asseoir là”	“I want to sit there”
* pas trouv[e]	Alizée, 2;8	“pas trouvé/er”	“not found/to find”

Table 2: A few examples of null subject non-finite clauses in early French

1.2 Why do we hypothesize only on null subjects in finite clauses?

Our study forwards an in-depth investigation of the null subject phenomenon within finite clauses. This does not mean however that we have left the other type of null subject aside arbitrarily. The reason for this restriction lies in our data. As can be seen in Table 3 hereunder, out of our 240 null subject sentences –22.4% of the entire corpus–, only 15 sentences display a non-finite verb. This figure amounts to the rate of a mere 1.4% of the entire corpus and 6.2% of the null subject sentences. A first obvious conclusion is that this type of null subject bears a very low rate of occurrence in our data³.

Verbal Sentences	Entire Corpus	With Subjects	Without Subjects	Null Subjects + Finite Verbs	Null Subjects + Non-finite Verbs
Numbers	1072	832	240	225	15
Percentages	100	77.6	22.4	21	1.4

Table 3: Null subjects and inflection in our corpus

³ See Section 1.4 for details on our data and approach.

Additionally, among these 15 sentences with non-finite verbs, it appears that 4 of these forms are closely preceded or followed by a finite form of the same verb or of a modal verb, as shown in (1).

- (1) Non-finite forms closely preceded or followed by a finite form:
- a. *Moi moi moi moi moi s' étais caché. Caché. (Raphaël, 2;6)
Me me me me me myself was hidden. Hidden.
“I hid myself”
 - b. *Moi pas tomb[e]. [...]. Ben moi i(l) m' a pas tomb[e]. (Raph., 2;7)
Me not fallen. [...]. Er me he/it me has not fallen.
“I didn't fall”
 - c. *Veux m' asseoir. M' asseoir là. (Raphaël, 2;7)
Want myself sit. Myself sit there.
“I want to sit there”
 - d. xx veux faire un ballon. [...]. Faire un ballon. (Thibault, 2;9)
xx want make a ball. [...] Make a ball.
“xx want to make a ball” (with modelling clay)

It also appears that other non-finite forms come along with finite forms of the same verb during the same recording session. This is illustrated in (2) hereunder.

- (2) Finite and non-finite forms during the same recording session:
- a. *Moi aussi faire un ballon. [...]. (Thibault, 2;9)
Me too make a ball. [...].
“I also (want to) make a ball”
*Moi aussi faire voiture. [...].
Me too make car. [...].
“I also (want to) make a car”
*Va faire. [...].
Will make. [...].
“He will make”
*A fait. [...].
Has made. [...].
“He has made”
*[e] fait pas.
Have [?] made not.
“I have not made”
 - b. *Arriv[e]. [...]. (Raphaël, 2;6)
Managed/to manage. [...].
“I managed (to do something)”
*Arrive pas à faire.
Manage not to do.
“I can not do (something)”

These two sets of examples illustrate the fact that this type of null subject is not a relevant characteristic of the children's utterances in our data. This is the reason why we do not consider these forms as representing an independent grammatical phenomenon. Rather, their low rate and their concomitant appearance with finite forms of the same verb suggest that these children have left this particular null subject stage and we hence analyse these forms as being the result of scarce errors. This is consistent with the ages of the children –between 2;3 and 3;1– as, indeed, this type of null subject is usually observed at an earlier stage –up to around 2– and is assumed to disappear before the other type of null subject. Additionally, we will see in our conclusion that the account we have just proposed for these non-finite clauses proves to be consistent with Wexler (1998)'s hypothesis on optional infinitives.

We are then left with null subjects in finite clauses and we have to explain why we do not apply the same reasoning –i.e. a performance error– to this second type of null subject. Again, the reason lies in our data. Indeed, as already seen in Table 3 above, their rate is much higher (21% of the entire corpus). This percentage hence illustrates the fact that this phenomenon is much more systematic and robust in the linguistic system of the children under investigation. And, indeed, this particular type of null subject has already been widely highlighted and analysed within the generative framework. In the next section, we hence briefly review some of the existing and renowned hypotheses on this matter. We also explain where we think the limits of these analyses are and what kind of new approach we develop in order to provide a fully comprehensive investigation into this phenomenon.

1.3 The existing hypotheses

Traditionally, within the Principles and Parameters framework, the null subject stage has been explained thanks to a child deficient competence or performance compared to the adult. And, indeed, seminal work has been carried out on this matter during the past twenty years or so. With regard to deficiency in competence, an initial mis-setting of a parameter has been argued for, e.g. the Pro-drop Parameter in Hyams (1986) or the Root Subject Drop Parameter in Rizzi (2002). Underspecification hypotheses have also been favoured, such as the initial optional underspecification of I in Hyams (1996). With regard to performance, limited memory, computation or production capacities have been put forward. Jakubowicz and Rigaut (1997) for instance proposed that children have a smaller production capacity than adults and that this induces a more or less important phonological reduction of the pronoun when the utterance is too heavy. Some of the latest hypotheses also look into pragmatics or at the syntax-discourse interface, e.g. Serratrice (this volume).

However, to our feeling, as the omission phenomenon only affects approximately one fifth to one quarter of the children's sentences, these different types of hypotheses all bring the same questioning. Indeed, within these hypotheses, how does one explain: (i) that the children who are producing these null subjects are perfectly capable of producing overt subjects concomitantly and (ii) that this latter

type of sentence represents the majority of their utterances. In our opinion, a hypothesis on the null subject phenomenon should answer this first twofold question. However, a comprehensive analysis should not limit itself to the above and there are two additional issues that we would like to account for: (iii) the omission rate, i.e. why approximately one fifth to one quarter of the sentences and (iv) the disappearance of the phenomenon, i.e. what happens in the child's linguistic system around 3;0 that explains it. So the aim of our investigation is to coherently and concomitantly answer all four of these aspects.

1.4 Our approach

Two major elements underpin our research: new methodology and new data. As far as our theoretical background is concerned, on the one hand, we keep in mind the Principles and Parameters framework, adhering to the Pro-drop Parameter (Chomsky 1981) along with the basic dichotomy between pro-drop vs. non pro-drop languages –as formalised by Rizzi (1997) for instance. Wexler (1998)'s Very Early Parameter Setting hypothesis is also followed. However, on the other hand, as mentioned in the previous section, this traditional generative framework has its limits and we therefore open this theoretical background to another linguistic field, namely dialectology. More precisely, with dialectology we intend to focus on two points. Firstly, the importance of the data. With regard to this matter, our provisos are to work from a *wide* range of *recent* and *homogeneous* information –in time and space–, gathered in *every day* surroundings. Indeed, in our opinion, working from one child or from a very small number of children brings the risk of hypothesizing from individual linguistic characteristics and hence not forwarding a comprehensive theory of acquisition. As far as homogeneity is concerned, our data had to be collected during a restricted and recent period of time. This is why we did not work from existing French corpora such as Patsy Lightbown's⁴, Madeleine Léveillé's⁵, or Christian Champaud's⁶, recorded respectively in 1974/1975, 1971 and 1988 and that hence provide information about French uttered by the previous generation of French native speakers. Neither did we work from the data gathered under the supervision of Bernadette Plunkett⁷ between 1997 to 1999 as this information comes from three children who are geographically heterogeneous –Belgium, Canada and France– and who hence provide data on three different varieties of French. So, besides the fact that it is always scientifically very enriching to confront new and older information, these are some of the reasons why our choice went to recording a new corpus. And, in order to work from a wide range of spontaneous information, we decided to record the children from a class in a French kindergarten. This

⁴ Lightbown (1977): two children (Nathalie and Daniel).

⁵ Suppes et al. (1973): one child (Philippe).

⁶ From the Laboratoire de Psychologie CNRS in Paris: one child (Grégoire). The Léveillé and Champaud corpora are available in the CHILDES database (MacWhinney 2000).

⁷ See De Cat (2002) for details. This corpus is also available in the CHILDES database (MacWhinney 2000).

allowed us to start with as many as 25 different children, nevertheless being from the same latest generation and uttering the same variety of French. From this initial class group, the utterances of 17 children were actually analysed. Indeed, 8 children were not taken into account either because both their parents were not native French speakers, or because they were too shy and did not wish to participate or because they were part of the older children of the class and hence had already left the null subject stage. Working from as many different children enabled us to spot out the individual linguistic characteristics of each child, such as specific constructions, particular verbs or nouns, etc. We hence applied filters to the data, leaving out the verbs used only by one single child for instance. This methodology brings diversity as well as levelling, ensuring us that we are building hypotheses from solid data. In addition to this cross-sectional perspective, our corpus is also longitudinal as it counts five recording sessions spread over four months –from October 2003 to January 2004. The ages of the children range between 2;3 –the youngest child during the first session– and 3;1 –the oldest child during the last session. The data are spontaneous and comprise interactions with the investigator –looking at books, playing games, etc.– as well as interactions between the different children. The utterances selected for this study on null subjects are de facto verbal clauses and they amount to a total of 1,072.

Our second point lies in the way these data are investigated. Indeed, we sought for an explanation to the null subject stage, not within the standardly assumed child deficiency in competence or performance, but rather within their linguistic system, hence going beyond the usual deficiency-based analyses. As an illustration of this reasoning, let's imagine the following situation. A dialectologist, after having gathered data, notices that the language he is studying presents a certain number of null subjects, despite the fact that this language has already been classified as a non pro-drop language. Would this dialectologist, for all that, propose an explanation based on the idea that the adults speaking this language show a deficient competence or a deficient performance? Certainly not. Rather, we suppose that this researcher would study the whole linguistic system –especially the verbal and pronominal sub-systems– and hence seek for an intra-linguistic explanation. This is exactly what we propose to do with the children's data. However, working on French oral subjects implies preliminary clarifications with regard to what the term 'subject' exactly means.

1.5 What does the term 'subject' cover in the 'null subject stage' appellation?

We all know that two types of elements, i.e. pronouns and nouns, can bear the role of subject. However, we also know that, in French, the former category divides into at least two traditional subcategories, i.e. clitics and strong pronouns⁸. Consequently, when it comes to oral data and languages such as French that display clitics and phenomena such as doubling or dislocation, there are many issues that

⁸ We keep to this traditional dichotomy and hence do not use the more recent tripartition proposed by Cardinaletti and Starke (1994) as we assume that it is not relevant here.

can be debated with regard to the notion of what the ‘subject’ of a sentence is. Indeed, an additional variable lies in the fact that nouns, clitics and strong pronouns can be found either on their own or combined in different ways, as shown in our child data in (3) hereunder.

(3) Different combinations of nouns, strong pronouns and clitics in our corpus:

- a. Les filles elles jouent là. (Raphaël 2;8)
“the girls they play there”
- b. Il dort le monsieur. (Raphaël, 2;6)
“He sleeps the man”
- c. Moi je fais ça. (Kassandra 2;9)
“Me I do that”
- d. Je fais le pain moi. (Julia, 3;0)
“I make the bread me”

We know that the issues are numerous with regard to these different sentences and that they are still under debate. It is obvious that, within the scope of this article, we can not address all of them. We hence have to leave the very interesting issue about the status of the clitics –arguments or verbal inflections– and how these constructions should be analysed –dislocations or doublings– aside⁹. The main issue we address here is: what does the term ‘subject’ stand for exactly in the appellation ‘null subject stage’? The first step we take in order to answer this question is to detail the different types of ungrammatical sentences with regard to the absence of a subject in our corpus. The results of our analysis are presented in Table 4 hereafter.

Types ¹⁰	Examples	Glosses	N	%
–C –N	* _ veux un camion	“(I) want a lorry”	151	14.1
+PrP –C	* moi _ fais	“me (I) do”	33	3.1
–C +PoP	* _ veux vert moi	“(I) want green me”	33	3.1
–C +PoN	* _ est fermé le manège	“(It) is closed the roundabout”	15	1.4
+PrP –C +PoP	* et moi _ veux ça moi	“and me (I) want that me”	8	0.7
Subtotal for ungrammatical subjects			240	22.4
Subtotal for grammatical subjects			832	77.6
Total entire corpus			1072	100.0

Table 4: The different types of ungrammatical ‘subjects’ in our corpus

Table 4 highlights that these different types of null ‘subject’ clauses can be subsumed under one slightly altered and more refined appellation, i.e. null ‘clitic’ clauses. Indeed, these sentences either display a pre- or post-verbal strong pronoun,

⁹ But see De Cat (2005) for a review on this matter.

¹⁰ +=uttered, -=not uttered, Pr=preverbal, Po=postverbal, C=clitic, N=noun, P=strong pronoun.

a post-verbal noun or a pre- and post-verbal strong pronoun. However, the ungrammaticality of all these different combinations appears to always be the same: the absence of a clitic. Let's now see in Table 5 how this new appellation fits in with the grammatical subjects uttered by the same children.

Types ¹¹	Examples	Glosses	N	%
+PrC	Elle est grande	“She is tall”	459	42.8
+PoC			0	0.0
+PrN			0	0.0
+PrN +C	Le renard il a mangé	“The fox he has eaten”	16	1.5
+C +PoN	Il est là la galette	“It is there the pie”	30	8.1
+PrP +C	Lui i(l) s'appelle Raphaël	“Him he is called Raphaël”	87	2.8
+C +PoP	Je peux jouer moi ?	“I can play me?”	64	6.0
+PrP +C +PoP	Moi j'ai un grand vélo moi	“Me I have a big bike me”	6	0.5
NI	attends !	“ wait!”	170	15.9
Subtotal for grammatical subjects			832	77.6
Subtotal for ungrammatical subjects			240	22.4
Total entire corpus			1072	100.0

Table 5: The different types of grammatical ‘subjects’ in our corpus

The above data allow us to observe a perfect complementary distribution between the absence of a clitic in the ungrammatical sentences in Table 4 and its recurrent presence in all the types of grammatical sentences in Table 5¹². The fine-grained appellation unveiled thanks to the first set of data is hence confirmed: the nature of the null element reduces to clitics. In parallel, Table 5 puts forward another very interesting characteristic of this child system: the total absence of preverbal nouns uttered on their own. Indeed, preverbal nouns always appear with an accompanying clitic. It can also be noticed that, even under this condition, these preverbal nouns remain very rare (16 sentences out of 1072). The pattern ‘noun + finite verb’ –considered as the canonical template for French written clauses– is hence not part of what a child under 3;0 utters spontaneously. This observation from child data is consistent with Blanche-Benveniste (1994, 2003)’s work on adult oral French. Indeed, Blanche-Benveniste (1994) for instance forwards a rate of 5.9% of preverbal nouns in her oral data and an additional interesting observation that most of these preverbal nouns appear in subordinate clauses. Other studies on oral French, such as Jeanjean (1981), had already given these directions. So, as adults do not produce many preverbal nouns either, it appears that the template ‘noun + finite verb’ is mainly applied only in written French. This is another investigation field that we have to leave aside here.

¹¹ Same codes as in Table 4 adding NI=null imperative.

¹² With the expected absence of a subject of any type in imperative clauses.

These results for child oral data lead us to assume that the bounds between what is grammatical and what is ungrammatical for children under 3;0 lies in the dichotomy ‘presence of a clitic’ vs. ‘absence of a clitic’. According to our data, the phenomenon we are talking about is not a null *subject* stage but rather a null *pronoun* stage and more particularly a null *clitic* stage as far as French is concerned. This is also why we favour the appellations *Pro-drop* Parameter and *pro-drop* or non *pro-drop* languages instead of the possible null *subject* appellations¹³.

2. Our analysis of this null pronoun phenomenon

2.1 A few words on input

We propose an analysis of our data in three steps: study of the children’s verbal system, study of their pronominal system and crossing of these first two analyses. This study hence mainly deals with the children’s output. However, input also has to be considered. Indeed, it must be noted that the data a child hears daily do not always conform with the French minus value of the *Pro-drop* Parameter. Micro-variation exists¹⁴ and French displays overt pronouns as well as grammatical null pronouns, particularly in all the imperative sentences and in some oral expressions with defective verbs such as *falloir* ‘have to’. A few examples of these grammatical null pronouns are displayed in (4) hereafter.

(4) Grammatical null pronouns in adult French:

- a. All the imperative verbs, whatever conjugation type the verb belongs to:
 - parler ‘talk/speak’:
 - _ parle ‘speak 2sg’
 - _ parlons ‘let’s speak’
 - _ parlez ‘speak 2pl’
 - faire ‘do/make’:
 - _ fais ‘do 2sg’
 - _ faisons ‘let’s do’
 - _ faites ‘do 2pl’
- b. Some expressions with defective verbs, such as *falloir* ‘have to’, both in matrix and in embedded clauses:
 - _ faut faire attention
 - ‘(one) has to be careful’
 - Quand _ faut y aller, _ faut y aller
 - ‘when (one) has to go, (one) has to go’¹⁵

These examples show that input does not constantly provide clear-cut information to the child with regard to the value of the *Pro-drop* Parameter and that

¹³ It has to be noticed that the English literature displays a choice whereas French only displays one translation, i.e. ‘Paramètre du *Sujet* Nul’.

¹⁴ See Kayne (1996) and Oliviéri (2004).

¹⁵ Example from Haegeman (1997).

this duality could hence be considered as misleading for a young child. Let's now turn to the children's output.

2.2 The children's verbal system

Table 6 presents an overview of the children's verbal system. All their verbs from all their different types of sentences –with and without a clitic– are noted and taken into account here. The reason for this lies in the fact that we found no difference between the verbs uttered with a clitic and those uttered without a clitic. Consequently, we could not follow a hypothesis such as Lorusso et al. (2005)'s study for early Italian, who, for this pro-drop language, establish a link between the presence of a subject and the type of verb in the sentence. Table 6 hereafter displays the list of the main verbs uttered in the corpus as well as their occurrence rates and their conjugation types.

Verbs	Glosses	%	Conjugation types	%
Etre	“be”	48.7	3 rd and être/avoir	94.9
vouloir	“want”	17.2		
faire	“make/do”	16.2		
avoir	“have”	7.0		
savoir	“know”	3.6		
voir	“see”	2.2		
manger	“eat”	1.5	1 st	5.1
rouler	“roll”	1.0		
jouer	“play”	0.9		
tomber	“fall”	0.7		
cacher	“hide”	0.5		
marcher	“walk/work”	0.5		
		100.0		100.0

Table 6: An overview of the children's verbal system

This table sheds light on a very interesting fact. Indeed, it can be noticed that nearly all the verbs used by these children –94.9% of them, in the upper part of the table– belong to the French third-conjugation type. This fact, in itself, is not really surprising despite the fact that we know that 90% of the French verbs belong to the first-conjugation group¹⁶. Indeed, this latter figure does not mean that 90% of the verbs uttered by adults in a conversation belong to this group and it also has to be considered that some of the verbs of the third-conjugation type are extremely frequent ones. Rather, we would like to focus the attention on the fact that third-conjugation verbs along with the verbs *être* “be” and *avoir* “have”, in their present tense paradigms, in French, display more phonetically different forms than first-

¹⁶ See Carelli et al. (1996).

group verbs do. Table 7 hereafter details these present tense paradigms and the number of phonetically different forms within each paradigm (in bold print): *être* and *avoir* display five different forms out of six, *vouloir*, *faire* and *savoir* have four, whereas *voir* and first-conjugation verbs only display three phonetically different forms out of six. This observation is interesting when we recall that the type of verb that is usually referred to when considering French as a non pro-drop language is the first-conjugation type. Indeed, Pierce (1992), Haegeman (1997), Jakubowicz and Rigaut (1997) or Rizzi (1997), among others, all refer to the first-conjugation verb *parler* “talk”.

	être	vouloir	faire	avoir	savoir	voir	1 st conj.
1sg	suis	veux	fais	ai	sais	vois	mange
2sg	es	veux	fais	as	sais	vois	manges
3sg	est	veut	fait	a	sait	voit	Mange
1pl	sommes	voulons	faisons	avons	savons	voyons	mangeons
2pl	êtes	voulez	faites	avez	savez	voyez	mangez
3pl	sont	veulent	font	ont	savent	voient	mangent
%	48.7	17.2	16.2	7.0	3.6	2.2	5.1
≠ forms	5	4	4	5	4	3	3

Table 7: The entire paradigms of the main verbs used by the children

This first step sheds light on two characteristics of this verbal system: (i) it displays nearly no first-conjugation verbs –only 5.1% of the data– and (ii) most of the verbs used by the children display four or five different forms out of six when their paradigms are complete –for 92.7% of the verbs. These peculiarities, compared to the adult system of the same language, lead us to approach the child French verbal system with a pro-drop verbal system, such as Italian. In fact, the French and Italian paradigms for the verb “be” –as shown in (5) hereunder– are quite relevant of this fact. Indeed, the bold print emphasizes that this verb displays five different forms out of six in both languages.

- (5) être: **suis**, **es** [e], est [e], **sommes**, **êtes**, **sont**
 essere: **sono**, **sei**, è, **siamo**, **siete**, sono

Thus the French verb *être* stands on a par with the Italian verb *essere* as far as rich verbal morphology is concerned. And bearing in mind that the French children use this verb in nearly half of all their sentences (48.7%), it would henceforth appear very contradictory to consider the children’s verbal system the same way as the adult system, i.e. as a system with poor verbal morphology. Rather, our first two conclusions are that: (i) the child verbal system is different from the adult verbal system, and (ii) the verbal system we are investigating here bears similar features to a pro-drop verbal system such as Italian.

2.3 The children's pronominal system

It would be interesting now to know if the children use all the different verbal forms displayed in Table 7 or only some of them. We hence turn to the second step of our analysis and investigate the children's pronominal system. Table 8 hereafter shows an overview of our study. It indicates which verbal forms the children used during our recordings –1sg, 2sg, 3sg, 1pl, 2pl and 3pl– and with which pronoun each verbal form was uttered. For 3sg forms for instance, the range of possible pronouns is: referential *il* “he”, expletive *il* “it/one/there”, *elle* “she”, *ce/c'* “it”, *ça* “this/that”, or *on* “one/we”. Both the overt pronouns sentences as well as the null pronoun sentences are included in this table. This means that, when a pronoun is null in the data, it is reconstituted according to context in order to classify all the verbal sentences. A null pronoun sentence such as * *_est vert* “(it) is green” for instance is part of the 23.4% of *c'* sentences and hence part of the 60.9% of 3sg verbal forms.

Pronouns	Glosses	%	Forms	%	%
je	“I”	35.0	1sg	35.0	98.8
tu	“you”	2.9	2sg	2.9	
il	“he”	23.4	3sg	60.9	
il expl	“it/one/there”	2.2			
elle	“she”	4.3			
c'	“it”	23.4			
ça	“this/that”	2.5			
on	“one/we”	5.1			
nous	“we”	0.0	1pl	0.0	1.2
vous plu	“you plural”	0.0	2pl	0.0	
vous pol	“you polite”	0.0			
Ils	“they masc”	0.7	3pl	1.2	
elles	“they fem”	0.5			
		100.0		100.0	100.0

Table 8: Overview of the children's pronominal system

Within the six possible verbal forms, thirteen different pronouns are detailed. The above rates unveil the following characteristics. Firstly, this system can be described as being mainly singular. Indeed, Table 8 shows that only a mere 1.2% of the clauses include a plural form. Additionally, we can see that not all of the plural forms are uttered by the children at this stage as the first and second plural pronouns are never used. Moreover, it has to be noted that these plural forms all come from the same child thus the feature [±Singular] does not really apply in the average system yet. Secondly, within the singular forms, it can also be noticed that not all of the different forms are uttered. Indeed, most of this system –95.9%– is composed of

two main different forms only: the first singular form –with the pronoun *je*– representing 35% of the sentences and the third singular form –with different pronouns– found in 60.9% of the clauses. The second singular form hence shows a very low occurrence rate of 2.9%. This system can consequently be characterized as displaying one single opposition between two singular forms: 1sg –which represents the feature [+Speaker]– vs. 3sg. –which can be defined by opposition with the feature [–Speaker].

As for the verbal system studied in the previous section, the data clearly show that the children’s pronominal system is not identical to the adult’s one. Indeed, even if we take into consideration the fact that, in oral French, the first plural form with the pronoun *nous* is often replaced with the third singular form with the pronoun *on*, adults use the second singular and third plural forms whereas children do not at this stage.

2.4 Crossing the analyses

What happens now if we cross the information gathered in Tables 7 and 8 above? We obtain Table 9 that displays the actual occurring verbal forms, listing the main verbs uttered by the children but reduced this time to their two main forms, i.e. 1sg and 3sg.

	être	vouloir	faire	Avoir	savoir	Voir	manger
	“be”	“want”	“do”	“have”	“know”	“see”	“eat”
1sg	suis	veux	fais	ai	sais	Vois	Mange
3sg	est	veut	fait	a	sait	Voit	Mange
%	48.7	17.2	16.2	7.0	3.6	2.2	1.5

Table 9: The actual occurring verbal forms in the data

The bold print in Table 9 highlights the fact that the verbs *être* and *avoir* –which appear in a total of 55.7% of the children’s sentences– display phonetically different forms in this system. This amounts to say that the 1sg and the 3sg forms for these two verbs display rich verbal morphology insofar as the oppositions between *suis* and *est* on the one hand and *ai* and *a* on the other hand are totally distinctive. This pattern is then exactly the same as the one traditionally found in pro-drop languages such as Italian, which presents, for the same verbs, the forms *sono* vs. *è* and *ho* vs. *ha*. Consequently, we argue that such verbal forms are reminiscent of the mechanism that licences null pronouns in pro-drop languages. So when a child like Alizée (2;8) says *_ suis là* “(I) am here”, we claim that, within the child’s system, this type of sentence finds an explanation in the fact that the only other form that the child uses at this stage is distinctive, i.e. it does not require the presence of a pronoun in order to be identified as the 1sg form of the verb *être*¹⁷. In a nutshell, it can be said that part of the child’s verbal system is similar to a pro-drop system.

¹⁷ This point is also made in Van Kampen (2006), with a different analysis though.

Consequently, the architecture of the French child verbal system does not yet correspond to the one found in the adult system, insofar as the former displays richer verbal morphology than the latter. We hence argue that this is precisely why the children's utterances include more null pronouns than the adult's ones and we identify this initial duality between the correct value of the parameter and the reality of the child system as the trigger to the null pronoun phenomenon.

This hypothesis is strengthened by the fact that, for some of the poor verbal morphology verbs such as *vouloir* "want", the two-person paradigm described above nearly reduces to a one-person system. Indeed, the data show that the verb *vouloir* for instance, when uttered with a null pronoun, is in 59 sentences out of 60 in the 1sg form –the only 3sg occurrence being *oui, _ veut voler* "yes, (she) wants to fly" (Raphaël, 2;6, talking about his friend Alisson). There is hence hardly no opposition in this paradigm. This observation is thus consistent with our hypothesis that the trigger to the phenomenon lies within the morphological richness of the child's reduced verbal paradigms. Moreover, the behaviour of the verb *vouloir* parallels the one of a defective verb such as *falloir* "have to". This latter verb does not appear in Table 6 as it is uttered by one single child –Raphaël. However the pattern is similar as, when it is used by Raphaël –11 times in the whole corpus–, it appears 9 times without a pronoun, such as in the sentence *_ faut mettre les palmes* "(one) has to wear the flippers". This verb is a defective verb, i.e. it is exclusively used in the 3sg form: *il faut* "one has to". Again, this observation is consistent with our analysis: this paradigm presents only one form so the presence of the pronoun is not necessary to identify the person. This is also consistent with the adult occurrences of this defective verb since *falloir* is also often used orally by adults without a pronoun.

3. Conclusion

The well-known child null 'subject' phenomenon is reconsidered here thanks to our new data and a different approach associating the generative framework with a dialectological point of view. Firstly, an in-depth investigation into the ungrammatical sentences uttered by the children allows us to forward a more fine-grained appellation to this phenomenon. Indeed, the data show that the missing element is not any type of subject but only pronouns, and more precisely clitics as far as French is concerned. We hence identify this phenomenon not as a null 'subject' phenomenon but as a null 'pronoun' one (cf. Section 1.5). Secondly, our methodology invites us to investigate the entire verbal and pronominal systems of the children. The outcome is that the children, at this particular stage, manipulate a linguistic system which is different from the adult system. Indeed, our data highlight that the former system is reduced compared to the latter in terms of different verbal forms as the children's paradigms mainly display only two forms –1sg vs. 3sg. Many of these verbal forms hence turn out to be distinctive –e.g. *suis* "(I) am" vs. *est* "(he) is"– and consequently do not require the presence of a pronoun to be identified. The child system presents a richer verbal morphology than the adult and this characteristic is reminiscent of the rich verbal morphology that is found in pro-drop languages. Consequently, we claim that it is this particular characteristic of the

child linguistic system that represents the trigger to the null pronoun phenomenon in his reduced system. However, this hypothesis only represents one piece of the whole puzzle so the last part of this conclusion aims at placing it correctly and coherently together with the rest of the picture.

Firstly, the children fix the value of the Pro-drop Parameter correctly, during their first months, thanks to the data they receive daily from their surroundings. This corresponds to what Wexler calls the Very Early Parameter Setting (VEPS) and this correct fixing of the parameter forwards an explanation to the approximately 75% of the children's grammatical sentences with regard to the presence of a pronoun. However, the VEPS hypothesis does not provide an explanation to the 25% of the sentences that do not display a pronoun. The picture is hence incomplete for the moment.

Secondly, on the one hand, this value of the Pro-drop Parameter corresponds to the adult verbal system that displays six different forms for most of the verbs and a majority of verbal paradigms that present a poor morphology. On the other hand, this value does not yet fully correspond to the architecture of the child reduced system insofar as this latter system displays two-person paradigms and a majority of distinctive forms. We hence claim that this discrepancy between the child and the adult systems is the trigger to the null pronoun phenomenon, i.e. the child drops some of his pronouns because they are not as necessary as in the adult system. This hypothesis hence explains not only the phenomenon in itself but also gives a clue to its occurrence rate of 25%. Indeed, our data show that not all the children's verbs provide this rich verbal morphology so the impact of this characteristic is not pervasive. So far our analysis answers the first three points mentioned in Section 1.3, i.e. (i) explain the null pronoun phenomenon, (ii) explain its concomitancy with overt pronouns and (iii) justify its occurrence rate. We also mentioned –in Section 1.2– that our analysis of the non-finite verbs in our corpus was, in a way, consistent with Wexler's hypothesis on Optional Infinitives. Indeed, Wexler assumes that there are Optional Infinitive languages, such as French, and non-Optional Infinitive languages, such as Italian. We have just argued that French children display a verbal system that is close to the one found in Italian for instance. It then follows that, within such a verbal system, no genuine Optional Infinitives should be found.

An additional advantage of our hypothesis is that it can be extended to other languages. Indeed, Benveniste (1946) points out the fact that the 3sg forms are different from the other forms in many languages throughout the world. English with its 3sg *-s* morpheme is a very relevant example. Null pronoun sentences¹⁸ such as *is broken* (Naomi), *get down* (Peter) or *goes front* (Peter) could not require a pronoun if the system were reduced to the 1sg and the 3sg forms. And if it turned out that the children worldwide, at this stage, manipulated a reduced two-person system, it would follow that, in many cases, the 3sg form would be distinctive in such reduced systems. Consequently, our hypothesis is trans-linguistic and this entails that it might shed light on certain aspects of UG.

¹⁸ These examples come from Pierce (1992).

We are now left with our aim (iv) of Section 1.3 concerning the disappearance of this null pronoun phenomenon. So far, we have often called it a ‘phenomenon’. However, it has to be noted that it only represents a stage in the evolution of the child’s linguistic development as its disappearance is usually attested around 3;0. An important aspect of a hypothesis on this phenomenon thus has to include an explanation to this disappearance. We hypothesize that the null pronoun phenomenon fades away as the child grows out of his reduced verbal and pronominal systems: (i) verbwise, i.e. when the child acquires additional verbs, especially verbs from the first conjugation type that do not display as many different forms in the present paradigm and (ii) pronounwise, i.e. when the child uses additional forms such as the 2sg form which is often phonetically similar to the 1sg form in French. On the whole, the child leaves this null pronoun stage when his linguistic system gets closer to the adult’s system, i.e. when the discrepancy between the two systems –which is identified as the trigger– fades away.

References

- Benveniste, E. (1946) ‘Structure des relations de personne dans le verbe’, in: *Problèmes de linguistique générale, 1*. Paris: Editions Gallimard, 1966, 225-236.
- Blanche-Benveniste, C. (1994) ‘Quelques caractéristiques grammaticales des "sujets" employés dans le français parlé des conversations’, in: M. Yaguello (ed.) *Subjecthood and Subjectivity. Proceedings of the Colloquium "The Status of the Subject in Linguistic Theory"*. Gap, Paris and Londres: Editions Ophrys, 77-107.
- Blanche-Benveniste, C. (2003) ‘Les formes grammaticales de réalisation des sujets et leur inégale représentation en français contemporain’, in: J.-M. Merle (ed.) *Le sujet*. Paris: Editions Ophrys, 73-90.
- Cardinaletti, A. and M. Starke (1994) ‘The Typology of Structural Deficiency: A Case Study of the Three Classes of Pronouns’, in: H. van Riemsdijk (ed.) *Clitics in the Languages of Europe*. Berlin and New York: Mouton de Gruyter, 1999, 145-233.
- Carelli, E. et al. (1996) *Le Robert and Nathan conjugaison*. Paris: Editions Nathan.
- Chomsky, N. (1981) *Lectures on Government and Binding: The Pisa Lectures*. Dordrecht: Foris.
- De Cat, C. (2002) *French Dislocation*. PhD Thesis, University of York.
- De Cat, C. (2005) ‘French Subject Clitics are not Agreement Markers’. *Lingua*, 115, 9, 1195-1219.
- Haegeman, L. (1997) ‘Adult Null Subjects in Non Pro-drop Languages’, in: M.-A. Friedemann and L. Rizzi (eds.) *The Acquisition of Syntax. Studies in Comparative Developmental Linguistics*. London: Longman, 2000, 129-169.
- Hyams, N.M. (1986) *Language Acquisition and the Theory of Parameters*. Dordrecht: D. Reidel Publishing Company.
- Hyams, N.M. (1996) ‘The Underspecification of Functional Categories in Early Grammar’, in: H. Clahsen (ed.) *Generative Perspectives on Language Acquisition*. Amsterdam and Philadelphia: John Benjamins, 91-127.
- Jakubowicz, C. and C. Rigaut (1997) ‘L’acquisition des clitiques nominatifs en français’, in: A. Zribi-Hertz (ed.) *Les pronoms : morphologie, syntaxe et typologie*. Saint-Denis: Presses Universitaires de Vincennes, 57-99.

- Jeanjean, C. (1981) 'L'organisation des formes sujets en français de conversation : étude quantitative et grammaticale de deux corpus'. *Recherches sur le français parlé*, 3, 99-136.
- Kampen, J. van (2006) 'Early Operators and Late Topic-drop/Pro-drop', in: V. Torrens and L. Escobar (eds.) *The Acquisition of Syntax in Romance Languages*. Amsterdam: John Benjamins, 203-223.
- Kayne, R.S. (1996) 'Microparametric Syntax. Some Introductory Remarks', in: R.S. Kayne (ed.) *Parameters and Universals*. Oxford: Oxford University Press, 2000, 3-9.
- Lightbown, P. (1977) *Consistency and Variation in the Acquisition of French*. PhD Dissertation, Columbia University.
- Lorusso, P. et al. (2005) 'Overt Subject Distribution in Early Italian Children', in: A. Brugos et al. (eds.) *Supplement to the Proceedings of the 29th Boston University Conference on Language Development (BUCLD)*. Boston, MA.
- MacWhinney, B. (2000) *The CHILDES Project. Tools for Analyzing Talk, Third Edition*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Oliviéri, M. (2004) 'Y a-t-il des frontières dialectales en syntaxe ?'. *Etudes corses*, 59, 77-93.
- Palasis-Jourdan, K. (2005) *Problèmes d'acquisition et le Paramètre du Sujet Nul*. Mémoire de Master, Université de Nice-Sophia Antipolis.
- Pierce, A. (1992) *Language Acquisition and Syntactic Theory: A Comparative Analysis of French and English Child Grammars*. Dordrecht: Kluwer Academic Publishers.
- Rasetti, L. (1996) 'Null Subjects and Root Infinitives in the Child Grammar of French'. *Geneva Generative Papers*, 4, 2, 120-132.
- Rizzi, L. (1993) 'Some Notes on Linguistic Theory and Language Development: The Case of Root Infinitives'. *Geneva Generative Papers*, 0, 2, 102-114.
- Rizzi, L. (1997) 'A Parametric Approach to Comparative Syntax: Properties of the Pronominal System', in: L. Haegeman (ed.) *The New Comparative Syntax*. Harlow: Addison Wesley Longman, 268-285.
- Rizzi, L. (2002) *On the Grammatical Basis of Language Development: A Case Study*. Ms., University of Siena.
- Serratrice, L. (this volume). 'Null and overt subjects at the syntax-discourse interface. Evidence from monolingual and bilingual acquisition'.
- Suppes, P. et al. (1973) 'The French Syntax of a Child's Noun Phrases'. *Archives de Psychologie*, 42, 207-269.
- Wexler, K. (1992) 'Optional Infinitives, Head Movement and the Economy of Derivation in Child Grammar'. *Occasional Paper*, 45, Center for Cognitive Science, Cambridge, MA: MIT.
- Wexler, K. (1998) 'Very Early Parameter Setting and the Unique Checking Constraint: A New Explanation of the Optional Infinitive Stage'. *Lingua*, 106, 1-4, 23-79.

The Acquisition of the Morphosyntax-Pragmatic Interface in French L1

Evidence from reference with articles and pronouns

Margot Rozendaal

Amsterdam Center of Language and Communication, University of Amsterdam

1. Introduction

One of the first uses to which children put language is reference. Reference can be defined as the use a linguistic expression to identify a person or object (Brown and Yule 1983). This is a crucial part of language acquisition, since it is the only way for children to make clear what they want to communicate without having to resort to deictic gestures, such as pointing. The linguistic subfields of morphosyntax and pragmatics are both closely involved in reference. The linguistic form that speakers choose to refer depends on various pragmatic aspects, for example on whether speaker and hearer are both familiar with the referent and on whether the referent is new or given in the current discourse. Thus, in French, an indefinite article indicates newness and cannot be used for referents that are already given in discourse. In that case, a definite noun or pronoun is used (see example 1).

- (1) Ma collègue a vu un comédien célèbre courant le marathon. *Le comédien/il/*
**un comédien* était en tête de la course.
'My colleague saw a famous comedian running the marathon. The comedian/he/
*a comedian was in the leading group.'

In language acquisition children are thus faced with a double task. They must not only learn the morphosyntactic forms for reference in their mother tongue, but also learn when to use these forms in relation to the pragmatics of reference. Earlier research on the interaction between the morphosyntax and pragmatics of reference in language acquisition has found that up until at least six years-of-age children experience difficulties in several aspects of pragmatic language use, such as taking into account the listener's perspective (Hickmann 2003; Power and Dal Martello 1986). However, early sensitivity to the (morpho)syntax-pragmatics interface has also been demonstrated for two-year-olds, for example in the field of object drop (Serratrice 2005).

The current study will address the morphosyntax-pragmatics interface of indefinite articles, definite articles and pronouns in the language of two- and three-year-old French children and in the input addressed to them. I will focus on how children, who are in the process of acquiring these forms, use these in relation to different pragmatic aspects. It is assumed that these pragmatic aspects have cognitive underpinnings, some of which are present before the start of article and pronoun use, for example the ability to distinguish whether something is new or given. It is, however, not known whether children are able to apply their cognitive

knowledge to linguistic reference from the moment that they start using morphosyntactic forms. Moreover, it is unclear whether children make the same associations between morphosyntax and pragmatics in articles and pronouns. The current paper tries to fill these gaps. French is an interesting language to investigate these questions, since French children are relatively young when they start using articles, especially in comparison to children acquiring Germanic languages (Chiercha, Guasti and Gualmini 2001). The question is however, whether they use articles and pronouns in similar ways as adults do in the input.

The paper starts with a description of the pragmatic aspects that play a role in the use of linguistic forms in adult French. Next, we will briefly describe earlier research on the acquisition of the morphosyntax-pragmatics interface in French and other languages. Finally, the method and results of the current research will be presented, followed by a discussion of the results.

2. Background

2.1 The morphosyntax-pragmatics interface in adult French

As was briefly mentioned above, indefinite articles, definite articles and pronouns cannot always be used in the same pragmatic contexts, since these forms convey different levels of accessibility or givenness (Ariel 1996; Gundel 1996). Indefinite articles indicate that the referent is highly inaccessible and new, whereas (zero-) pronouns¹ form the other end of the accessibility scale: they are used for referents that are in focus and given. Several aspects influence the accessibility or givenness of a referent. Three pragmatic aspects are investigated in the current study, these are: (1) whether the referent is specific or non-specific to the speaker (based on the speaker's familiarity with the referent, Lyons 1999: 173), (2) whether the referent is new or has already been mentioned in the discourse, (3) whether the referent is familiar to the listener (based on mutual knowledge / perceptual availability).

The first distinction is between non-specific reference and specific reference, based on whether the speaker (presumably) does not have or does have a particular entity in mind. Specificity is thus interpreted as a pragmatic notion in this research and not as a semantic one as in, for example, Enç (1991). In example (2)² the

¹ In French a difference can be made between full pronouns and subject / object clitics. According to van Kampen (2004), clitics do not refer deictically but depend on discourse instead, but see Van Kampen and Pinto (this volume). On the basis of their common property of high accessibility to the listener however, full pronouns and clitics are combined in the current research. They will be henceforth referred to as 'pronominal forms' or 'pronouns' here. A different distribution over pragmatic functions of these forms is however not excluded and needs to be investigated further.

² Examples will come primarily from the input to the children or the children's language production whose language is analyzed in this research (CHILDES). Target references are indicated with italics. The examples are presented as they occur in the CHILDES-database as much as possible, including hesitations, self-corrections etc.

discourse context indicates that the speaker does not (yet) have particular necklaces in mind. She is referring to necklaces that might be made. The entity is not (yet) familiar to the speaker, nor to the hearer, who can therefore only identify the type of entity described. This is termed here ‘non-specific reference’. In this research other typical examples of non-specific reference include proposals to construct something and reference to one instance out of many. Non-specific reference is usually indicated with an indefinite article in adult French, although definite articles can also be used, especially in generics (Lyons 1999).

(2) Non-specific reference (input to Anne, 3;3)

- %sit: The investigator is asking Anne what she usually does at school.
- *INV: Ah tu fais des petites perles?
‘Oh, you make little beads?’
- *INV: Tu fais *des colliers* alors?
‘And do you then make necklaces?’

If the discourse context indicates that the speaker has a specific entity in mind, the form might be new or given in the current discourse (see Table 1 at the end of this section). A referent that is mentioned for the first time in the current discourse is termed ‘discourse-new’. Form choice then depends on the familiarity of the referent to both speaker and hearer. There might be ‘mutual knowledge’ (MK) about a referent between speaker and hearer on the basis of, amongst other things, shared (world) knowledge (‘the queen’) or inference (‘the wheels of a car’) (Clark and Marshall 1981). For MK-referents both indefinite and definite articles can be used to mention a referent for the first time. An example is given in (3) where the mother refers back to a mutually known elephant that she and the child saw in the zoo some time ago.

(3) Specific discourse-new referent with mutual knowledge (input to Philippe 2;3)

- %sit: Philippe is showing his mother a toy elephant and asks her for his trunk.
- *MOT: Il a un nez très très long, ça s' appelle une trompe.
‘He has a very long nose, that is called a trunk.’
- *MOT: Tu te souviens, tu l' as vu au jardin un jour *le gros gros elephant* qui mangeait avec son [//], sa trompe.
‘Do you remember, you have seen a very, very big elephant once in the zoo, who ate with his trunk.’

If on the other hand there is ‘no mutual knowledge’ (NMK), an indefinite article must be used to indicate the newness of the referent to the hearer, as is shown in example (4). It is possible for the hearer to become familiar with the referent, because it is physically present and therefore perceptually available. As long as there is shared attention between speaker and hearer for a perpetually available referent, pronouns can be used to introduce the referent in the discourse. This is the case in example (5) where the mother refers to hot coffee with a demonstrative pronoun.

- (4) Specific discourse-new referent, no mutual knowledge (Philippe 3;3)
 %sit: the investigator is asking Philippe about Mother's Day.
 *INV: A l' école tu avais pas préparé quelque chose?
 At school, didn't you make something for her??
 *CHI: Si, *un carnet*.
 'Yes, a booklet.'
- (5) Specific discourse-new referent that is perceptually available to the hearer (input to Philippe, 2;3)
 %sit: Philippe is very interested in the steam (which he calls 'smoke') coming from a hot drink.
 *CHI: c' est la fumée +...
 'It is smoke...'
 *MOT: C' est la fumée parce que c' est chaud.
 'It is smoke because it (the coffee) is hot.'

Referents that have already been mentioned before in discourse are termed 'discourse-given'. Indefinite articles cannot be used for given referents, since this form indicates newness. Furthermore, form choice for given referents is influenced by the syntactic closeness between a subsequent mention and its previous mention (Wilson and Sperber 2004). In this research, a subsequent mention that is similar to the immediately previous mention is termed 'maintenance'. Pronouns are the most felicitous forms for maintenance, since the use of a nominal form can be seen as redundant (compare example 6a and 6b). However, if another referent intervenes between the two subsequent mentions ('referent shift'), definite nouns are equally appropriate or even better than pronouns (see example 7).

- (6) a. Discourse-given-maintenance (input to Grégoire, 2;3)
 *MOT: Pour de vrai comment il te dit Adrien?
 'Seriously, how does Adrien call you?'
 *MOT: *Il*_i t' a inventé un tout gentil petit nom qu'il te donne.
 'He came up with a very nice name that he gave you.'
- b. Discourse-given – maintenance (Input to Grégoire, 2;3)
 *MOT: Moi je l'aime bien ton petit avion_i.
 'I like your little airplane very much.'
 *MOT: Cherche *ton petit avion*_i.
 'Go find your little airplane.'
- (7) Discourse-given – shift (input to Léa, 3;3)
 %sit: Léa is playing in the sink, with various objects.
 *CHI: Je pense que on va un peu laver la tasse de café_i.
 'I think that I will just wash the coffee cup.'
 %sit: she reaches for it after speaking, but her grandmother takes it away.
 *GRM: Ah non pas [] pas dans l' eau que [//] où tu viens de te laver les pieds.
 'Oh no, not in the water where you have just washed your feet.'

*GRM: Hein, Léa! Maman mettra *la tasse de café*; # euh dans le lave-vaisselle, hein.
 Hey, Léa! Mummy will put the coffee cup in the dish washer.’

Table 1 gives an overview of how indefinite articles, definite articles and pronouns can be used for the pragmatic aspects investigated in adult French.

The main question in this paper is if young children, who are in the process of acquiring morphosyntax, are also sensitive to these pragmatic aspects in their use of indefinite articles, definite articles and pronouns. Moreover, form-function associations in the input will be examined to establish the evidence that children get from the input.

	indefinite article	definite article	pronoun
Non-specific reference	+	+	-
Specific Reference			
Discourse-new NMK	+	-	-
MK—not perceptually avail.	+	+	-
MK – perceptually avail.	+	+	+
Discourse-given Maintenance	-	-	+
Shift	-	+	+

Table 1: Relevant form-function combinations for this research
 += form is correct for function; -= form is incorrect for function.

2.2 Studies on the development of the morphosyntax-pragmatics interface

The acquisition of the morphosyntax of reference has been widely studied in French L1. The earliest forms of articles appear to be filler syllables that are often considered to be proto-articles (Veneziano and Sinclair 2000). The development of article use in obligatory contexts proceeds fast in French. Several authors report that French children provide articles in around 80% or more of the obligatory contexts at 2;6 (e.g. Bassano, Maillochon and Mottet 2005; Pannemann 2006; Van der Velde, Jakubowicz and Rigaut 2002). The first pronominal forms are produced around 2;0. Furthermore, many studies have found that the acquisition of subject clitics precedes that of object clitics, which are generally produced from 2;6 onwards (Hamann, Rizzi and Frauenfelder 1996; Van der Velde et al. 2002). The present research will not take object drop into account. Rather, the focus will be on the association between overt forms and pragmatic functions.

The acquisition of pragmatic properties of articles and pronouns has also been studied for French (e.g. Hickmann 2003). The results indicate that children seem to master some pragmatic properties relatively late. For example, up until at least six years-of-age, they erroneously use definite articles in discourse-new-NMK contexts (see Table 1). Most studies on the morphosyntax-pragmatics interface of reference have been conducted with children older than four years-of-age. However, French

children start using articles and pronouns around two-years-of-age and therefore the investigation of the morphosyntax-pragmatics interface can and should start earlier.

Studies of the development of socio-cognition suggest that infants develop cognitive concepts necessary for the pragmatics of reference before the age of two. Before twelve months they are able to distinguish between stimuli that are, for them, novel or familiar and around fourteen months children are also aware of what is new for others, even if it is not new for the children themselves (Roder, Bushnell and Sasseville 2000; Tomasello and Haberl 2003). Complete understanding that others' minds are separate from one's own is part of the development of a Theory of Mind, which seems to take at least until age four (Ruffman and Perner 2005).

It is not clear whether children use their acquired cognitive knowledge in the pragmatics of reference as soon as they start to use articles and pronouns, although a few studies found evidence for early sensitivity to pragmatics. It has been shown that children associate indefinite articles with non-specific reference in English (Schaeffer and Matthewson 2005). Moreover, in various languages children drop subjects and objects for given referents but use full forms for new referents (Guerriero, Oshima-Takane and Kuriyama 2006; Serratrice 2005, this volume). These findings indicate that young children distinguish between new and given in their choice of linguistic forms. This distinction might then also be apparent in their choice between indefinite versus definite articles and nouns versus pronouns, as the results here will hope to clarify.

Various studies have found that in story telling children have troubles in applying MK and NMK in article production until a late age. However, children seem to be better able to take the interlocutor's knowledge into account in more natural situations (O'Neill 2005). Since this study investigates spontaneous speech, some sensitivity to other's knowledge is therefore expected in the use of articles (MK/NMK) and pronouns (perceptually available/unavailable). Errors are also expected here, however, since taking the listener's perspective into account is part of the development of a Theory of Mind, which continues into the pre-school years.

In sum, in adult French there is a clear pattern of which linguistic forms can be used for the pragmatic aspects of specificity, givenness in discourse and the familiarity of the referent to the listener. Some but not all of the cognitive concepts for these pragmatic aspects are present before two years-of-age. The question is whether children associate these pragmatic aspects with the same linguistic forms as French adults do in the input. Moreover, do children apply the same pragmatic aspect to both articles and pronouns? The current study will thus focus on the following questions:

1. Do children acquiring French distinguish between non-specific/specific reference, new/given in discourse and familiar / not familiar to the listener in the use of indefinite articles, definite articles and pronouns?
2. Do the children show sensitivity or insensitivity to the same pragmatic aspect(s) in using indefinite versus definite articles and in using pronouns versus nouns?
3. How can the children's acquisition behavior be related to the use of linguistic forms for pragmatic aspects in the input?

3. Method

3.1 Subjects

The data were taken from the CHILDES-database (MacWhinney 2000) and needed to satisfy two criteria. Firstly, the transcripts had to include speech from both conversational partners, as this allows us to track the diverse pragmatic aspects (non-specific reference, givenness etc.). Secondly, the transcripts needed to be coded for external, non-linguistic events and context information to facilitate the coder's analysis of the pragmatic aspects. Video-recordings of the conversation would have facilitated the interpretation even more. Unfortunately, these tapes did not exist for all CHILDES-data or were not available through CHILDES at the time of coding. Therefore, videos have not been used in this research

Data from four monolingual children acquiring French were taken between 2;0 and 3;3 (Anne and Léa from the York-corpus, Grégoire from the Champaud-corpus, Philippe from the Leveillé-corpus). The data were analyzed with three-monthly intervals, at 2;0, 2;3, 2;6, 2;9, 3;0 and 3;3. For the child Philippe there were no data available before 2;3. Furthermore, at the time of coding there was no third French child available in CHILDES whose data covered the total age range 2;0-3;3. Therefore a mixed longitudinal-cross-sectional design was used to obtain more data. Grégoire's data range from 2;0-2;6 and Léa's data range from 2;9-3;3. For each child a sample of input language of a (grand)parent and/or an investigator was analyzed at the ages of 2;3 and 3;3.

Since a small number of children are studied, it was important to know that the three children per language group fall within the normal range of general linguistic development. The subjects' MLU in words (MLU_w) was therefore compared to the MLU_w of a group of ten children of the same age. The subjects were comparable to the norm group: at only one time point in one child was there an MLU_w more than 1 s.d. from the norm (Grégoire 2;6). Since the language level of the four children is rather similar, the data from the different children will be pooled per age point in the analyses reported here.

3.2 Analysis

For each child a sample of 600 utterances was analyzed in order to achieve a similar amount of discourse diversity across the different children. If there were no or insufficient data available at exactly the target age, as many additional data as needed were used from recordings made within one month before or after the target-age. In no case were there more than five weeks between the different samples of one child for a particular target-age. Moreover, 300 utterances of input language from the samples of each child were coded at both 2;3 and 3;3. The same analysis procedure was used for the child and adult utterances.

Nominal and pronominal references to persons, objects and concrete substances were selected from the utterance sample. References were excluded if they occurred in singing, partly uninterpretable utterances, unfinished utterances and imitations. Generic locations (*dans la cuisine*) or nouns that are part of a (fixed) verbal

construction (*aller en voiture*) were excluded. Noun phrases in such constructions show idiosyncratic behaviour with regard to the presence and type of article. The total proportion of such excluded nominal references is less than 5% of the data.

The morphosyntactic analysis focused on articles and pronouns. In the study reported here, the use of indefinite articles will be contrasted with the use of definite articles for different pragmatic functions. Attributively used demonstrative pronouns are combined with the definite articles in the analysis, since these forms behave similarly with respect to the pragmatic aspects investigated. A category ‘other’ contains ungrammatical bare nouns (**ø chaise*), possessive (*ma etc chaise*), numeral (*deux chaises*) and partitive articles (*du sucre*) as well as fillers (*ə chaise*)³. The analysis of pronouns contains personal, demonstrative, possessive, numeral, relative and reflexive pronouns. In the current analysis, the category of pronouns as a whole, including both full and clitic pronouns, is contrasted with the use of nouns and proper names for pragmatic functions. Table 2 shows the total numbers of indefinite, definite/demonstrative articles, ‘other’ nouns, pronouns and proper names produced by each subject. These data form the basis of all further analyzes.

Child	Indefinite articles	Definite articles	‘Other’ nouns	Pronouns	Proper names	Total
Anne	173 (10%)	367 (21%)	236 (14%)	777 (45%)	180 (10%)	1733
Grégoire	54 (8%)	101 (15%)	229 (35%)	152 (23%)	120 (18%)	656
Léa	95 (13%)	141 (19%)	82 (11%)	353 (47%)	83 (11%)	754
Philippe	320 (18%)	513 (29%)	173 (10%)	658 (38%)	5% (90)	1754

Table 2: Child data used in study

The pragmatic aspects discussed in section 2.1 are the basis for the pragmatic functions analyzed (see also Table 1). The pragmatic function of ‘labelling’ was however added to the analysis, since in child language, utterances in which the speaker categorises, names or identifies a specific entity frequently occur. Labelling often appears after a WH-question from the interlocutor or occurs in a predicating or existential construction (example 8). Also, single word utterances that are not elaborations from a previous utterance or elaborated upon in subsequent utterances by the child are classified as labelling.

(8) Labelling – (Grégoire, 2;5)

*CHI: *Ça c'est un gros camion rigolo.*
‘That is a big, funny truck.’

³ Transcriptions of *e/a* in front of nouns were interpreted as fillers in the data analysis.

Following Table 1, referents were then coded for whether they referred to a non-specific or specific referent. If reference was analyzed as specific, it could then be coded as discourse-given or discourse-new. Referents that are specific and new in discourse can be either MK or NMK to the interlocutors. Moreover, discourse-new referents are coded for whether they are perceptually available or not. Finally, referents that have already been mentioned before in discourse, either by the child or the interlocutor, are coded as discourse-given. A distinction is made between maintenance and shift. In maintenance the target reference refers to the same entity as the reference immediately previously. The referent thus stays in focus. In referent shift, the distance between two references is larger: there are one or more references to other entities in between. For references that occur in dislocated (topic) position, Hickmann's analysis (2003) was followed: the dislocated element is seen as the referential element and is therefore coded morphosyntactically and pragmatically. The resumptive (clitic) pronoun is not analyzed separately.

The author coded the data. To determine the reliability of the coding scheme, a trained research assistant coded 10% of the child data independently. The mean percentage of agreement between the two coders was 85% for pragmatics and more than 98% for morphosyntax.

In the statistical analysis Pearson's chi-square was used to determine whether morphosyntactic forms were used differently for pragmatic functions at each age point. The significance level was set at $p < 0.05$. Chi-square was not calculated if more than 20% of the cells had a value lower than 5, since this reduces the power of the test. To determine the variables that contribute to a significant chi-square value, the adjusted standardized residual ('asr') was used. More specifically, the asr indicates how a particular form is used for a particular pragmatic function relative to other forms for that function and also how this particular form is used for other functions. As such, it indicates if a particular form is associated with one or more pragmatic functions and/or disassociated with others. Asr-scores between 2 and 3 are seen as major contributors to the overall (significant) chi-square value. In this research, asr-scores higher than 2.5 are reported.

4. Results

4.1 The morphosyntax-pragmatics interface in the input

In this section the morphosyntax-pragmatics interface in the input will be discussed. The use of form for functions in the input creates expectations for the morphosyntax-pragmatics interface in child French.

Table 3 shows the use of articles for pragmatic functions in the adult input to the children. The focus lies on the use of indefinite versus definite/demonstrative articles. Other forms are included in the category 'other'. With respect to the new-given distinction, there is no need to distinguish between discourse-given-maintenance and discourse-given-shift in analyzing articles, since for both pragmatic functions indefinite articles cannot be used. Maintenance and shift are therefore combined into one category 'discourse-given' in Table 3. The distinction between MK and NMK is relevant for the speaker's sensitivity to the listener's perspective.

Indefinite articles must be used for NMK, whereas both indefinite and definite articles can be used for MK. There is no need to distinguish between perceptual availability in DP-use, since as long as the referent is mutually known, both indefinite and definite articles are allowed for perceptually unavailable referents.

The figures in bold designate form-function combinations that have a strong association (indicated with >) or strong disassociation (indicated with <) compared to the expectation based on equal distribution of the forms over all pragmatic functions (asr-calculation). For example, the pragmatic function of discourse-given is strongly associated with the use of definite/demonstrative articles (80% of the cases), and disassociated with indefinites (3%).

	indefinite article	def/dem article	'other' noun
Labelling	51% (48)>	35% (33)<	15% (14)
Non-specific reference	62% (49)>	22% (17)<	17% (13)
Discourse-given	3% (6)<	80% (160)>	22% (46)
Discourse-new MK	21% (27)	55% (73)	24% (32)
Discourse-new-NMK	(1)	(0)	(0)

Table 3: Use of articles for pragmatic functions in the French input data. Raw figures are given in brackets. Bold=cell is a major contributor to the significant chi-square value for form-function associations. >=Adjusted standardized residual greater than 2.5, the morphosyntactic form is more strongly associated with this function than other forms and more strongly with this function than with other functions. < has the reverse interpretation.

Table 3 shows that there are clear form-function associations (χ^2 , $p < 0.001$). Indefinite determiners are strongly associated with labelling and non-specific reference and disassociated with discourse-given references. Definite/demonstrative determiners are also used for labelling and non-specific reference, but these are dissociated with these functions, since they are most strongly associated with discourse-given referents. Although the category of discourse-new does not contribute to the overall effect, the adults use significantly more definite determiners than indefinite or 'other' nouns for MK (χ^2 , $p < 0.001$). Finally, the pragmatic function of discourse-new-NMK hardly occurs in the input (n=1). The children thus seem to receive little positive evidence on how to express this pragmatic aspect.

Table 4 shows the use of pronouns for pragmatic functions. The pragmatic aspect of givenness can be investigated by examining the choice of morphosyntactic forms for discourse-given referents versus discourse-new referents (perceptually available and unavailable). Moreover, sensitivity to different *degrees* of givenness is investigated by examining form choice for discourse-given-maintenance versus discourse-given-shift. Sensitivity to the listener's perspective is examined by looking at the use of pronouns to introduce referents in discourse. A distinction is made here between introducing referents that are either physically present or absent. In the latter case the referent is perceptually unavailable to the interlocutor and pronouns cannot be used to introduce the referent in discourse.

	Pronoun	Noun	Proper name
Discourse-new-perc. avail	30% (46) <	64% (100) >	6% (10)
Discourse-new-not perc. avail	2% (1) <	73% (33) >	24% (11) >
Discourse-given-maintenance	81% (218) >	16% (43) <	3% (8) <
Discourse-given-shift	53% (251)	35% (169)	12% (58) >

Table 4: Use of pronouns for pragmatic functions in the French input data.
For explanation: see Table 3.

It is clear from Table 4 that there are also strong form-function associations for pronouns as compared to nouns and proper names (χ^2 , $p < 0.001$). Pronouns are disassociated with both perceptually available and perceptually not available discourse-new-referents. There is, however, an association between pronouns and discourse-given maintenance. In contrast, nouns are disassociated with maintenance and associated with both forms of discourse-new-referents. The pragmatic function of referent shift is not particularly associated with either nouns or pronouns. However, pronouns are more often used for maintenance than shift, indicating that adults differentiate between degrees of givenness. Finally, proper names are most strongly associated with discourse-new referents that are not perceptually available and with referent shift.

4.2 The morphosyntax-pragmatics interface of articles in child French

The French children in this study had already started using articles before 2;0, the age at which the current investigation starts. The percentage of ungrammatical bare nouns is down to 2% or less at 2;9. Table 5 gives an overview of the percentage of realized articles in obligatory contexts at the six age points investigated.

	2;0	2;3	2;6	2;9	3;	3;3
Realized articles in obligatory cont.	60%	84%	91%	98%	99%	98%

Table 5: Percentage of realized articles and bare nouns in the French child data.

In Table 5, the category ‘realized articles’ includes indefinite, definite, demonstrative, possessive, partitive and fillers. In the following analysis, the focus is on how children use indefinite and definite/demonstrative articles for different pragmatic functions. The remaining forms are included in the category ‘other’, together with the ungrammatical bare nouns. Table 6 shows the children’s use of morphosyntactic forms for pragmatic functions per age point. Associations (>) and disassociations (<) of forms with functions per age point (asr-calculation) are again indicated in bold. It is important to keep in mind that the asr does not only take into account the distribution of forms within one particular function, but also the distribution of a particular form over all other functions.

	Form	2;0	2;3	2;6	2;9	3;0	3;3
Labelling	Indefinite	10% (8)	31% (35)>	36% (33)>	61% (54)>	58% (61)>	53% (53)>
	Def/dem.	9% (7)	30% (34)<	32% (30)<	17% (15)<	27% (28)<	30% (30)<
	Other	81% (64)	40% (45)	32% (30)	23% (20)	15% (16)	17% (17)
non-specific reference	Indefinite	33% (2)	50% (26)>	26% (10)	52% (33)>	71% (65)>	60% (32)>
	Def/dem.	(0)	8% (4)<	53% (20)	33% (21)<	13% (12)<	19% (10)<
	Other	67% (4)	42% (22)>	21% (8)	14% (9)	16% (15)	21% (11)
discourse-given	Indefinite	(0)	1% (1)<	2% (5)<	2% (2)<	4% (4)<	4% (4)<
	Def/dem.	19% (13)	64% (86)>	65% (147)>	83% (87)>	77% (79)>	64% (90)>
	Other	81% (55)	35% (47)	33% (74)	15% (16)	19% (20)	32% (45)>
discourse-new-MK	Indefinite	13% (2)	17% (18)	25% (31)>	21% (24)<	19% (24)<	31% (36)
	Def/dem.	19% (3)	54% (59)>	51% (62)	57% (64)	53% (67)>	45% (52)
	Other	69% (11)	29% (32)<	24% (29)	22% (25)	28% (36)>	24% (27)
discourse-new-NMK	Indefinite	(0)	13% (2)	36% (8)	20% (2)	58% (11)	63% (5)
	Def/dem.	50% (4)	44% (7)	32% (7)	70% (7)	21% (4)	25% (2)
	Other	50% (4)	44% (7)	32% (7)	10% (1)	21% (4)	13% (1)

Table 6: Use of articles for pragmatic functions in the French child data.

Raw figures are given in brackets. Bold=cell is a major contributor to the significant chi-square value for form-function associations. >=Adjusted standardized residual greater than 2.5, the morphosyntactic form is more strongly associated with this function than other forms and more strongly with this function than with other functions. < has the opposite interpretation.

At 2;3 for example, the use of indefinite articles, is strongly associated with labelling, whereas definite/demonstrative articles are disassociated with this function, even though the percentage of use for labelling itself hardly differs between the two forms. Since the children use the definite/demonstrative article more often used for specific-given referents than for labelling, there is a

disassociation between definitive/demonstrative articles and labelling and an association with indefinite articles.

The statistical analysis could not be carried out at 2;0, since the raw figures were too low. From 2;3 onwards, there are clear form-function associations in the children's language at every age point (χ^2 , $p < 0.001$). These associations have the same pattern as found in the input. That is, the children associate indefinite articles strongly with non-specific reference (see example 9) and disassociate definite/demonstrative articles with this function. The same pattern is found for labelling. In contrast, definite/demonstrative articles are associated with specific discourse-given referents. There is a disassociation between indefinite articles and specific discourse-given at all age-points.

(9) Non-specific reference with indefinite article (Philippe, 3;3)

- %sit: Philippe and the investigator are talking about the shape of an object.
*INV: Il est comment le mien alors?
'Then how is mine?'
*CHI: Il est ovale.
'It is oval.'
*INV: Il est ovale?
'It is oval?'
*INV: Ah bon.
'ok'
*CHI: Comme *un oeuf*.
'Like an egg.'

Overall, there are no clear associations or disassociations between either type of article and discourse-new-MK. However, if the form choice is analyzed separately for discourse-new-MK, it appears that definite/demonstrative articles are used significantly more frequently than indefinite articles at all age points, as was also found in the input (χ^2 , $p < 0.05$, see example 10).

(10) Discourse-new-MK with definite article (Léa 2;9)

- %sit: Léa is playing in the sink with her puppet.
*GRM: Comment dis tu?
'What did you say?'
*CHI: Je veux laver *les mains* de *la poupée*.
'I want to wash the puppet's hands.'

However, the children use indefinite articles significantly more frequently for discourse-new referents than for discourse-given referents at all age points (χ^2 , $p < 0.001$). This indicates that they do differentiate between new and given in article use. The French children do not significantly differentiate their use of morphosyntactic forms for MK and NMK in the age range 2;0-2;6 (χ^2 , $p = 0.21$, age points combined due to low cell frequencies). However, they have started to use more indefinite articles than definite/demonstrative articles for NMK at 2;9-3;3 (χ^2 , p

<0.01, age points combined due to low cell frequencies). The error of using a definite article for NMK is, however, still quite frequent at all ages. The error rate ranges between 21% and 50%. In example (11), Grégoire uses a definite plural article to introduce some horses he has seen the other day. The investigator tries to elicit more information, but Grégoire's answers are not informative enough, indicating that he does not take the investigators knowledge into account

(11) Discourse-new-NMK with definite article (Grégoire, 2;6)

- *CHI: Non, non, moi j'étais, moi j'ai été regarder *les chevaux*.
 'No, no, I went, I went to see the horsies.'
 [...]
- *INV: Y'a longtemps que tu as été voir les chevaux?
 'Is it a long time ago that you went to see the horsies?'
- *CHI: Et pis j'en a pas des oranges des chevaux.
 'And then there weren't any orange horses.'
 [...]
- *INV: Quand quand tu es allé voir les chevaux?
 'When, when did you go to see the horsies?'
- *CHI: J'ai été, y'a longtemps.
 'I went, is long time ago.'

The French children use articles for referents that are new or given in discourse in an adult-like way from an early age. There are no differences between the children and the adults in the use of articles for discourse-new-MK ($\chi^2, p < 0.96$). Moreover, already at 2;3 the children hardly use indefinites for discourse-given, a pattern also found in the input language. Moreover, by showing an association between indefinites and non-specific reference and a disassociation of this function with definite articles, the French children demonstrate that they also make a difference between non-specific and specific reference. The difference between MK and NMK in article use is emerging in these children. Before the end of the period under investigation, i.e. at 3;0, they have reached adult levels of the associations of forms with the aspects of specificity and givenness in discourse, but with the presence or absence of mutual knowledge.

4.3 The morphosyntax-pragmatics interface of pronouns in child French

In this section, the pragmatic aspects of givenness and familiarity to the listener (here defined as perceptual availability) will be discussed in relation to pronoun use. Table 7 shows that from 2;3 onwards, the children make clear form-function combinations; pronouns, nouns and proper names are used differently for the four pragmatic functions at all age points ($\chi^2, p < 0.001$). Nouns are associated with discourse-new and disassociated with discourse-given-maintenance from 2;3 onwards. In contrast, pronouns are strongly associated with discourse-given maintenance at all age points (example 12).

	Form	2;0	2;3	2;6	2;9	3;0	3;3
Discourse- new- perc. available	Pronoun	35% (8)	27% (30)	37% (50)	34% (48) <	33% (50) <	32% (46) <
	Noun	65% (15)	73% (82) >	63% (84) >	65% (91) >	66% (102) >	65% (93) >
	ProperN	(0)	(0) <	(0) <	1% (2) <	1% (2) <	3% (4)
Discourse- new- <i>not</i> perc. avail.	Pronoun	8% (1)	2% (1) <	8% (6) <	2% (1) <	7% (4) <	11% (5) <
	Noun	75% (9)	88% (43) >	78% (60) >	76% (32) >	73% (44) >	67% (30) >
	ProperN	17% (2)	10% (5)	14% (11)	21% (9)	20% (12) >	22% (10) >
Discourse- given- maintenance	Pronoun	30% (15)	54% (57) >	63% (144) >	86% (171) >	85% (175) >	85% (210) >
	Noun	60% (30)	41% (43) <	34% (78) <	8% (15) <	11% (23) <	13% (31) <
	ProperN	10% (5)	5% (5)	4% (8) <	7% (13) <	3% (7) <	2% (5) <
Discourse- given- shift	Pronoun	13% (13)	20% (30) <	40% (124)	51% (149)	61% (182)	62% (244)
	Noun	38% (38)	60% (91)	48% (148)	31% (90)	27% (80) <	28% (109) <
	ProperN	49% (49)	20% (31) >	13% (39) >	18% (53) >	12% (35) >	11% (42) >

Table 7: Use of articles for pragmatic functions in the French child data.

Raw figures are given in brackets. Bold=cell is a major contributor to the significant chi-square value for form-function associations. >=Adjusted standardized residual greater than 2.5, the morphosyntactic form is more strongly associated with this function than other forms and more strongly with this function than with other functions. < has the opposite interpretation.

(12) Discourse-given-maintenance with pronoun (Léa, 2;9)

MOT: Et Luc_i, où est il en ce moment?

‘And Luc, where is he now?’

CHI: Euh *i*_i fait dodo.

‘Uh, he is doing a nap.’

The children’s sensitivity to new and given in pronoun use is, however, not completely adult-like at 2;3 and therefore, seems to be developing up to 2;9. That is, only from 2;9 onwards the children disassociate pronouns with discourse-new referents that are perceptually available as do the adults. Furthermore, the use of pronouns in maintenance strongly increases to around 80% at 2;9, a figure also

found in the input. Finally, at this age, the children also start to use more pronouns and fewer nouns for discourse-given-shift.

The choice of different forms for referent maintenance versus referent shift indicates whether children distinguish between different degrees of givenness. It appears that from 2;3 onwards, the children use pronouns, nouns and proper names significantly differently for maintenance and shift (χ^2 , $p < 0.001$). They prefer to use pronouns for maintenance, whereas nouns and proper names are more often used for shift (example 13).

(13) Discourse-given-shift with definite article (Philippe 2;3)

- MOT: Si la vitre_i est cassée, y a pas que ça.
'If the window is broken, only that is left.'
- MOT: Les roues aussi, mon pauvre chat.
'The wheels too, my poor darling.'
- CHI: Réparer *la vitre*_i.
'Repair the window.'

The erroneous use of pronouns for discourse-new referents that are not perceptually available varies between 2% and 11%. This error is thus less frequent than the use of definite articles for discourse-new-NMK (see section 4.2). Moreover, the children significantly differentiate their use of pronouns, nouns and proper names according to whether the referent is perceptually available or not at both 2;0-2;6 and 2;9-3;3 (χ^2 , $p < 0.01$, age points combined due to low cell frequencies). Pronouns are more often used for referents that are physically present. If the referent is absent, nouns or proper names are preferred. This pattern is highly similar to what was found in the input.

In sum, the children associate pronouns with given referents from 2;3 onwards. Moreover, they prefer to use pronouns over nouns if a discourse-given referent is highly in focus (maintenance), indicating that they also distinguish between different degrees of givenness. Their sensitivity to givenness seems however to be developing to the adult level until 2;9. Only at this age do the children disassociate pronouns with discourse-new referents that are also perceptually available. From 2;3 onwards, the children appear to take account of the listener's perspective, since they prefer to use nouns or proper names if the referent is not perceptually available. Whether this really indicates sensitivity to the perspective of the listener will be discussed in the next section.

5. Discussion and conclusion

This study has investigated reference to persons and objects with articles and pronouns longitudinally in two- and three year old children acquiring French. In adult French the choice for an indefinite or definite article and noun or pronoun is, amongst other things, influenced by three pragmatic aspects: (1) whether the referent is non-specific or specific, (2) whether the referent is new or given in discourse and (3) whether the referent is familiar to the hearer, based on mutual knowledge or

perceptual availability. Studies on children's socio-cognitive knowledge indicate that some of the cognitive underpinnings of these three pragmatic aspects are already present when children start to produce articles and pronouns, that is around two years-of-age. The goal of this study was to examine if and how young children take account of these pragmatic dimensions in using indefinite and definite/demonstrative articles and pronouns when they are in the process of acquiring these forms, that is between two and three years-of-age. The role of the input was also taken into account. The children in this study have reached an adult level of producing articles in obligatory contexts by 2;9.

With respect to article use, the results show that the children apply both the non-specific/specific distinction and the distinction between new/given in discourse in an adult-like way from 2;3 onwards. That is, indefinite articles are associated with non-specific reference and labelling. Definite/demonstrative articles are associated with specific referents, both discourse-new and discourse-given. Moreover, within the category of specific referents, the children do use indefinite articles for discourse-new referents, but hardly make the error of using indefinites for discourse-given referents. This indicates that from at least 2;3 onwards, French children also distinguish between new and given in discourse for article use. For the pragmatic aspects of givenness and specificity, the French children thus already make adult-like form-function associations by 2;3. With respect to the perspective of the listener however, the children need more time to use articles correctly. They often use definite articles for not mutually known referents, also at 3;3.

The French children are also sensitive to the distinction between new and given in discourse in using pronouns compared to nouns (and proper names). From 2;3 onwards, nominal forms are associated with new referents, whereas pronouns are associated with given referents. Furthermore, the children use pronouns more frequently when the referent is syntactically close to its antecedent (maintenance) than when the syntactic distance is larger (shift). There is however also a developmental pattern in the children's sensitivity to new and given in pronoun use: only from 2;9 onwards the percentage of pronouns for maintenance is at an adult level. Moreover, the children start to disassociate pronouns with discourse-new-referents that are perceptually available at this age. The children seem to take the perspective of the listener into account in using pronouns already at 2;0. They hardly use pronouns to introduce referents that are not physically present and not perceptually available. Rather, they correctly prefer to use nouns or proper names to refer to these entities. Whether this pattern of pronoun use really indicates sensitivity to the listener will be discussed below.

The study also considered if the children apply pragmatic aspects in similar ways to articles and pronouns. The results are somewhat difficult to interpret. Firstly, the children distinguish between discourse-new and discourse-given in article use in an adult-like way from 2;3 onwards. In contrast, it takes the children to 2;9 to reach the adult pattern in using pronouns versus nouns and proper names with respect to givenness in discourse. The consistency of form-function combinations in the input might play a role here. Indefinite determiners are hardly used for discourse-given referents in the input (3%). This might give children a strong cue on the restricted use of this form. The pattern of use of nouns and pronouns for pragmatic functions is, however, much less clear. Both forms are used to a

considerable extent for both new and given referents in the input. For example the adults use 30% of pronouns for discourse-new referents that are perceptually available. They use nouns in 35% of the cases in discourse-given-shift. These less systematic form-function associations in the input might give children a less strong cue on how to use nouns and pronouns for given referents. They seem to need more time to acquire these form-function associations.

The results are also contradictory with respect to the children's sensitivity to the listener's perspective. Up to 2;9 the children do not differentiate their use of articles with respect to MK or NMK and at 3;3, the children still make the error of using an definite article to refer to a referent that is not mutually known between speaker and listener. In contrast, the children seem to be more sensitive to the listener's needs in using pronouns. From 2;3 onwards, the children disassociate pronouns with discourse-new referents that are not perceptually available to the listener and use this form erroneously in only about 7% of the cases. However, this apparent sensitivity to the listener's needs might be explained by the deictic properties of pronouns. That is, the children may be using pronouns to introduce referents to discourse only if the referent is perceptually available to the children themselves. This suggestion needs to be investigated further, preferably in an experimental situation. Results by Matthews, Lieven, Theakston and Tomasello (2006) suggest this is the case. They investigated two-year-old children's use of nouns and pronouns in an experimental situation where the child could always see the referent and the listener either could or could not see it. In these situations two-year-old children did not differ their use of morphosyntactic forms with respect to the perceptual availability of the referent to the listener, possibly because the referent was visible to the children.

To conclude, the current study has shown that children between two and three acquiring French already apply pragmatic aspects to the use of articles and pronouns. Different aspects of the morphosyntax-pragmatics interface in reference are, however, acquired at different rates. The distinctions between non-specific/specific and new/given in discourse are both acquired and used in an adult-like way before the end of the period under investigation here (3;3). The children have not yet learned to correctly take the listener's perspective into account by this age. Furthermore, the morphosyntax-pragmatics interface is acquired on a form-by-form basis. Articles are used according to the specificity of the referent and according to whether the referent is new or given in discourse from 2;3 onwards. For pronouns, deixis seems to be important. Only by 2;9, the French children have developed an adult-like use of pronouns and nouns with respect to the pragmatic aspect of givenness.

References

- Ariel, M. (1996) 'Referring expressions and the +/- coreference distinction', in: T. Fretheim and J.K. Gundel (eds) *Reference and Referent Accessibility*. Amsterdam/Philadelphia: John Benjamins, 13-37.
- Bassano, D., I. Maillochon and S. Mottet (2005) 'Noun grammaticization in French: prosodic and lexical factors on determiner use in children's speech'. Paper presented at *ELA (Emergence of Linguistic Abilities)*, Lyon, 8-10 December.
- Brown, G. and G. Yule (1983) *Discourse Analysis*. Cambridge: Cambridge University Press.

- Chiercha, G., M.T. Guasti and A. Gualmini (2001) 'Nouns and articles in child grammar and the syntax/semantics map', ms. University of Milan / University of Siena / University of Maryland, College Park.
- Clark, H.H. and C.R. Marshall (1981) 'Definite reference and mutual knowledge', in: A.K. Joshi, I.A. Sag and B.L. Webber (eds) *Elements in Discourse Understanding*. Cambridge: Cambridge University Press, 10-63.
- Enç, M. (1991) 'The semantics of specificity'. *Linguistic Inquiry* 22, 1-25.
- Guerriero, A.M.S., Y. Oshima-Takane and Y. Kuriyama (2006) 'The development of referential choice in English and Japanese: a discourse-pragmatic perspective'. *Journal of Child Language* 33, 823-857.
- Gundel, J.K. (1996) 'Relevance theory meets the givenness hierarchy', in: T. Fretheim and J.K. Gundel (eds) *Reference and Referent Accessibility*. Amsterdam/Philadelphia: John Benjamins, 141-153.
- Hamann, C., L. Rizzi and U.H. Frauenfelder (1996) 'On the acquisition of subject and object clitics in French', in: H. Clahsen (ed), *Generative Perspectives on Language Acquisition: empirical findings, theoretical considerations and cross-linguistic comparisons*. Amsterdam/Philadelphia: John Benjamins, 309-334.
- Hickmann, M. (2003) *Children's Discourse: Person, Space and Time across Languages*. Cambridge: Cambridge University Press.
- Kampen, J. van (2004) 'Learnability order of the French pronominal system', in: R. Bok-Bennema, B. Hollebrandse, B. Kampers-Mahne and P. Sleeman (eds) *Romance Languages and Linguistic Theory: Selected papers from 'Going Romance' 2002*. Amsterdam/Philadelphia: John Benjamins, 163-181.
- Kampen, J. van and M. Pinto (this volume) 'Germanic and Romance discourse devices in acquisition'.
- Lyons, C. (1999) *Definiteness*. Cambridge: Cambridge University Press.
- MacWhinney, B. (2000) *The CHILDES Project: Tools for analyzing talk*. Hillsdale, N.J.: Lawrence Erlbaum.
- Matthews, D., E. Lieven, A. Theakston and M. Tomasello (2006) 'The effect of perceptual availability and prior discourse on young children's use of referring expressions'. *Applied Psycholinguistics* 27, 403-422.
- O'Neill, D.K. (2005) 'Talking about new information: the given/new distinction and children's developing theory of mind', in: J. Wilde Astington and J.A. Baird (eds) *Why Language Matters for Theory of Mind*. Oxford: Oxford University Press.
- Pannemann, M. (2006) 'More variability in French L1: consequences for theories of DP-acquisition', in: M. Vliegen (ed), *Linguistik International 16. Variation in Sprachtheorie und Spracherwerb* Amsterdam: Lang, 223-232.
- Power, R.J.D. and M.F. Dal Martello (1986) 'The use of the definite and indefinite articles by Italian preschool children'. *Journal of Child Language* 13, 145-154.
- Roder, B.J., E.W. Bushnell and A.M. Sasseville (2000) 'Infants' preferences for familiarity and novelty during the course of visual processing'. *Infancy* 1, 491-507.
- Ruffman, T. and J. Perner (2005) 'Do infants really understand false belief?' *Trends-in-Cognitive-Sciences* 9, 462-463.
- Schaeffer, J. and L. Matthewson (2005) 'Grammar and pragmatics in the acquisition of article systems'. *Natural Language and Linguistic Theory* 23, 53-101.
- Serratrice, L. (2005) 'The role of discourse pragmatics in the acquisition of subjects in Italian'. *Applied Psycholinguistics* 3, 437-462.
- Serratrice, L. (this volume) 'Null and overt subjects at the syntax-discourse interface: Evidence from monolingual and bilingual acquisition'.

- Tomasello, M. and K. Haberl (2003) 'Understanding attention: 12- and 18-month-olds know what is new for other persons'. *Developmental Psychology* 39, 906-912.
- Velde, M. van der, C. Jakubowicz and C. Rigaut (2002) 'The acquisition of determiners and pronominal clitics by three French-speaking children', in: I. Lasser (ed) *The Process of Language Acquisition*. Frankfurt/Berlin: Peter Lang Verlag, 115-132.
- Veneziano, E. and H. Sinclair (2000) 'The changing status of 'filler syllables' on the way to grammatical morphemes'. *Journal of Child Language* 27, 461-500.
- Wilson, D. and D. Sperber (2004) 'Relevance', in: G. Ward and L. Horn (eds) *Handbook of Pragmatics*. Oxford: Blackwell, 607-632.

The Relationship Between Word Prosodic Structure and Sentence Prosody

(Non)evidence from Brazilian Portuguese

Raquel S. Santos and Paula Fikkert
Universidade de São Paulo and Radboud University Nijmegen

1. Introduction

In recent years a number of studies have addressed the acquisition of prosodic structure, in particular word stress (Fikkert 1994, Demuth 1996, Demuth and Fee 1995, Gerken 1994, Archibald 1995, Santos 2001, 2003, Grimm 2004, among others). The study of the acquisition of word stress has been subject to large debates. A number of researchers reported an early trochaic bias in the acquisition of particularly Germanic languages, both in perception (Jusczyk, Cutler and Redanz 1993) and in production (Allen and Hawkins 1978, 1980, Echols and Newport 1992, Gerken 1994, Fikkert 1994, among other). More recently many researchers have convincingly argued that this bias is not innate, but reflects language-specific knowledge (Vihman, DePaolis and Davis 1998, Santos 2001, 2003, 2006 among others). In our earlier work, we have compared word prosodic structure in the acquisition of Dutch and Brazilian Portuguese, where Dutch children have trochaic word patterns at a very early stage, while Brazilian Portuguese children seem to favor iambic word patterns (Fikkert and Santos 2005, Santos 2006).

One issue that has been raised in the literature on the acquisition of stress, but has not been discussed in much detail, is whether the acquisition of word stress is a process that is entirely a bottom up process, in the sense that the child's word template is being extended in the course of development, as proposed by Fikkert (1994), or whether top-down processes also influence the prosodic shape of early words, such as argued by Santos (2001, 2003) and Grimm (2006). In the latter view children start with larger prosodic units (utterances), which have their own intonation and intonational boundaries, which correlate with prosodic prominence. Santos and Grimm have hypothesized that the prosodic structure of words in isolation in fact reflects the prosodic structure of utterances, rather than word stress. If this is the case, the prosodic structure of words in utterances larger than a single word may be largely dependent on the prosodic structure of the utterance in which they occur.

In this paper we pursue the issue of whether the rhythmic structure of the utterance in which a word occurs could explain variation in the prosodic structure of the first words. In particular, we will address the question whether the iambic bias that has been reported in the acquisition of Brazilian Portuguese (Santos 2001, 2003, 2006, Bonilha 2005) could be due to prosodic context-effects in utterances.

The paper is organized as follows. Section 2 gives a global description of the prosodic structure of words in Brazilian Portuguese. Section 3 presents an overview of the studies on acquisition of primary word stress in Brazilian Portuguese. In

section 4 we formulate hypotheses that can be made if we assume top-down influences on word prosodic structure. In particular, we discuss predictions with respect to syllable deletion and insertion as a function of different rhythmic conditions in the utterance. Section 5 describes the methodology used in this study. Section 6 presents the results. Finally, section 7 presents some concluding remarks.

2. Word prosodic patterns in Brazilian Portuguese

In Brazilian Portuguese, word stress falls on one of the last three syllables of the word. The distribution of word prosodic patterns is different for nouns and verbs. According to Cintra (1997), in nouns, 18% has final stress (cf. 1a), 63% penultimate stress (cf. 1b), and 7% of the nouns have stress on the third syllable from the end (cf. 1c). Nouns, such as those in (2), which are often used in child-directed speech and hence are common in children's early vocabularies, almost always have final stress, as can be seen in the examples in (2a–c). On the other hand, nouns with a diminutive suffix, which also are very common in child language, change word stress to the penultimate syllable, as can be seen in (1d–f) and (2d–f).^{1,2}

- | | | | | | | | | |
|--------|------|---------|----------|---------|-----------|-------------|----------------|---------------|
| (1) a. | caFÉ | [ka fÉ] | “coffee” | d. | caféZInho | [kafÉ zi≠U] | “small coffee” | |
| | b. | CAsa | [kaza] | “house” | e. | caSInha | [ka zi≠a] | “small house” |
| | c. | Ônibus | [onibus] | “bus” | f. | onibuZInho | [onibu zi≠U] | “small bus” |
| | | | | | | | | |
| (2) a. | xiXI | [Si Si] | “pee” | d. | xixiZInho | [SiSi zi≠U] | “small pee” | |
| | b. | coCO | [ko ko] | “poo” | e. | cocoZInho | [koko zi≠U] | “small poo” |
| | c. | neNÊ | [ne ne] | “baby” | f. | neneZInho | [nene zi≠U] | “small baby” |

Verbs do not have a predominant pattern. Infinitives and the first and third person singular forms of the simple past tense have final stress (3a–c), while imperatives, gerunds, and the first and third person singular forms of the present tense have penultimate stress (4a–c):

- | | | | | | | |
|--------|-------|-----------|---------------|------|---------------------|--------------|
| (3) a. | faLAR | “to talk” | (4) a. | fale | “talk (imperative)” | |
| | b. | faLEI | “(I) talked” | b. | falo | “(I) talk” |
| | c. | faLOU | “(he) talked” | c. | fala | “(he) talks” |

It is worth mentioning that in Brazilian Portuguese the main acoustic correlate of primary stress is duration (e.g., Major 1985, Moraes 1987, Massini-Cagliari 1992).

¹ The discussion of whether in stress in words with a diminutive suffix is lexical or post-lexical is beyond the goal of this paper. We refer to Lee (1995) for a relevant discussion.

² Stressed syllables are indicated with capital letters.

3. The acquisition of Brazilian Word Stress

Claims on the acquisition of word prosodic patterns and word stress are often based on truncation patterns. Based on experimental evidence from truncation data, Rapp (1994) claims that Brazilian children display a trochaic pattern in the early words. This contrasts to a large number of studies on word stress acquisition in Brazilian Portuguese, which have reported an iambic bias at the early stages of the acquisition process (Santos 2001, 2003, 2006, Bonilha 2005, Baia to appear).

Rapp's claims were based on experimental data. She conducted an experiment that induced the deletion of weak syllables and found that children between 1;6 to 2;0 years old produced 51% of the words as trochees and only 38% as iambs, which led her to argue in favor of a trochaic bias. However, the experiment had more targets where the deletion of unstressed syllables resulted in trochees, than targets that would result in iambs. A reanalysis of her data presents a completely different picture: target iambs, such as *caFÉ* "coffee", are produced correctly in 87,2% of the cases, while for trochees, such as in *CAsa* "house" the percentage correctly produced forms was 82%. Target iambs were produced as monosyllables in 10,6%, while this was 10,2% for trochees. Finally, stress errors were rare: for target iambs the percentage of stress errors was 2,1%, while for target trochees it was 2,6%. From this reanalysis we can draw the conclusion that children do not treat iambs and trochees differently. In other words, these data do not show a bias for either a trochee or an iamb.

Based on analyses of spontaneous longitudinal data from Brazilian children Santos (2001, 2003, 2006), Fikkert and Santos (2005), Bonilha (2005), and Baia (to appear) argued that at the early stages these children showed a predominance of iambic word forms. They give different explanations for this finding. Santos (2001, 2003, following Scarpa 1999) argued that the predominance of iambic patterns reflects sentence prosody, rather than word prosody. In particular, based on the analysis of whole sentences, she claimed that children's iambic word patterns are the result of intonational prominence. This prominence falls at the right edge of an utterance boundary, and hence looks iambic.

Santos (2006), Fikkert and Santos (2005), Bonilha (2005), and Baia (to appear) looked at the prosodic structure of words out of context. Santos (2006) analyzed isolated words in the same corpus as presented here, analyzed them according to the same method as used in Fikkert (1994) and compared the results with those reported for Dutch in Fikkert (1994). Fikkert (1994) showed that Dutch children produce trochees correctly from a very early stage onwards; while iambs are truncated to monosyllables (WS >> S) and trisyllabic words with medial stress are truncated to trochees (WSW >> SW). In other words, the initial unstressed syllable often is not produced. On the other hand, monosyllables do sometimes have an inserted syllable to the right edge (S >> SW), giving rise to a trochaic pattern. The results of Santos' study show that Brazilian children correctly produce iambs until the age of 1;7, but they often truncate trochees to monosyllables (SW >> W). If they insert syllables these appear to the left edge of monosyllables and disyllabic trochees (S(W) >> WS(W)), while trisyllabic words with medial stress are truncated to iambs (WSW

>> WS). This pattern seems to be the exact opposite of that of the Dutch children. Whereas Dutch children seem to aim at producing trochees, Brazilian children aim at producing iambs. In short, there is no evidence for a universal trochaic bias.

4. Hypotheses

In this paper we investigate the role of sentence prosodic structure on the realization of word prosodic structure. From earlier research we know that stressed syllables are more prominent, and that children are more likely to pay attention to prominent syllables (e.g., Waterson 1971, Jusczyk, Cutler and Redanz 1993). Moreover, in general syllables at the end of intonational phrases are lengthened (Hayes 1995), whereas syllables at the beginning of phonological phrases are more carefully produced (Cho and Keating 2001). Furthermore, languages try to optimize the rhythmic structure of utterances (Nespor and Vogel 1986). Based on these insights, our hypothesis is that the position of the word in the utterance may influence the prosodic shape of children's first words, which is an alternative account for syllable deletion and insertion in early child data.

In this paper we analyze children's realization of the prosodic structure of target words in three contexts: First, words in one-word utterances, such as exemplified in (5); second, words at the edge of an intonational boundary, like in (6); and third, words that are not at an intonational boundary, as in (7):³

- | | | |
|-----|---|------------------|
| (5) | [ca VAl _o] _{IP} | “horse” |
| (6) | [o ca VAl _o] _{IP} | “the horse” |
| (7) | [o ca VAl _o saiu] _{IP} | “the horse left” |

The first context is a neutral condition, in the sense that there can be no clashes or lapses in this context. However, in a one-word utterance, the word-initial syllable is also at the beginning of an intonational phrase, and hence, may be produced more correctly, while in this context the final syllable will be lengthened due to phrase-final lengthening.

The second context should allow us to see whether children take adjacent words, and in particular, adjacent syllables (weak or strong) into account. For instance, if a weak syllable of a word is preceded by a weak syllable in the preceding word, and hence is in a lapse context, this may more frequently lead to truncation of that syllable than in a neutral context.

Finally, the third context should allow us to investigate the influence of edges of intonational phrases. It allows us to test whether weak syllables are more often retained in phrase-initial or phrase-final position than elsewhere. To summarize, we will test the following contexts.

³ The relevant context is in bold.

First, contexts that favor the retention of unstressed syllables are given under (8a). These include contexts where a weak syllable at the edge of a word is preceded or followed by a strong syllable of another word. We do not expect children to add or delete syllables, as the resulting structure is rhythmically optimal. Examples are given in (9a). A similar prediction is made for the contexts given in (8b), where a strong syllable at the edge of a word is preceded or followed by a weak syllable of another word. Examples are given in (9b). Again, we do not expect children to add or delete syllables, as the resulting structure is already rhythmically optimal. On the other hand we predict weak syllables to be prone to deletion in the context where a weak syllable at the edge of a word is preceded or followed by a weak syllable of another word, as in (8c), and the examples in (9c). In these contexts there is a rhythmic lapse (two adjacent weak syllables).⁴ Finally, the insertion of an additional syllable is most likely to occur in the contexts given in (8d), in which the initial stressed syllable of a word is preceded by the stressed syllable of the preceding word, or the final stressed syllables of a word is followed by the stressed syllable of the next word. In these contexts there is a stress clash, and hence, we predict that these contexts are favorable for syllable insertion to undo the clash. Examples are given in (9d).

- (8) a. S – WS(W), where a WS(W)⁵ word is preceded by a strong syllable
 (W)SW – S, where a (W)SW pattern is followed by a strong syllable
 b. W – S(W), where a S(W) word is preceded by a weak syllable
 (W)S – W, where a (W)S word is followed by a weak syllable
 c. W – WS(W), where a WS(W) word is preceded by a weak syllable
 (W)SW – W, where a (W)SW word is followed by a weak syllable
 d. S – S(W), where a S(W) word is preceded by a strong syllable
 (W)S – S, where a (W)S word is followed by a strong syllable
- (9) a. TÁ feCHAdo “it’s closed”
 SAco FEIo “ugly bag”
 b. a CAsa “the house”
 caFÉ peQUEno “small coffee”
 c. a meNIa “the girl”
 CAsa verMEIha “red house”
 d. TÁ FEIo “it’s ugly”
 caFÉ FORte “strong coffee”

⁴ In the context (W)SW-W there are two possible deletions: the deletion of the weak syllable of the target word (WSØ-W) and the deletion of the weak syllable of the adjacent word (WSW-Ø). However, sometimes the adjacent word is a monosyllabic word. Therefore, if children deleted that syllable, there would be no trace to postulate its existence. Therefore, we only consider cases where the deletion involved a syllable of the target word.

⁵ Parentheses indicate optionality. In the contexts described here, this indicates that the target word can either be bisyllabic or trisyllabic. A similar situation holds for the context.

5. Methodology

The corpus that forms the basis for our investigation is part of the *Projeto de Aquisição da Linguagem* of the Universidade Estadual de Campinas (Lemos 1995) and the *Projeto de Aquisição do Ritmo* of the Universidade de São Paulo (Santos 2005). Our corpus consists of production data from two Brazilian children, from 1;4 to 2;0 years of age. Both children come from the state of São Paulo. Spontaneous interactions of the children with their parents were audio-recorded on a monthly basis in half-hour sessions. The data were phonetically transcribed by the first author and later double-checked by other trained native speakers. Only data for which full agreement between the transcribers was reached are taken into consideration here. We selected words that appeared minimally eight times, so that the same words may be analyzed in different contexts.

The following prosodic word patterns were attested:

- | | | | |
|---------|-------------------------|--------|----------|
| (10) a. | SW (trochee) | GAtó | “cat” |
| b. | WS (iamb) | caFÉ | “coffee” |
| c. | WSW | meNIna | “girl” |
| d. | S (strong monosyllable) | pé | “foot” |

The word pattern in (10a) and (10b) are particularly important to analyze, because these could be target to truncation (SW, WS >> S), or stress errors (SW >> WS, or vice-versa). The insertion of syllables is also possible (SW >> WSW; WS >> WSW). A trisyllabic word with medial stress (as in (10c)) could be truncated either to a trochee (WSW >> SW) or to an iamb (WSW >> WS). Finally, a monosyllabic word, such as in (10d) could be changed to a trochee or an iamb, depending on the position of the inserted syllable (S >> SW; S >> WS). In (11) possible realizations of different target prosodic patterns are illustrated.

- | | | | |
|---------|--------|----|---|
| (11) a. | CArro | >> | [^l ka.u] ~ [ka] ~ [ka ^l ka] “car” |
| b. | miGUEL | >> | [mi ^l ge] ~ [ge] ~ [^l mige] <i>proper name</i> |
| c. | meNIno | >> | [mi ^l ni] ~ [^l ninu] “boy” |
| d. | PÉ | >> | [p ^l] ~ [a ^l p ^l] “foot” |

In total, there were 1332 tokens taken into account. Two types of words were left out of the analysis. First, the reduplicative ‘familiar’ words, such as those in (2a–c), which invariably have final stress, and second, words with a diminutive suffix, which invariably have penultimate stress.

The words were classified according to the context in which they appear (see also (8)): whether the word forms a one-word utterance, whether the word occurs at an intonational boundary, and whether adjacent syllables were strong or weak. In

Table 1 an example for each of these contexts is provided to illustrate the way in which a word can be classified:⁶

	Target word /ka.ˈva.lu/ “horse”	No. of adjacent syllables	Left syllable	Right syllable	IP boundary
a.	[ta. <u>ka</u> .ˈ <u>ka</u> .la.ˈtej]	2	W	S	no
b.	[ka.ˈla.le]	0	∅	∅	yes (RL)
c.	[ka ka]	0	∅	∅	yes (RL)

Table 1: Classification of the target words

6. Results

Below we present the results for the individual contexts as sketched in (8) above. We discuss the contexts, which are most susceptible to change, because the target word either forms a lapse (section 6.1) or a clash (section 6.2) with the adjacent syllables. In 6.3 the situation in which target words and surrounding syllables form optimal rhythmical patterns (8ab). In this context target words should be least prone to change. Finally, in 6.4 we present the results for targets that are at initial or final intonational boundaries.

6.1 Adjacent weak syllables (W – W) – Contexts with a lapse

In Table 2 below we have given the raw numbers of instances where a weak syllable is deleted or maintained. This is graphically represented in Graph 1. However, one should bear in mind that the percentages are indicative of the development only, as the numbers at early stages of development are usually too low to justify the use percentages.⁷ The table and graph show the results for the context in which a lapse occurs, at respectively the left (W – WS(W)) and the right ((W)SW – W) edge of the target word. Deletion at the left edge results in a trochee or a monosyllable (W – ∅ S(W)). As can be seen, deletion at the left edge was more common around the age of 1;6 and 1;7, but in the majority of the cases the children produced the weak syllable. On the other hand, deletion of weak syllables at the right boundary would result in iambs or monosyllables ((W)S∅ – W). As can be seen, deletion was rare, and only appeared after 1;9. An example of a target word in a lapse context at the left boundary is presented in (12).

(12) [ta ka | kala | tej] W – WSW caVAlo “horse” R. 1;6

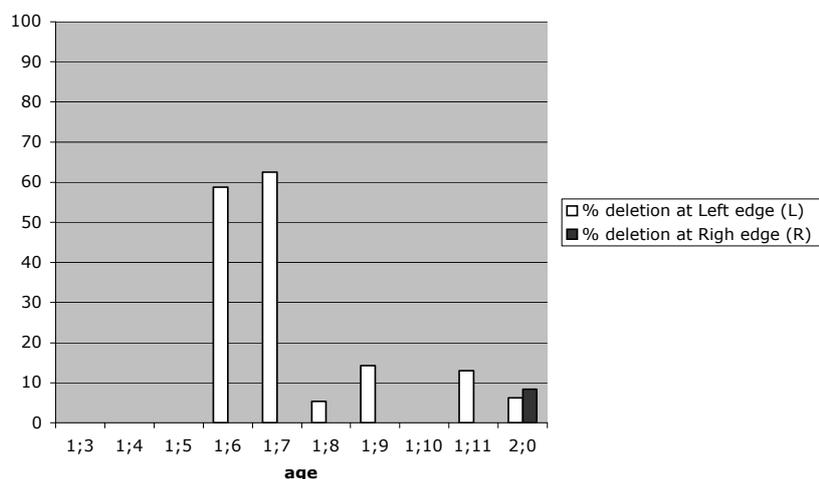
⁶ When there are adjacent syllables or words, the target word is underlined, as in (a).

⁷ A “*” marks cases, in which the number is too low to give percentages.

Age	1;3	1;4	1;5	1;6	1;7	1;8	1;9	1;10	1;11	2;0
Syllable deletion L	0	0	2*	10	5	1	1*	0	3	1
Maintenance L	0	5*	0	7	3	18	6*	0	20	15
% del L				58,8	62,5	5,3			13	6,3
Syllable deletion R	0	0	0	0	0	0	1*	1*	1*	1
Maintenance R	0	0	0	1*	0	2*	4*	1*	6*	11
% del R										8,3

Table 2: Raw numbers of instances of weak syllable deletion and maintenance of weak syllables at the L(ef) or R(ight) word boundary, given per month

Adjacent weak syllable at left/right word boundary



Graph 1: Adjacent weak syllables at the left and right boundary of a WS(W) target

6.2 Adjacent strong syllables (S – S) – Contexts with a clash

Table 3 and Graph 2 show the resulting patterns in the context in which the strong syllable of the target is adjacent to another strong syllable at respectively the left (S – S(W)) and right edge of the target word ((W)S – S). An example of such a context is given in (13). In this context, the insertion of a syllable would create an iamb (S – wS), whereas the insertion at the right boundary would create a trochee (Sw – S).

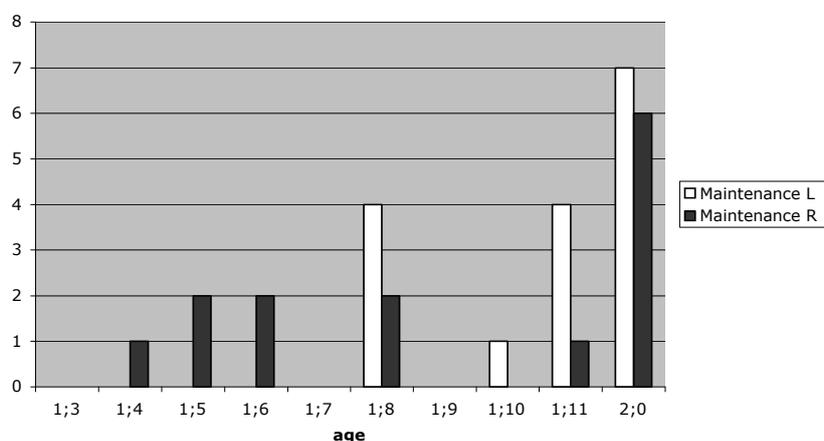
(13) [ka | i. | **bi**Su] S – SW caIR BIcho “fell animal” L.2;0

(14) [avopoesaki] S – SW eu **VOU** POR “I’m gonna R.2;0
Esse aQUI put this one here”

Age	1;3	1;4	1;5	1;6	1;7	1;8	1;9	1;10	1;11	2;0
Syllable insertion L	0	0	0	0	0	0	0	0	0	0
Maintenance L	0	0	0	0	0	4	0	1	4	7
Syllable insertion R	0	0	0	0	0	0	0	0	0	0
Maintenance R	0	1	2	2	0	2	0	0	1	6

Table 3: Raw numbers of instances of syllable insertion and maintenance of clash situation at the L(ef) and R(ight) word boundary, given per month

Adjacent strong syllable at left/right word boundary



Graph 2: Adjacent strong syllables at the left/right boundary of a S(W) / (W)S target

6.3 Adjacent syllables in rhythmically optimal contexts

Target word can also appear in various contexts that are rhythmically optimal because strong and weak syllables alternate. Table 4 and graphs 3 show the result in the context where a strong syllable precedes a target word that starts with a weak syllable (S – WS(W)), and the context where a strong syllable follows a target word that ends with a weak syllable (SW – S). If children would delete or insert a syllable, they would create a lapse (S – wWS(W)) or a clash (S – \emptyset S(W)). As we can see below, this never happens. Children did not delete or insert syllables, as shown in (15), except in three instances, one of which is shown in (16).

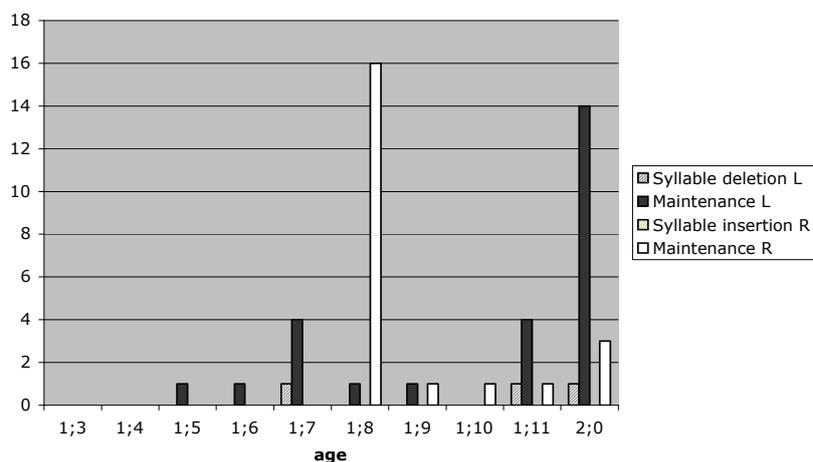
(15) [pEgaeli] PEga Ele “Get him!” R.2;0

(16) [esitadaj] esse esTÁ doDÓi “This is hurting” L.1;11

Age	1;3	1;4	1;5	1;6	1;7	1;8	1;9	1;10	1;11	2;0
Syllable deletion L	0	0	0	0	1*	0	0	0	1*	1
Maintenance L	0	0	1*	1*	4*	1*	1*	0	4*	14
Syllable deletion R	0	0	0	0	0	0	0	0	0	0
Maintenance R	0	0	0	0	0	16	1*	1*	1*	3*

Table 4: Raw numbers of instances of syllable deletion and maintenance of a rhythmically optimal situation at the L(ef) (S – WS) and R(ight) (SW – S) word boundary, given per month

Deletion and maintenance in rhythmically optimal contexts



Graph 3: Adjacent strong syllable at the left boundary of a target starting with a weak syllable and the right boundary of a target ending with a weak boundary

6.3 Intonational phrase boundaries

Finally, in this section children's production of the target words at intonational boundaries are presented. First, we present the results of syllable deletion when the target word was at the beginning or at the end of an utterance containing more than one word. Then, the results for words that were produced in isolation – therefore, presenting intonational boundaries on both sides – are discussed.

6.3.1 Intonational phrase at one side of the word

Table 5 and Graph 4 show the result for the context in which the initial weak syllable of target words is at the beginning of the utterances (left edge), and hence at the beginning of the intonational phrase. Deletion in this context would create

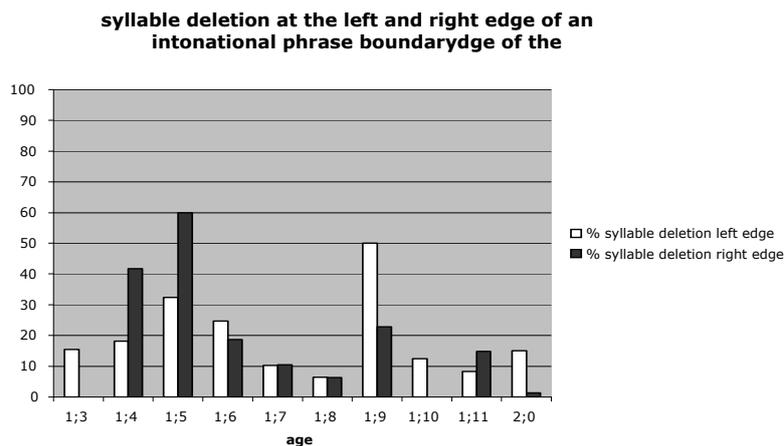
monosyllables ($_{IP}[WS \gg _{IP}[S]$) or trochees ($_{IP}[WSW \gg _{IP}[SW]$). As we can see from Table 5 and Graph 4, from the onset of speech children more often maintained the weak initial syllables than that they deleted those syllables. Some examples are given in (17) and (18).

Graph 4 also depicts the resulting production forms for target words with a weak final syllable at the end of the intonational phrase. In this context deletion would result in monosyllables ($(SW)_{IP} \gg S_{IP}$) or iambs ($(WSW)_{IP} \gg WS_{IP}$). As can be seen below, until 1;5 there was almost an equal proportion of deletion and maintenance of final weak syllables. After 1;6, the final weak syllable is more often kept than deleted. Some examples are given in (19) and (20).

- (17) [abila] aBRIR lá “open it there” R. 2;0
 (18) [kOlati] saCOla aQUI “bag here” R. 1;6
 (19) [adelabOla] caDÊ a BOla “where’s the ball?” R. 1;6
 (20) [abo] a BOla “the ball” R. 1;7

Age	1;3	1;4	1;5	1;6	1;7	1;8	1;9	1;10	1;11	2;0
Syllable deletion L IP	2	6	12	18	6	3	18	1	3	8
Maintenance L IP	11	27	25	55	52	43	18	7	33	45
Syllable deletion R IP	6*	25	24	17	4	5	21	0	7	1
Maintenance R IP	1*	35	16	74	34	73	71	5*	40	78

Table 5: Raw numbers of instances of syllable deletion and maintenance at the L(ef) (S – WS) and R(ight) (SW – S) edge of an intonational phrase boundary



Graph 4: Weak syllable deletion at the edges of intonational boundaries

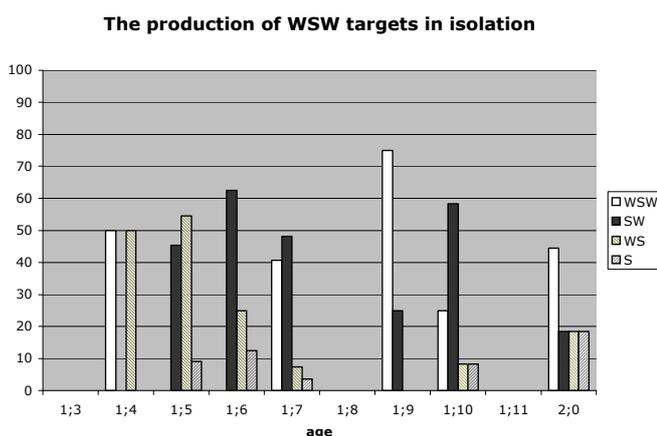
6.3.2 Targets in one-word utterances

Table 6 and Graph 5 show the results of a sub-group of the data analyzed in graph 4: WSW target words that are produced in isolation, and therefore have an IP boundary at both the right and the left edge of the word. Although this group is not large, it is of interest as here, deletion at the left boundary would create trochees ($[WSW]_{IP} \gg [SW]_{IP}$), and deletion at the right boundary iambs ($[WSW]_{IP} \gg [WS]_{IP}$). Of course, children could in principle also delete both weak syllables and produce a monosyllable: $[WSW]_{IP} \gg [S]_{IP}$. As we can see in Graph 5, until 1;5 children more often deleted the final weak syllable, thereby producing iambs, than the initial weak syllable (as also shown by Santos 2006, and Fikkert and Santos 2005). However, from 1;6 onwards trochees outnumbered iambs, until at 1;7 WSW productions appeared in the children's output. There were only a few cases in which both weak syllables were deleted. Examples are given in (21)–(24).

- (21) [ˈsa] saCOla “bag” R. 1;6
 (22) [kaˈwa] caVAlo “horse” R. 1;6
 (23) [ˈkOla] saCOla “bag” R. 1;7
 (24) [miˈniːna] meNIːna “girl” R. 2;0

Age	1;3	1;4	1;5	1;6	1;7	1;8	1;9	1;10	1;11	2;0
WSW	0	5	0	0	11	0	6	6	1*	12
SW	1*	0	5	5	13	2*	2	14	0	5
WS	6*	5	6	2	2	0	0	2	0	5
S	0	0	1	1	1	0	0	2	0	5

Table 6: Raw numbers of instances of realization of WSW target, in isolation



Graph 5: Targets with both weak initial and final syllables at IP boundaries

7. Conclusion

The data presented in this paper show that in contexts that would be rhythmically improved by either syllable deletion or insertion, children do not use this strategy systematically. Deletion is not very common in either context. Insertion of a syllable is very rare. In section 3 we raised the hypothesis that the shape of early words could be influenced by the context in which these words appear. We based our hypothesis on the fact that (a) children pay attention to the salient characteristics of the utterances (strong syllables are more prominent than weak ones, lengthened syllables are more prominent than syllables that do not undergo final lengthening), and (b) the assumption that children will optimise the rhythmic structure of an utterance (and hence avoid producing lapses and clashes). Many studies (for example, Cutler and Butterfield 1990, 1992; Cutler and Norris 1988, Mehler, Dommergues, Frauenfelder and Segui 1981, and Otake, Hatano, Cutler and Mehler 1993, among others) have pointed out that adults use strategies based on the rhythmic properties of their native language to segment speech. Other studies have shown that from an early developmental point, children can distinguish rhythmic classes (see Nazzi, Bertoni and Mehler 1998, among others) and discriminate words based on word stress (see Sansavini et al 1997, among others).

From the results in this we can conclude that the context in which target words appear does not influence the prosodic shape of early word production by children acquiring Brazilian Portuguese. The words in two and multiword utterances show the same pattern as those in isolation. They are either produced correctly, or favor iambic patterns. Of course, one-word utterances outnumber the two- and multiword utterances, but the hypothesis that word stress should in fact be interpreted as sentence stress is not confirmed by our analyses. Hence, it seems that for sentence prosody a different strategy is used, which is independent from the acquisition of word stress. We therefore argue that the word prosodic patterns are not based on sentence stress, but reflect children's knowledge of the stress patterns of words.

References

- Allen, G. D. and S. Hawkins (1978) 'The development of phonological rhythm', in: A. Bell and J.B. Hooper (eds.) *Syllables and Segments*. Amsterdam: North-Holland, 173–185.
- Allen, G. D. and S. Hawkins. (1980) 'Phonological rhythm: definition and development', in: G. Yeni-Konishian, J.F. Kavanagh and C.A. Ferguson (eds.) *Child Phonology-production 1*, New York: Academic Press, 227–256.
- Archibald, J. (1995) 'The acquisition of stress', in: J. Archibald (ed.) *Phonological Acquisition and Phonological Theory*. Hillsdale, N. J.: Lawrence Erlbaum, 81–109.
- Baia, M. F. (to appear) 'O pé trocaico frente a dados infantis do português brasileiro', to appear in: *Letras de Hoje: Anais do VII ENAL*, PUC–RS.
- Bonilha, G. F. G. (2005) *Aquisição Fonológica do Português Brasileiro: uma abordagem conexionalista da teoria da otimidade*. PhD dissertation, Pontifícia Universidade Católica do Rio Grande do Sul.
- Cho, T. and P. Keating (2001) 'Articulatory strengthening at the onset of prosodic domains in Korean'. *Journal of Phonetics* 28, 155–190.

- Cintra, G. (1997) 'Distribuição de padrões acentuais no vocábulo em português'. *Confluência* 5 (3), 83–92.
- Cutler, A. and S. Butterfield (1990) 'Durational cues to word boundaries in clear speech'. *Speech Communication* 9, 485–495.
- Cutler, A. and S. Butterfield (1992) 'Rhythmic cues to speech segmentation: Evidence from juncture misperception'. *Journal of Memory and Language* 31, 218–236.
- Cutler, A. and D. Norris. (1988) 'The role of strong syllables in segmentation for lexical access'. *Journal of Experimental Psychology: human perception and performance* 14, 113–121.
- Demuth, K. (1996) 'The prosodic structure of early words', in: J. Morgan and K. Demuth (eds.) *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*. Hillsdale, N. J.: Lawrence Erlbaum, 171–184.
- Demuth, K. and J. Fee (1995) 'Minimal words in early phonological development', ms. Brown University and Dalhousie University.
- Echols, C. H. and E. Newport (1992) 'The role of stress and position in determining first words'. *Language Acquisition* 2, 189–220.
- Fikkert, P. (1994) *On the Acquisition of Prosodic Structure*. PhD dissertation Leiden University, Holland Institute of Generative Linguistics.
- Fikkert, P. and R. S. Santos (2005) 'The Acquisition of word stress: bottom-up or top-down: a cross-linguistic perspective'. Paper presented at the *V Workshop on Phonological Acquisition*. Radboud University Nijmegen, The Netherlands.
- Gerken, L-A. (1994) 'A metrical template account of children's weak syllable omissions from multisyllabic words'. *Journal of Child Language* 21, 565–584.
- Grimm, A. (2004) 'The prosodic pattern of words and phrases in the acquisition of German'. Paper presented at the *9th Conference on Laboratory Phonology*.
- Grimm, A. (2006) 'Intonational patterns and word structure in early child German', in: D. Bamman, T. Magnitskaia and C. Zaller (eds.) *Proceedings of the 30th annual Boston University Conference on Language Development Volume 1*. Somerville, MA: Cascadilla Press, 237–248.
- Hayes, B. (1995) *Metrical Stress Theory – Principle and Case Studies*. Chicago: The University of Chicago Press.
- Jusczyk, P., A. Cutler and N. Redanz (1993) 'Preference for the predominant stress patterns in English words'. *Child Development* 64, 675–687.
- Lee, S-H. (1995) *Morfologia e Fonologia Lexical do Português Brasileiro*. PhD dissertation, Universidade Estadual de Campinas, Campinas.
- Lemos, C. T. G. (1995) 'Aquisição da Linguagem'. *Report to FAPESP 1976/1384 FAPESP 1985/3325-6 CNPq 404672-88.4 FAEP 190/90*, ms. Universidade Estadual de Campinas.
- Major, R. (1985) 'Stress and Rhythm in Brazilian Portuguese'. *Language* 61 (2), 259–282.
- Massini-Cagliari, G. (1992) *Acento e Ritmo*. São Paulo: Contexto Ed.
- Mehler, J., J.Y. Dommergues, U. Frauenfelder and J. Segui (1981) 'The syllable's role in speech segmentation'. *Journal of Verbal Learning and Verbal Behavior* 20, 298–305.
- Moraes, J. A. (1987) 'Correlats acoustiques de l'accent de mot en Portugais Brésilien' *Proceedings of the International Congress of Phonetic Sciences* 3, 313–316.
- Nazzi, T., J. Bertoncini and J. Mehler (1998) 'Language discrimination by newborns: toward an understanding of the role of rhythm'. *Journal of Experimental Psychology: Human perception and performance* 24 (3), 756–766.
- Nespor, M. and I. Vogel (1986) *Prosodic Phonology*. Dordrecht: Foris Publications.
- Otake, T, G. Hatano, A. Cutler and J. Mehler. (1993) 'Mora or syllable? Speech segmentation in Japanese'. *Journal of Memory and Language* 32, 258–278.

- Rapp, C. (1994) *A Elisão de Silabas Fracas nos Estágios Iniciais da Aquisiãõ da Fonologia do Português*. Master Dissertation, Universidade Federal da Bahia.
- Sansavini, A., J. Bertoncini and G. Giovaneli (1997) 'Newborns discriminate the rhythm of multisyllabic stressed words'. *Developmental Psychology* 33 (1), 3–11.
- Santos, R. (2001) *A Aquisiãõ do Acento Primário em Português Brasileiro*. PhD dissertation. Departamento de Lingüística, Universidade Estadual de Campinas, Campinas.
- Santos, R. (2003) 'Bootstrapping in the acquisition of word stress in Brazilian Portuguese'. *Journal of Portuguese Linguistics* 2(1), 93–114.
- Santos, R. (2006) 'Word Stress Acquisition: A Comparison between Brazilian Portuguese and Dutch', in: D. Bamman, T. Magnitskaia and C. Zaller (eds.) *Proceedings of the 30th annual Boston University Conference on Language Development Volume 2*. Somerville, MA: Cascadilla Press, 525–536.
- Scarpa, E. M. (1999) 'Sons preenchedores e guardadores de lugar: relações entre fatos sintáticos e prosódicos na aquisição da linguagem', in: E. Scarpa (ed.) *Estudos de Prosódia* Campinas: ed. Unicamp, 253–284.
- Vihman, M., R. DePaolis and B. Davis (1998) 'Is there a trochaic bias in early word learning? Evidence from infant production in English and French'. *Child Development* 69, 935–949.
- Waterson, N. (1971) 'Child Phonology: a prosodic view'. *Journal of Linguistics* 71, 179–211.

Null and Overt Subjects at the Syntax-Discourse Interface

Evidence from monolingual and bilingual acquisition

Ludovica Serratrice
The University of Manchester

1. Introduction

Referential expressions such as definite descriptions and pronouns are crucial to the processes of language comprehension and production. Understanding which definite description a pronoun refers to, and knowing when to use a pronoun instead of a definite description, are integral to a speaker's linguistic competence, and a fundamental aspect of language development. Choosing the most appropriate referring expressions poses a special challenge inasmuch as there is no one-to-one correspondence with entities in the real world. For example, the same individual (e.g. William Shakespeare) can be identified as 'Shakespeare', 'the bard', 'he', etc. Which expression will be used in a given context will depend on a range of discourse factors including the accessibility of the referent, its uniqueness, and its topicality, to name but a few. Moreover, in null-subject languages, there is the additional option to have a null subject pronoun whose distribution is similarly regulated by discourse factors.

In this paper I will investigate the extent to which children understand and produce lexical noun phrases (NPs) and null and overt pronouns in appropriate discourse contexts in two languages like English and Italian which behave similarly in the distribution of lexical NPs, but differ in terms of the availability of null and overt pronominal forms. I will present evidence from comprehension and production, and I will draw upon both corpus and experimental data to illustrate the role played by different discourse cues. I will also use data from English-Italian and Spanish-Italian bilingual children to explore the robustness of discourse-pragmatic competence in situations of language contact. Before reviewing the results of these studies (section 4 and section 5), I will first outline the cross-linguistic differences between the distribution of null and overt pronominal subjects in English and Italian (section 2), and I will present a brief overview of studies on subject argument realization in the acquisition literature (section 3). In section 6 I will draw a number of conclusions from the evidence reviewed, and I will present some issues for future research.

2. Cross-linguistic differences in the distribution of pronominal subjects

With a few notable exceptions, usually regarded as instances of topic drop in root contexts illustrated in (1b) (Haegeman 1997), English is a language where third person pronominal subjects need to be expressed overtly as shown by the grammaticality contrasts in (1):

- (1) a. Laura_i said good bye to her friend and she_i got out of the car.
 b. Laura said good bye to her friend and got out of the car.
 c. Laura_i said good bye to her friend while she_i got out of the car.
 d. *Laura said good bye to her friend while got out of the car.

In Italian, by contrast, null subjects (*pro*) are not only allowed in both root (e.g. (2b)) and non-root contexts (e.g. (2d)), but they are also the only pragmatically felicitous option in certain discourse conditions, as exemplified in (2a) and (2c):

- (2) a. ??Laura_i ha salutato la sua amica e lei_i è uscita dalla macchina.
 Laura has said bye to her friend and she is gone out from the car
 “Laura said good bye to her friend and she got out of the car”
 b. Laura_i ha salutato la sua amica e *pro*_i è uscita dalla macchina.
 Laura has said bye to her friend and *pro*_i is gone out from the car
 “Laura said good bye to her friend and (she) got out of the car”
 c. ??Laura_i ha salutato la sua amica mentre lei_i è uscita dalla macchina.
 Laura has said bye to her friend while she is gone out from the car
 “Laura said good bye to her friend while she got out of the car”
 d. Laura_i ha salutato la sua amica mentre *pro*_i è uscita dalla macchina.
 Laura has said bye to her friend while *pro*_i is gone out from the car
 “Laura said good bye to her friend while (she) got out of the car”

Grimshaw and Samek-Lodovici (1998) originally captured the complementary distribution of *pro* and overt subject pronouns in Italian in terms of the status of the antecedent. Null subjects are the pragmatically optimal option when they are co-referential with a topic antecedent. More recently, on the basis of off-line and on-line psycholinguistic evidence, Carminati (2002) proposed that overt and null subject pronouns are distinguished by the bias towards different syntactic positions of their antecedent, a proposal formalized in the Position of Antecedent Strategy (PAS):

“The null pronoun prefers an antecedent in the sentence which is in the Spec IP position, while the overt pronoun prefers an antecedent which is not in the Spec IP position.” (Carminati 2002: 41)

Carminati (2002) showed that even in the absence of semantic or discourse pragmatic cues, adult Italian readers prefer a subject antecedent in sentences like (3a) 81% of the time, while they opt for an object antecedent in sentences like (3b) 83% of the time:

- (3) a. Marta_i scriveva frequentemente a Piera quando *pro*_i era negli Stati Uniti.
 Marta wrote frequently to Piera when *pro* was in the United States
 “Marta wrote frequently to Piera when (she) was in the United States”

- b. Marta scriveva frequentemente a Piera_i quando lei_i era negli Stati Uniti.
Marta wrote frequently to Piera when she was in the United States.
“Marta wrote frequently to Piera when she was in the United States”

The PAS captures both the syntactic and the discourse pragmatic constraints that regulate the complementary distribution of overt and null pronominal subjects in Italian. From a syntactic point of view the position of the antecedent is identified in the canonical subject position, i.e. the spec of IP. From a discourse-pragmatic perspective the choice between an overt and null pronominal is justified in terms of the relative accessibility of their antecedent (Ariel 1994). More salient subject antecedents are later referred to by using a maximally reduced form such as *pro*; less salient non-subject antecedents require a form that will signal the shift of topic. Interestingly, Carminati (2002) suggests that the PAS does not only hold for Italian, but applies cross-linguistically. In non-null-subject languages like English the only difference is that unstressed pronouns cover the ground that in Italian is shared between *pro* and overt pronouns. At the same time there is evidence that unstressed subject pronouns in English are biased towards a subject antecedent (Gernsbacher and Hargreaves 1989; Hudson-D’Zmura and Tanenhaus 1998). In conclusion, in both languages the minimal referential expression available in the language, *pro* in Italian and unstressed personal pronouns in English, privileges a subject antecedent.

3. Approaches to subject omission and realization in child language

Given the linguistic background outlined in the previous section, the obvious question from a developmental perspective is to ask whether children are sensitive to cross-linguistic differences in the syntactic and discourse-pragmatic constraints that regulate the distribution of null and overt subjects in the adult language. Researchers have known for a long time that, regardless of whether null subjects are grammatical or not in the adult target, children go through an initial phase in which they omit subject arguments. Over the past twenty years the research focus has mostly been on children learning non-null-subject languages like English, where, with rare exceptions, subject omission is ungrammatical. A number of explanations have been proposed to account for this pervasive phenomenon in child language, ranging from syntactic limitations, performance limitations, to phonological limitations (see Guasti 2002 for a recent review). At the same time, with a few exceptions (Grinstead 2000; Bel 2003; Lorusso, Caprin and Guasti 2005), much less attention has been devoted to subject omission in null-subject languages in which argument ellipsis is syntactically licensed but not always pragmatically appropriate.

In Hyams’s (1986) original proposal, English-speaking children’s failure to realize subjects overtly in obligatory contexts was ascribed to the mis-setting of the *pro*-drop parameter to the positive value. Hyams’s argument was that these children were essentially treating English as if it were Italian, and therefore considered null subjects to be a viable syntactic option. Subsequent work started to cast doubts on the mis-setting hypothesis as this approach was not consistent with the absence of null subjects in non-root contexts such as *wh*-questions and subordinate clauses

(Valian 1991). Working from a cross-linguistic perspective Valian (1991) and Valian and Eisenberg (1996) crucially showed significant differences in the rate of subject omission in children learning a non-null-subject language like English, and in children learning either Italian or Brazilian Portuguese, both null-subject languages. According to Valian and colleagues this discrepancy in subject omission was consistent with children's sensitivity to the frequency of null subjects in the input coupled with limited processing capacity and working memory limitations.

Another attempt to account for the nature of children's early null subjects in non-null subject languages is Rizzi's (1994) truncation hypothesis where he captures the co-occurrence of null subjects with non-finite forms in root contexts. In this proposal early grammars differ from adult grammars inasmuch as they allow for clausal truncation below the topmost Complementizer Projection and for an antecedentless Null Constant in the specifier of the Inflectional Phrase.

Other accounts of early null subjects have focused more on children's limited processing capacities as the underlying cause for the omissions (Bloom 1990). In this view children omit subjects because of their clause-initial position which is also the position with the highest processing demands.

As mentioned earlier, a small number of studies have specifically investigated the omission of subjects in languages where subject omission is indeed grammatical in the adult language. Grinstead (2000) examined subject argument realization in children learning Catalan and Spanish and related children's omissions to the inactivity of the topic-focus field in the Complementizer system. Grinstead's data and interpretation was later questioned by Aguado-Orea and Pine (2002) and by Bel (2003) who did report an alternation between null and overt subjects from the earliest stages of verb production. Bel (2003) also presented data showing that the Spanish-speaking children and the Catalan-speaking children in her study produced subjects both in the preverbal and the postverbal position, the latter being favoured in the case of unaccusative verbs. Lorusso, Caprin and Guasti (2005) found similar results for subject realization in Italian with the SV order being more frequent in the case of transitive and unergative verbs, and the postverbal position being favoured in the case of unaccusative verbs.

Although the studies reviewed so far on the distribution of null and overt subjects in Spanish, Catalan and Italian child language have focused on slightly different aspects of the grammar of the three languages, one common denominator is the lack of reference to the discourse-pragmatic contexts in which subject argument were overtly realized or phonologically null. Crucially, however, the choice of referential expressions is constrained both by the morpho-syntactic availability of a range of forms in the language (i.e. definite and indefinite NPs, demonstrative pronouns, overt personal pronouns, *pro*, etc.), and by the discourse-pragmatic status of the referent in question. Focusing exclusively on the syntactic aspect of argument realization and neglecting the role of discourse-pragmatics can provide but a partial picture of the phenomenon of referential choice.

Over the last few years a number of researchers have started to investigate more carefully the role played by discourse pragmatics, and in particular by the informativeness status of a referent, in predicting whether an argument will be

realized overtly or not in child language (Clancy 1993; Allen 2000; Allen and Schröder 2003; Serratrice, Sorace and Paoli 2004; Serratrice 2005; Guerriero 2006; Hughes and Allen 2006; Skarabela 2006; Skarabela 2007). The concept of informativeness goes back to Greenfield and Smith's (1976) formulation of the principle according to which informative arguments, i.e. arguments whose referents are not highly salient and/or accessible, are more likely to be realised overtly than uninformative arguments, i.e. arguments with highly salient and accessible referents. The notion of a referent's salience and accessibility is in turn understood in terms of a number of binary features such as person, activation and disambiguation which can be systematically coded for informativeness. First and second person referents are always highly salient and accessible as they identify the speaker and the addressee. As a consequence these referents' informativeness status will be low and they are more likely to be omitted by children. In contrast, the search space for third person referents is considerably wider, their salience and accessibility is inevitably lower than for first and second person referents, and their informativeness status is higher. Because of their high informativeness status, the likelihood that third person referents will be omitted is therefore lower.

Activation is a notion deriving from Chafe's work on the use of referring expressions (Chafe 1994, 1996). Chafe argues for the existence of three activation states in the hearer's mental representation of discourse: active, semi-active and inactive. An active referent is one that is salient in the hearer's consciousness at a particular moment in time. An active referent is highly accessible either through recent prior mention, or through being the focus of the speaker' and hearer's joint attention. A semi-active referent is one that is in the hearer's peripheral consciousness, he or she has a background awareness of it, but it is not currently the focus of attention. Finally, inactive referents have neither been introduced linguistically, nor are they physically present. Semi-active and inactive referents have a high informativeness status because they are in greater need to be identified, and as such they are less likely to be omitted than active referents which can be easily recovered from the context.

Finally, whenever the discourse includes two or more referents with the same activation state, subsequent reference to either of them needs to be disambiguated by a maximally informative overt referring expression, generally an NP.

Clancy (1993) led the way in the systematic investigation of the impact of a number of aspects of informativeness, or discourse prominence in her own terms, on the realisation of argument referents in Korean child language. Besides person, Clancy included five other discourse-pragmatic features: newness, contrast, query, absence, and animacy to code the children's arguments for discourse prominence. Each argument was coded as being either discourse-prominent or non-discourse-prominent with respect to the above features. Her findings corroborated Greenfield and Smith's (1976) prediction and showed that arguments that were third person, new to the discourse, in contrast with other arguments, the focus or response to a question, absent from the immediate physical surroundings, and inanimate, i.e. informative arguments, were significantly more likely to be realised overtly than arguments that were non-discourse-prominent, or uninformative.

Work by Allen and her associates (Allen 2000; Allen and Schröder 2003; Skarabela and Allen 2002; Hughes and Allen 2006) has made a substantial contribution to the analysis of argument realization in Inuktitut and in English child language in terms of informativeness. Allen (2000) and Allen and Schröder (2003) showed that the six discourse pragmatic factors initially identified by Clancy, plus two additional ones introduced by Allen (2000) (differentiation in discourse and differentiation in context), were reliable predictors of argument form as shown by a series of binary logistic regression analyses where argument form (null = 0; overt = 1) was the dependent variable, and the eight informativeness features were treated as the independent variables. Specifically, Allen (2000) reported that third person arguments were sixteen times more likely to be realised overtly than either first or second person arguments. Along similar lines, Skarabela and Allen (2002) analyzed argument realization in connection with joint attention in the spontaneous production of four Inuktitut-speaking children and convincingly showed that children were more likely to omit arguments in the presence of joint attention.

This section presented an overview of different approaches to subject argument realization in child language. My contention is that an account of the distribution of null and overt subjects cannot be satisfactory unless it considers the interface between syntax and discourse pragmatics. Only a proposal that works on both of these two levels can predict where children are more or less likely to resort to null forms.

In the following section I will present data from my own work over the last five years showing that this kind of approach is crucial not only to understanding spontaneous production data in monolingual children, as in the studies reviewed so far, but that it is equally important to the interpretation of comprehension data in both monolingual and bilingual children.

4. The role of discourse pragmatics in monolingual children's argument realization

With the exception of Paradis and Navarro (2003), who analysed the spontaneous production of a Spanish-English bilingual child and of three Spanish-speaking monolingual children between the ages of two and three, there is a dearth of studies analysing subject realization and omission in Romance null-subject languages from a discourse pragmatic perspective (but see Orsolini, Rossi and Pontecorvo 1996, for argument realization in Italian children's narratives). In view of this gap in the literature, in Serratrice (2005) I set out to investigate to what extent Italian-speaking children are sensitive to the discourse-pragmatic constraints on subject omission in a language where null subjects account for approximately 70% of all subject contexts in the adult input. The fact that the majority of subject arguments in adult Italian are phonologically null, coupled with children's limited processing capacity might lead us to expect that, at least in the initial stages, children may omit subjects across the board. Should this be the case they would not be contravening the syntactic constraints of their language where null subjects are indeed always syntactically grammatical, but they might well violate the discourse

pragmatic constraints that regulate the complementary distribution of null and overt subjects. The aim of the study was therefore to explore whether Italian-speaking children do actually use null subjects indiscriminately, or whether they resort to the null option only in those cases in which it is a pragmatically felicitous option.

The data used in the analyses were taken from the Calambrone corpus (Cipriani et al. 1989) available on CHILDES (MacWhinney 2000), a collection of 79 recordings of spontaneous interaction between six children, aged 1;7-3;3, and adult caregivers. I coded the subject arguments of 3667 non-imperative verbs in fully intelligible utterances including a declarative sentence. I then divided the children's output into three MLUw stages (Stage I: MLUw 1.5-2, Stage II: MLUw: 2.0-3.0; Stage III: MLUw 3.0-4.0; Stage IV: MLUw > 4.0), and I coded the subject arguments for overtness and for three informativeness features: person, activation state, and disambiguation. My hypothesis was that children are indeed sensitive to the discourse-pragmatic constraints that regulate the distribution of referential expressions from an early age. I therefore predicted that informative arguments would be significantly more likely to be realized overtly than uninformative arguments. The results confirmed this hypothesis and showed that from as early as MLUw 2.0, referents that were third person, highly active, and with more than one potential antecedent, were realized overtly significantly more often than first and second person, semi-active, inactive and unambiguous referents.

Neither a syntactic approach (Grinstead 2000), nor a performance deficit account (Bloom 1990; Valian 1991; Valian and Eisenberg 1996), can offer a satisfactory explanation for the selective omission of subjects in contexts where the features person, activation and disambiguation are uninformative. From a syntactic point of view, null subjects in Italian are always grammatical, therefore an approach that is based solely on syntax cannot provide a principled explanation of why subjects are more likely to be realised overtly in third person contexts than in non-third person contexts, or in contexts in which the referent is inactive or ambiguous.¹

The pattern of subject drop in the data is also problematic for a developmental account that relies exclusively on a generic performance limitation (Bloom 1990;

¹ In an original syntactic approach to subject omission in child language, van Kampen (2004, 2006) proposes a tripartite division of verbal predicates into bare predicates (type a), situation-bound predicates (type b), and discourse-bound predicates (type c). According to van Kampen's analysis, only type c predicates allow for real instances of pro-drop in third person contexts. Type c predicates are finite verbs that show full Agreement and follow the acquisition of D-marking (articles, possessives, demonstratives). Earlier occurrences of null subjects are instead treated as instances of mood-implied subjects, with first person subjects implied by the optative mood, and second person by the imperative mood. In addition, unlike type c third person null subjects, type b ones are lexically restricted to a limited number of verbs. Although van Kampen's approach is couched in morpho-syntactic terms, it also draws on the semantics of the predicates' mood, on the discourse-pragmatic contexts in which the different type of predicates occur, and on the lexical specificity with which some predicates occur. It is therefore a more articulate proposal than the one proposed by Grinstead (2000) which relies exclusively on Case marking, and therefore a more viable account of the observed differences in the rate of subject omission.

Valian 1991; Valian and Eisenberg 1996). A performance limitation account would predict that heavier nominals such as demonstratives, definite NPs, and proper nouns should be more likely to be omitted than lighter monosyllabic first and second person subject pronouns like 'io' and 'tu'. Although it is sensible to assume that children will be initially constrained by somewhat limited processing abilities, both in comprehension and in production, nevertheless the data clearly show that these limitations can be overridden in cases in which the discourse and pragmatic conditions require the expression of crucial information. In the case of argument realization in Italian this translates into the use of an overt rather than a null subject when the informativeness status of the referent is high.

The findings from the analysis of the spontaneous production of Italian pre-schoolers in Serratrice (2005) provided a useful starting point for a better understanding of subject argument realization in null-subject languages. There are however obvious limitations in the kind of analyses one can conduct on data coming from unstructured conversations between young children and adult caregivers. Firstly, communication tends to revolve around the here and now. In this type of setting the vast majority of referents are typically accessible to both interlocutors. It is therefore difficult to know how children would behave in contexts in which they have to take into account their listener's point of view in the identification of a referent that is accessible to them but not to their listener. Secondly, in this kind of naturalistic data with young children, topic continuity is often an issue. Frequent changes of topic make it difficult to track referents over time as the discourse unfolds. To overcome these two main limitations of naturalistic data I recently conducted two studies investigating argument realization in Italian and in English. The first is a narrative study of referent introduction, re-introduction and maintenance in the Frog Story in eight-year-olds (Serratrice 2007a), the second includes two experiments in which I manipulated variables such as the type of questions pre-schoolers were asked by an adult interlocutor, the accessibility of the referent to the listener, and the number of referents that the children were required to identify linguistically (Serratrice, in press).

In Serratrice (2007a) I elicited narratives from three groups of eight-year-old children using the Frog Story: Italian-speaking monolinguals, English-speaking monolinguals and English-Italian bilinguals. The main aim of the paper was to assess whether bilingual and monolingual children differ in the way in which they use global and local markings of the new-given distinction in each of their two languages in the context of referent introduction, re-introduction and maintenance. In terms of global marking I was interested in the clause-initial vs. clausal-final position of arguments. As far as local marking was concerned I looked at the type of referential expressions used to code both subject and object arguments. All referents in each fully intelligible clause were coded for argument status (argument, adjunct, post-copular predicate), syntactic function (subject, direct object), morpho-syntactic form (null subject pronoun, overt subject pronoun, strong object pronoun, clitic object pronoun, definite/indefinite NP), and discourse function (referent introduction, re-introduction and maintenance). Subject arguments were also coded for word order with respect to the verb (pre-verbal, post-verbal). The first mention

of a referent was coded as an introduction. Subsequent mentions of the same referent were divided into two categories: re-introduction and maintenance. A referring expression was coded for re-introduction if it expressed a subject or an object argument that was not mentioned in the immediately preceding clause, and/or if it expressed a subject argument whose immediate antecedent was in object position. In all other cases a subsequent mention of a referent was coded as maintenance.

The inclusion of two typologically different languages (English and Italian), and two populations of language learners (monolinguals and bilinguals) allowed for cross-linguistic comparisons between English and Italian, and for within-language comparisons between children learning one vs. two languages. In the following I will focus on the cross-linguistic differences and similarities in terms of children's use of word order and morphosyntactic forms to mark the new-give distinction. I will comment on the similarities and differences between the bilinguals and the monolinguals later on in this paper.

At the level of global marking there were statistically significant cross-linguistic differences, especially for referent introduction. In Italian referents were largely introduced in post-verbal subject position while in English the preference was for the pre-verbal position, although there were a number of stylistically marked postverbal subjects even in English (e.g. "And out popped/came the owl"). This confirms the status of English as a rigid SVO language as opposed to Italian where word order is more flexible, and the post-verbal position is typically associated with narrow focus, the ideal position for the placement of new information. In both languages referents were typically re-introduced and maintained in the preverbal subject position although I did find significant differences at the level of the local marking in terms of the type of referential expressions used. Referents were generally re-introduced by lexical NPs in both languages; however in Italian a non-negligible proportion of null subjects were also used to re-introduce referents. As suggested elsewhere by Orsolini et al. (1996), the use of null subjects for semi-active referents might have been justified by the presence of other semantic or pragmatic cues that allowed for the identification of the referents in question, although I did not specifically code for this aspect of the data. The different null-subject status of the two languages was especially clear in the frequency of overt pronominal subjects used for referent maintenance. In English, overt pronouns for this discourse function were employed around 60% of the time, while in Italian they were used in only 3% of contexts.

The use of elicited narratives in this study allowed me to overcome one of the aforementioned restrictions of spontaneous production with young children, i.e. the rather limited scope of topic continuity, and hence the difficulty to track the identification of referents over a substantial number of turns. The findings clearly show that both bilingual and monolingual children use language-specific discourse-appropriate referential expressions in the realization of subject arguments in the course of a complex narrative.

In a second study (Serratrice, in press) I addressed another limitation of naturalistic data: the lack of scope to assess children's sensitivity to their interlocutor's access to referents. In the typical situation in which recordings of

child-adult interaction are conducted, conversations with young children tend to revolve around referents that are generally physically present and accessible to both interlocutors. In this kind of setting the children's perspective mostly coincides with the adults', there is therefore little scope for exploring how children would behave in the identification of referents when their point of view is different from their interlocutors'. In two production experiments with English-speaking three- five- six-year-olds, and adults, I addressed this issue and manipulated the accessibility of the referent (accessible to both child and experimenter/ accessible only to child) and two other variables: the number of referents (one/two) and the type of question the experimenter asked the child (general/specific) to investigate whether children of different ages can indeed use the most appropriate referential expressions in a range of different discourse contexts.

During both tasks the participants were presented with a series of colour drawings of simple transitive events with a human subject acting upon an inanimate object (e.g. a man peeling a carrot, a girl eating an ice cream) and had to answer a question posed by the experimenter. The results of the first experiment confirmed that, like adults, children of all ages were sensitive to the type of question they were asked. They produced significantly more lexical NPs than pronouns when they answered a general question like "What's happening here?" as opposed to a specific question like "What's that person doing?". The assumption behind the general question was that both the predicate and the arguments constitute new information, as such they need to be realized by a maximally informative referential expression like a lexical NP. By contrast, the presupposition behind a question like "What's that person doing?" is that only the predicate constitutes new information, the realization of the subject argument by the lexical NP "that person" allows for the use of a reduced referential expression like an overt pronoun in English, or even null reference. The effect of accessibility on subject argument realization was less clear-cut. Whether the experimenter could see the referent or not did not make any difference for the three-year-olds, but it did for the older children. Even when the question was specific, and a pronoun would have been justified in terms of the type of information solicited, both the five- and the six-year-olds produced more lexical NPs when the experimenter could not see the picture. Their proportion of lexical NPs in this condition was still significantly lower than the adults', but they were able to override the cue provided by the question when the referent was not accessible to their interlocutor. In the second experiment the question was always specific; alongside accessibility of the referent to the experimenter, I manipulated the number of human subjects in the pictures: either one or two people of different gender. The results showed that the three-year-olds did not produce more informative expressions in the two-referent contexts, while the adults and the older children did. I take this as evidence that, from at least five onwards, children are sensitive to properties of the referential scene when they choose referring expressions. The presence of two referents made it necessary to use a more informative referential expression for disambiguation purposes and they complied with this extra-linguistic constraint. One surprising finding was the fact that rather than producing gender-marked personal pronouns, both the adults and the children

tended to use maximally informative lexical NPs in the two-referent condition. This is possibly motivated more by the speaker's internal processing needs rather than by the need to provide an unambiguous referent for the benefit of the listener, as suggested by Arnold and Griffin (2007). Overall the second experiment confirmed that, like the adults, the older children were sensitive to the accessibility of the referent to the listener and used more lexical NPs when the experimenter could not see the pictures.

In sum, data from naturalistic and experimental studies, typologically different languages, different ages, and different learner populations repeatedly show that subject omission in child language can be reliably predicted by the informativeness status of the referent. In the following section I look more specifically at the use of null and overt subject pronouns in bilingual children in relation to the issue of cross-linguistic influence.

5. Overt and null pronominal subjects in bilingual children

As illustrated in Section 2, English and Italian vary with respect to the distribution of null and overt pronominal subjects. In English, overt pronouns are appropriate both in referent maintenance and referent re-introduction contexts. By contrast, in Italian there is a division of labour between null pronouns for referent maintenance, and overt pronouns for referent re-introduction. Evidence from production data (Serratrice 2005, 2007a) shows that monolingual children consistently use language-specific pronominal forms for referent maintenance: null pronouns in Italian and overt pronouns in English. In the case of bilingual children who are simultaneously learning the two languages, the question is whether the division of labour between null and overt pronouns in Italian might be obliterated by the fact that such division does not exist in English. Although there is good evidence that bilingual children by and large treat their two languages as separate phonological, lexical and morpho-syntactic systems from early on (Meisel 1994; De Houwer 1990; Genesee, Nicoladis and Paradis 1995; Bosch and Sebastián-Gallés 2001), nevertheless some researchers have advocated the possibility that a degree of permeability may exist between the two under certain conditions (Döpke 1998; Hulk and Müller 2000; Müller and Hulk 2001; Bernardini 2003; Nicoladis 2006). With specific reference to the production of subject pronouns a number of studies have recently shown that children who are learning a null-subject language together with a non-null subject language have a tendency to use overt pronominal subjects in contexts in which a null subject would be pragmatically appropriate (see Paradis and Navarro 2003 for English-Spanish; Serratrice, Sorace and Paoli 2004 for English-Italian; Hacothen and Schaeffer, in press for English-Hebrew). This evidence has been taken to show that subject argument realization in the language with the less economical system (i.e. the language with both null and overt pronouns, where overt pronouns are associated with focus and topic shift) is vulnerable to cross-linguistic influence from the language with the more economical system (i.e. the language with only overt pronouns which are underspecified for focus and topic shift). In representational terms Serratrice et al. (2004) proposed that prolonged contact with

the underspecified setting of English might lead to the underspecification of the discourse pragmatic features associated with overt pronominal subjects in Italian. Because of the cross-linguistic influence from English, in the grammar of a bilingual child overt pronominal subjects are no longer specified for focus and topic shift in Italian. Due to their underspecified nature, overt subject pronouns in Italian would then be used in both focus or topic shift contexts and in no-focus and no-topic shift contexts, similarly to what has recently been observed for attrited L1 speakers of Italian and Greek (Tsimpli, Sorace, Heycock and Filiaci 2004). More recently, in Serratrice (2007b) I argued for an alternative processing account of cross-linguistic influence in the overuse of overt subject pronouns in English-Italian bilingual children. The claim is that the routine processing of English structures in which overt pronominal subjects are co-referential with a subject antecedent will increase the likelihood that overt pronouns will be processed in the same way in parallel Italian structures in what might be considered an instance of cross-linguistic priming. This proposal is in line with the notion of structural priming, i.e. the fact that both in comprehension and in production speakers have a tendency to produce sentences with previously heard or produced syntactic structures (see Ferreira and Bock 2006; Branigan 2006 for a recent overview). A number of recent studies have indeed shown that, provided that word order is the same cross-linguistically, this tendency to “re-use” syntactic structures persists across the two languages of a bilingual speaker (Loebell and Bock 2003; Hartsuiker, Pickering and Veltkamp 2004). Cross-linguistic structural priming and the probability that a speaker will strive for convergence via congruent lexicalization (Muysken 2002) is likely to be significantly affected by the extent to which the speaker/comprehender is operating in a bilingual vs. monolingual mode. With specific reference to the overuse of subject pronouns in null-subject languages, in a case study of two adult bilingual English-Spanish speakers Toribio (2004) showed that her participants used more overt subject pronouns in Spanish when they were in “bilingual mode”, i.e. when they were code-switching freely between English and Spanish in a narrative task, than when they were retelling a narrative in Spanish “monolingual mode”. The high activation of structures containing an overt subject pronoun in English in the bilingual mode primed the use of an overt subject when the participant switched to Spanish.

The different level of activation of the language driving convergence between the bilingual and the monolingual mode would also contribute to an explanation of a discrepancy between different sets of results in the production of English-Spanish and Italian-Spanish bilingual children. Contrary to what has been reported for three- and four-year-olds in naturalistic contexts I did not find any significant differences between the English-Italian and the monolingual Italian children in the proportion of overt subject pronouns in referent maintenance contexts in the Italian narrative task in Serratrice (2007a). The only significant difference was the less frequent use of clitic object pronouns for referent maintenance in Italian in the group of bilingual children. One of the reasons for the absence of a difference between the bilingual and the monolingual children in the narrative production task may well be the fact that the bilingual children in that study were specifically asked to tell the story in

either Italian or English to a silent experimenter, and as a consequence they were required to enter a “monolingual mode”. In contrast, the children in the other studies mentioned were in principle freer to switch back and forth between the two languages. Another possible explanation for these divergent findings is to be found in the fact that for ten of the twelve bilinguals in Serratrice (2007a) Italian was the language of the community, and therefore the language they heard and spoke more often, while in all the other cases the children were being raised in an environment in which English was the language of the community.

If structural priming is not only to do with the activation of recently processed structures as argued by some (Pickering and Branigan 1998; Branigan, Pickering and Cleland 2000), but it also involves implicit learning, as argued by others (Bock and Griffin 2000; Chang, Dell, Bock and Griffin 2000), then it is reasonable to assume that the more frequently a structure is encountered together with a certain meaning, the more likely it is that that meaning will be encoded using that structure. In terms of the cross-linguistic priming of pronouns, if referent maintenance is structurally encoded by an overt pronoun virtually 100% of the time in English, the language in which the bilingual child receives quantitatively more input, it is likely that in Italian too overt pronouns will have an increased likelihood of being treated as pragmatically appropriate forms for referent maintenance. If, vice versa, Italian is the language where the bilingual child receives most of her input, it is sensible to assume that the English co-referential structure will have less of a longer term impact on the child’s processing of pronouns in Italian, and hence we would expect fewer pragmatically inappropriate choices than in the former case of a bilingual child growing up with English as the community language.

To assess the role played by input in a more systematic way, in a recently completed grammaticality judgement study on null and overt subject pronouns in Italian (Sorace, Serratrice, Filiaci and Baldo, under review) we included two groups of English-Italian bilingual children: one in the UK and one in Italy. In line with Argyri and Sorace’s (2007) findings for UK-based and Greece-based Greek-English bilinguals, we expected that, because of the greater use of English both in comprehension and in production, the children living in the UK would be more likely than their counterparts in Italy to accept pragmatically infelicitous overt subject pronouns. This was indeed the case, but the effect of language of the community (English vs. Italian) also interacted significantly with age. The bilinguals in the UK accepted significantly more pragmatically inappropriate overt pronouns than both the Italian monolinguals and the bilinguals in Italy only at the age of 6-7; by the age of 8-10 the differences between the three groups were no longer significant. Besides language of the community and age, in this study we also manipulated language combination to try and tease apart the effect of cross-linguistic influence from English from the more general effect of bilingualism *per se*.² We therefore included another group of bilingual children who were learning

² Pinto (2006) independently proposed that children learning language pairs where the distribution of subjects follows similar \pm topic shift constraints, such as Spanish-Italian/Dutch-Italian, should make fewer errors in the distribution of overt pronouns than

two null-subject languages: Spanish and Italian. If even children learning two typologically similar languages accepted more pragmatically overt pronouns than monolingual Italian children, we would have to conclude that cross-linguistic influence cannot be the only reason for the English-Italian bilingual children's over-acceptance of redundant overt pronouns. The inclusion of the Spanish-Italian bilingual group showed that this was indeed the case. Even this group of children accepted more overt pronouns than their monolingual peers, although they did so only in the older age group (8-10-year-olds). Thus it seems that, regardless of the language combination, being bilingual increases the chances of choosing a pragmatically inappropriate option. Interestingly this was not only the case of overt pronouns in referent maintenance contexts, as we also found a significant difference between monolingual children and bilingual children – regardless of language of the community and of language combination – in the proportion of null subjects chosen in referent switch contexts; bilingual children chose more pragmatically infelicitous null subjects than their monolingual peers.

6. Conclusions and directions for future research

From the evidence presented here there are a number of conclusions to draw and possible directions to suggest for future research in the area of referential choice in monolingual and bilingual children. In terms of approaches to argument realization in child language, it should be clear that a principled account of how child speakers and comprehenders choose and interpret referential expressions, and especially third person pronouns, must take into consideration the discourse pragmatic constraints that regulate the distribution of these forms in the adult language. Syntactic accounts, such as Rizzi's (1994) truncation hypothesis may be necessary to explain cross-linguistic differences in the rate of subject omissions in root and non-root contexts, for example. Performance limitations are also undoubtedly going to affect children's output (e.g. Bloom 1990) and cannot be ignored. Ultimately however, an approach that assesses the referent's informativeness status is going to provide a more adequate explanation of the empirical data together with the opportunity to make falsifiable predictions. For example, neither a syntactic deficit nor processing limitations can explain why three-year-old English-speaking children are more likely to omit a subject when answering a question like 'What's this person doing?' than when answering a question like 'What's happening here?' (Serratrice, in press). If a syntactic layer is unavailable, the child should omit a subject regardless of the type of question she answers. By the same token, if a child is constrained by the higher processing load at the beginning of an utterance she should omit a clause-initial subject answering either of these questions. Moreover, a lexical NP including both a determiner and a noun should have a higher processing cost than a monosyllabic pronoun, and hence fewer chances of being realized overtly, but the exact opposite was found

children learning a language pair like English and Italian in which only one language (Italian) obeys the \pm topic shift constraint.

(Serratrice, in press). Children were more likely to produce a full lexical NP when answering a sentence-focus question like ‘What’s happening here?’ than a monosyllabic pronoun answering a predicate-focus question like ‘What’s this person doing?’. The information structure of the question clearly biased the children towards subject omission (younger children), the use of pronouns (younger children and adults) or of lexical NPs (older children and adults) in different discourse contexts. In terms of future research a way forward to tease apart the contribution of different aspects of children’s competence in the use of referential expressions could be to use binary logistic regression to investigate how much of the variance in subject argument realization is accounted for by syntax (e.g. root vs. non-root contexts), performance limitations (e.g. clause-initial vs. non-clause-initial contexts) and discourse pragmatics (informative vs. non-informative referents) (see Allen 2000, and Skarabela 2006 for a similar approach with different types of informativeness features).

Alongside the issue of argument realization in English- and in Italian-speaking monolingual children, this paper focused on the production and comprehension of null and overt subject pronouns in Italian in English-Italian and Spanish-Italian bilingual children. A number of studies including bilingual children learning a null-subject language and a non-null-subject language have shown that bilinguals tend to produce significantly more overt pronouns in the null-subject language in contexts in which their monolingual peers tend to use null pronouns. I initially presented an underspecification approach according to which prolonged and sustained contact with a language like English, where overt pronouns are not specified for either topic shift or focus, would lead to the underspecification of the topic shift and focus features associated with overt pronouns in Italian. Another possibility is that the grammatical representation of overt pronouns is not actually affected, but that the redundancy of overt pronouns in the production and comprehension of Italian is due to cross-linguistic structural priming. None of the studies in the bilingual literature have yet systematically addressed the possibility that redundant pronominal subjects may be the result of the activation of structures containing overt pronouns in the non-null-subject language, or even the outcome of implicit learning over time. At the same time there are no published studies in the priming literature that I am aware of testing whether the use of an overt pronoun in a referent maintenance context can be primed by the previous comprehension and/or production of an overt pronoun.³ A study exploring this possibility in bilingual speakers is the obvious next step. Preliminary findings from two studies investigating the priming of overt pronouns in monolingual Italian-speaking six- and eight-year-olds (Serratrice and Pinto, in preparation) indicate that children are indeed more likely to produce an overt

³ Clancy (in press) recently explored the possibility that priming might be involved in the production of morphologically marked arguments in Korean child language. In her analysis of the use of nominative, topic, and accusative markers in two Korean-speaking two-year-olds, Clancy reported that marked arguments were significantly more likely after the production of a morphologically marked argument in the preceding maternal utterance and she concluded that priming can decrease the child’s processing burden and hence facilitates production.

pronoun in a referent maintenance context when they have listened to and repeated, or even just listened to, a structure containing an overt pronoun. In one experiment the children listened to an adult's description of a picture (e.g. 'La bambina spruzza la mamma mentre lei prende il sole' 'The girl splashes her mother while she is sunbathing') and repeated it before describing a similar picture of their own. When the children's picture set up a referent maintenance context the six-year-olds used 25% more pragmatically redundant overt pronouns when the adult prime contained an overt pronoun (referent switch context) than when it contained a null pronoun (referent maintenance context). The comparable advantage for the eight-year-olds was 38%. We found a similar priming advantage in a second experiment in which the children listened to eight adult primes without repeating them and then proceeded to describe eight pictures themselves. The effect of priming was particularly strong considering that there was no lexical overlap between the prime and the target sentences, that in half of the trials the gender of the pronoun in the prime was different from the gender of the pronoun in the target, and the effect of gender was not found to be significant. These results suggest that monolingual children can be primed to produce a pragmatically redundant overt pronoun, and the next question to ask is whether this effect can be replicated across languages in bilingual children. Future research should address this issue both in naturalistic (see Travis 2007 for an analysis of priming in different spoken genres), and in experimental settings.

Another important finding in the studies addressing the issue of cross-linguistic influence relates to the results from the Spanish-Italian bilinguals in Sorace et al. (under review). We found that even children who are learning two null-subject languages chose significantly more redundant overt pronouns in contexts of referent maintenance than their monolingual Italian peers. This result has important implications for our understanding of differences between monolingual and bilingual acquisition. Recently discrepancies between bilinguals and monolinguals have been ascribed to the effect of exposure to the other language. For example in the case of object omission in Dutch-French and German-Italian bilingual children Müller and Hulk (2001) argued that the increased rate of null arguments in French and in Italian is due to the high frequency of null objects in Dutch and German respectively. Similar arguments for cross-linguistic influence of language A on language B have been made in many other studies. What is novel in the Sorace et al.'s study is the idea that bilingual children might behave differently from monolingual peers with respect to a certain construction even when their two languages do not differ in any obvious way with respect to the target construction. Future studies should address this question by comparing different bilingual populations on the same target construction; only in this way can the effects of language-specific cross-linguistic difference be teased apart from the more general effects of growing up with two languages.

Finally, another direction in which the study of referential expressions in bilingual first language acquisition should move is the use of on-line tasks that tap into children's interpretative processes. All the studies reviewed so far have either included naturalistic data or off-line tasks which largely draw on post-interpretative

processes. Experiments using the visual world paradigm (Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy 1995) would give us an insight into language processing as it happens in real time and would thus allow us to determine whether bilingual children actually treat anaphoric expressions any differently from their monolingual counterparts.

References

- Aguado-Orea, J. and J. Pine (2002) 'There is no evidence for a 'no overt subject' stage in early child Spanish: A note on Grinstead (2000)'. *Journal of Child Language* 29, 865-874.
- Allen, S. (2000) 'A discourse-pragmatic explanation for argument representation in child Inuktitut'. *Linguistics* 38, 483-521.
- Allen, S. and H. Schröder (2003) 'Preferred argument structure in early Inuktitut spontaneous speech data', in: J. W. Du Bois, L. Kumpf and W. Ashby (eds.) *Preferred Argument Structure: Grammar and Architecture for Function* Amsterdam/Philadelphia: John Benjamins, 301-338.
- Argyri, E. and A. Sorace (2007) 'Cross-linguistic influence and language dominance in older bilingual children'. *Bilingualism: Language and Cognition* 10, 79-99.
- Ariel, M. (1994) 'Interpreting anaphoric expressions: a cognitive versus a pragmatic approach'. *Journal of Linguistics* 30, 1-42.
- Arnold, J. and Z. Griffin (2007) 'The effect of additional characters on choice of referring expressions: Everyone counts'. *Journal of Memory and Language* 56, 521-536
- Bel, A. (2003) 'The syntax of subjects in the acquisition of Spanish and Catalan'. *Probus* 15, 1-26.
- Bernardini, P. (2003) 'Child and adult acquisition of word order in the Italian DP' in: N. Müller (ed.) *(In)vulnerable Domains in Multilingualism* Amsterdam/Philadelphia: John Benjamins, 41-81.
- Bloom, P. (1990) 'Subjectless sentences in child language'. *Linguistic Inquiry* 21, 491-504.
- Bock, K., and Griffin, Z. (2000) 'The persistence of structural priming: Transient activation of implicit learning?'. *Journal of Experimental Psychology: General* 129, 177-192.
- Bosch, L. and N. Sebastián-Gallés (2001) 'Early language differentiation in bilingual infants', in: J. Cenoz and F. Genesee (eds.) *Trends in Bilingual Acquisition*. Amsterdam/Philadelphia: John Benjamins, 71-93.
- Branigan, H. (2006) 'Syntactic priming'. *Language and Linguistics Compass* 1, 1-16.
- Branigan, H., Pickering, M. and A. Cleland (2000) 'Syntactic coordination in dialogue'. *Cognition* 75, B13-B25.
- Carminati, M. N. (2002) *The Processing of Italian Subject Pronouns*. UMass, Amherst.
- Chafe, W. (1994) *Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago: The University of Chicago Press.
- Chafe, W. (1996) 'Inferring identifiability and accessibility', in: T. Fretheim and J. Gundel (eds.) *Reference and Referent Accessibility*. Amsterdam/Philadelphia: John Benjamins, 37-46.
- Chang, F., Dell, G., Bock, K. and Z. Griffin (2000) 'Structural priming as implicit learning: A comparison of model sentences'. *Journal of Psycholinguistic Research* 29, 217-230.

- Cipriani, P., Pfanner, P., Chilosi, A. M., Cittadoni, I., Ciuti, A., Maccari, A. et al. (1989) *Protocolli Diagnostici e Terapeutici nello Sviluppo e nella Patologia del Linguaggio* (No. 1/84 Ministry of Health). Pisa: Stella Maris Foundation.
- Clancy, P. (1993) 'Preferred argument structure in Korean acquisition', in: E. Clark (ed.) *The Proceedings of the Twenty-fifth Annual Child Language Research Forum*. Stanford: CSLI, 307-314.
- Clancy, P. (in press) 'Dialogic priming and the acquisition of argument marking in Korean', in: J. Guo, E. Lieven, S. Ervin-Tripp, N. Budwing, S. Ozçaliskan and K. Nakamura (eds.) *Crosslinguistic Approaches to the Psychology of Language: Research in the Tradition of Dan Isaac Slobin*. Hillsdale, NJ: Lawrence Erlbaum.
- De Houwer, A. (1990) *The Acquisition of Two Languages from Birth: A Case Study*. Cambridge: Cambridge University Press.
- Döpke, S. (1998) 'Competing language structures: the acquisition of verb placement by bilingual German-English children'. *Journal of Child Language* 25, 555-584.
- Ferreira, V. and K. Bock (2006) 'The functions of structural priming'. *Language and Cognitive Processes* 21, 1011-1029.
- Genesee, F., Nicoladis, E. and J. Paradis (1995). Language differentiation in early bilingual development. *Journal of Child Language* 22, 611-631.
- Gernsbacher, M. A. and D.J. Hargreaves (1989) 'Accessing Sentence Participants: The Advantage of First Mention'. *Journal of Memory and Language* 27, 699-717.
- Greenfield, P. and J.H. Smith (1976) *The Structure of Communication in Early Language Development*. New York: Academic Press.
- Grimshaw, J. and V. Samek-Lodovici (1998) 'Optimal subjects and subject universals', in: P. Barbosa, D. Fox, P. Hagstrom, M. McGinnis and D. Pesetsky (eds.) *Is the Best Good Enough? Optimality and Competition in Syntax*. Cambridge, MA: The MIT Press, 193-219.
- Grinstead, J. (2000) 'Case, inflection and subject licensing in child Catalan and Spanish'. *Journal of Child Language* 27, 119-156.
- Guasti, M. T. (2002) *Language Acquisition. The Growth of Grammar*. Cambridge: The MIT Press.
- Guerriero, S. (2006) *The Development of Argument Representation: A Cross-linguistic Discourse-Pragmatic Analysis of Study of English and Japanese Child Language*. Montreal: McGill University.
- Hacohen, A. and J. Schaeffer (2007) 'Subject realization in early Hebrew/English bilingual acquisition: The role of cross-linguistic influence'. *Bilingualism: Language and Cognition* 10, 333-344.
- Haegeman, L. (1997) 'Register variation, truncation and subject omission in English and in French'. *English Language and Linguistics* 1, 233-270.
- Hartsuiker, R., Pickering, M. and E. Veltkamp (2004) 'Is syntax separate or shared between languages? Cross-linguistic syntactic priming in Spanish-English bilinguals'. *Psychological Science* 15, 409-414.
- Hudson-D'Zmura, S. and M. Tanenhaus (1998) 'Assigning antecedents to ambiguous pronouns: the role of center of attention as the default assignment', in: M. Walker, A. Joshi and E. Prince (eds.) *Centering in Discourse*. Oxford: Oxford University Press.
- Hughes, M. and S. Allen (2006) 'A Discourse-Pragmatic Analysis of Subject Omission in Child English', in: D. Bamman, T. Magnitskaia and C. Zaller (eds.) *Proceedings of the Thirtieth Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press, 293-304.

- Hulk, A. and N. Müller (2000) 'Bilingual first language acquisition at the interface between syntax and pragmatics'. *Bilingualism: Language and Cognition* 3, 227-244.
- Hyams, N. (1986) *Language Acquisition and the Theory of Parameters*. Dordrecht: Reidel.
- Loebell, H. and K. Bock (2003) 'Structural priming across languages'. *Linguistics* 41, 791-824.
- Lorusso, P., Caprin, C. and M.-T. Guasti (2005) 'Overt subject distribution in early Italian children', in: A. Brugos, M. R. Clark-Cotton and S. Ha (eds.) *A Supplement to the Proceedings of the Twenty-ninth Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press.
- MacWhinney, B. (2000) *The CHILDES Project: Tools for Analyzing Talk* (Third edition). Hillsdale, NJ: Lawrence Erlbaum.
- Meisel, J. M. (1994) *Bilingual First Language Acquisition : French and German Grammatical Development*. Amsterdam/Philadelphia: John Benjamins.
- Müller, N. and A. Hulk (2001) 'Cross-linguistic influence in bilingual language acquisition: Italian and French as recipient languages'. *Bilingualism: Language and Cognition* 4, 1-21.
- Muysken, P. (2002) 'Modeling bilingual speech: Otra vez'. Ms. Radboud University Nijmegen.
- Nicoladis, E. (2006) 'Cross-linguistic transfer in adjective-noun strings by preschool bilingual children'. *Bilingualism: Language and Cognition* 9, 15-32.
- Orsolini, M., Rossi, F. and C. Pontecorvo (1996) 'Re-introduction of referents in Italian children's narratives'. *Journal of Child Language* 23, 465-486.
- Paradis, J. and S. Navarro (2003) 'Subject realization and cross-linguistic interference in the bilingual acquisition of Spanish and English: What is the role of the input?'. *Journal of Child Language* 30, 1-23.
- Pickering, M. and H. Branigan (1998) 'The representation of verbs: Evidence from syntactic persistence in written language production'. *Journal of Memory and Language* 39, 633-651.
- Pinto, M. (2006) 'Subject pronouns in bilinguals: Interference or maturation?', in: V. Torrens and L. Escobar (eds.) *The Acquisition of Syntax in Romance Languages*. Amsterdam/Philadelphia: John Benjamins, 331-350.
- Rizzi, L. (1994) 'Early null subjects and root null subjects', in: T. Hoekstra and B. D. Schwartz (eds.) *Language Acquisition Studies in Generative Grammar*. Amsterdam/Philadelphia: John Benjamins, 151-176.
- Serratrice, L. (2005) 'The role of discourse pragmatics in the acquisition of subjects in Italian'. *Applied Psycholinguistics* 23, 437-462.
- Serratrice, L. (2007a) 'Referential cohesion in the narratives of bilingual English-Italian children and monolingual peers'. *Journal of Pragmatics* 39, 1058-1087.
- Serratrice, L. (2007b) 'Cross-linguistic influence in the interpretation of anaphoric and cataphoric pronouns in English-Italian bilingual children'. *Bilingualism: Language and Cognition* 10, 225-238.
- Serratrice, L. (in press) 'The role of discourse and perceptual cues in the choice of referential expressions in English pre-schoolers, school-age children and adults'. *Language Learning and Development*.
- Serratrice, L. and M. Pinto (in preparation). Priming subject pronouns in Italian children.
- Serratrice, L., Sorace, A., and Paoli, S. (2004) 'Cross-linguistic influence at the syntax-pragmatics interface: subjects and objects in English-Italian bilingual and monolingual acquisition'. *Bilingualism: Language and Cognition* 7, 183-205.

- Skarabela, B. (2006) *The Role of Social Cognition in Early Syntax: The Case of Joint Attention in Argument Realization*. Boston, MA: Boston University.
- Skarabela, B. (2007) 'Signs of early social cognition in children's syntax: The case of joint attention in argument realization in child Inuktitut'. *Lingua* 117, 1837-1857.
- Skarabela, B. and S. Allen (2002) 'The role of joint attention in argument realization in child Inuktitut', in: B. Skarabela, S. Fish and A. H.-J. Do (eds.) *Proceedings of the Twenty-sixth Annual Boston University Conference on Language Development*. Sommerville, MA: Cascadilla Press, 620-630.
- Sorace, A., Serratrice, L., Filiaci, F. and M. Baldo (under review) 'Discourse conditions on subject pronoun realization: testing the linguistic intuitions of older bilingual children'. *Lingua*.
- Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K.M. and J. Sedivy (1995) 'Integration of visual and linguistic information during spoken language comprehension'. *Science* 268, 1632-1634.
- Toribio, A. J. (2004) 'Convergence as an optimization strategy in bilingual speech: Evidence from code-switching'. *Bilingualism: Language and Cognition* 7, 165-173.
- Travis, C. (2007) 'Genre effects on subject expression in Spanish: Priming in narrative and conversation'. *Language Variation and Change* 19, 101-135.
- Tsimpli, I.-M., Sorace, A., Heycock, C. and F. Filiaci (2005) 'First language attrition and syntactic subjects: a study of Greek and Italian near-native speakers of English'. *International Journal of Bilingualism* 8, 257-277.
- Valian, V. (1991) 'Syntactic subjects in the early speech of American and Italian children'. *Cognition* 40, 21-81.
- Valian, V. and Z. Eisenberg (1996) 'The development of syntactic subjects in Portuguese-speaking children'. *Journal of Child Language* 23, 103-128.
- van Kampen, J. (2006) 'Early operators and late topic-drop/pro-drop', in: V. Torrens and L. Escobar (eds.) *The Acquisition of Syntax in Romance Languages* Amsterdam/Philadelphia: John Benjamins, 203-227.

Referring Expressions in Early Italian

A study on the use of lexical objects, pronouns and null objects in Italian pre-school children

Roberta Tedeschi
Utrecht University

1. Introduction

This paper discusses the results of an experimental study investigating the use of lexical objects, object clitics and object omissions in monolingual Italian-speaking children of different ages and adult controls.

Recent studies have tested the hypothesis that the integration of syntax and discourse-pragmatic knowledge affects production in early stages of development (see Serratrice, Sorace and Paoli 2004; Rizzi 2002; Schaeffer 2000; Roeper 1999; among others). This paper focuses in particular on pre-school children's sensitivity to discourse cues and on the influence of discourse-pragmatics on the realization/omission of syntactically obligatory material (specifically the omission of syntactic objects). I discuss different hypotheses about the way syntax and discourse are integrated in early stages of language development. I address the proposal that omission of obligatory syntactic material, typical of child language, results from difficulties in integrating syntactic information within the appropriate discourse framework (Avrutin 1999; Serratrice, Sorace and Paoli 2004; Serratrice 2005; among others), together with the claim that children rely on discourse for the interpretation of missing syntactic material. In this respect, the phenomenon usually known as clitic omission is reconsidered in light of recent findings about argument drop in early Italian (Serratrice, Sorace and Paoli 2004). I take into account previous findings about object clitic omission in Italian child language. In particular, I discuss Schaeffer's (2000) proposal, in which clitic omission is seen as a syntax-pragmatics interface phenomenon. I compare Schaeffer's methodology and results to the findings of the present experiment. I suggest that competition of syntactic and pragmatic requirements in young children results in over-reliance on (linguistic and extra-linguistic) discourse for the encoding of information. Over-reliance on discourse is proposed to be the reason for the phenomenon of clitic/object omission. Children are thus expected to omit when the referents are salient and highly accessible in the discourse. The results of the experiment partially confirm this hypothesis. They show that children are sensitive to discourse cues in their use of referring expressions from very early on.

In the next sub-section, I will present data about clitic omission in early Italian, together with some proposals that have been made to account for this phenomenon. Section 2 introduces some hypotheses about the integration of syntax and discourse-pragmatics in early stages of development. In the third section I will present and discuss the results of the present experiment, followed by the conclusions.

1.1 Clitic omission in early Italian

In this paper I investigate to what extent children are sensitive to discourse factors in their use of referring expressions from very early on. This implies children's ability to choose referring expressions based on the accessibility status of the referents in discourse¹. Among the options that children have in their choice of referring expressions, the choice of a null element is also present. If the "null" option is ungrammatical in a certain language, the dropped element (in our case the object) can be considered as an omission. In Italian, a language that does not allow object *pro*, object clitics are often omitted in early stages of development (Antelmi 1997; Cipriani et al. 1993; Guasti 1993/94; Schaeffer 2000). Examples of omissions are given in (1) and (2) below, from the Calambrone corpus (Cipriani et al. 1989), available on the CHILDES database (MacWhinney 2000). In the examples, "0w" indicates that a pre-verbal (direct) object clitic has been omitted.

(1) penchè penchè 0w ha usato lui.
because because has used he
"because he has used"
(Raffaello, 2;6.13)

(2) poi 0w metto qua
Then put here
"then I put here" (while playing with toys)
(Diana, 1;11)

Object omission in child Italian is optional. When children produce clitics, they place them correctly (Guasti 1993/94; Schaeffer 2000; Cardinaletti and Starke 2000). Acquisition research has mostly focused on structural properties of clitics (Wexler, Gavarró and Torrens 2004; Hamann, Rizzi and Frauenfelder 1996; Jakubovicz et al. 1998; among others). However, recent accounts view the interaction of syntax and discourse-pragmatics as a crucial factor affecting omission (see Guasti 1993/94, Schaeffer 2000 and Serratrice this volume for Italian).

It is commonly assumed that children differ from adults because they optionally produce constructions that are obligatory in adult language. According to Schaeffer (2000), this phenomenon is due to the optional marking of syntactic features, driven by the immaturity of children's pragmatic system. Schaeffer provides a new approach for the phenomenon of clitic omission, relating her findings to children's capacities in other cognitive domains. In particular, she relates children's poor performance in production to the development of the pragmatic knowledge. Schaeffer claims that at early stages, children may ignore the distinction between discourse-related and non-discourse related referentiality. The reason for this would

¹ I adopt the term 'accessibility' from Ariel (1990). According to Accessibility theory, the choice of a referring expression depends on the degree of activation of its antecedent, i. e. on the ease with which its antecedent can be retrieved from memory.

be that at this stage they lack a particular pragmatic rule, the Concept of Non-Shared Knowledge. Children do not always realize that speaker and hearer knowledge are distinct entities. When they fail to do so, they do not make a distinction between discourse-related and non-discourse related referentiality. As a consequence, referentiality is optionally marked in early grammar. Schaeffer predicts that syntactic processes driven by referentiality (such as direct object clitic placement in Italian) will not always take place in early child language. The shift from optionality of syntactic referentiality marking to adult-like grammatical marking in her study is found to take place around the age of 3. Although three-year olds do not show a perfect adult-like behaviour, there is a clear developmental difference when they are compared to the two-year olds (62% vs 22 % overt object clitics).

Schaeffer's approach places the phenomenon of clitic omission at the syntax-discourse interface. When children do not mark referentiality in syntax, they refer directly to an entity in the (model of the) world, similarly to what adults do when knowledge is shared between speaker and hearer. However, as I will discuss later, in some studies children were found to be sensitive to the level of saliency and accessibility of the referents in the discourse when using referring expressions, including null objects in Italian (see Serratrice, Sorace and Paoli 2004). Some discourse-pragmatics abilities seem to be present from very early on. For example, children seem to drop arguments only in contexts that respect some pragmatic requirements. This suggests an alternative explanation of children's omission in early stages of development from a pragmatic point of view. Omission of obligatory material could in fact follow from competition between syntactic and pragmatic requirements, where discourse "wins". Rather than lacking a pragmatic principle, it is possible that children over-rely on one. Two possible explanations for how this phenomenon occurs are presented in the following section.

1. The integration of syntax and discourse in early stages of development

2.1 Features of Informativeness and argument realization

The hypothesis that omission of obligatory syntactic material follows from competition of syntactic and pragmatic requirements has been discussed in work by Serratrice et al. (2004), Serratrice (this volume) and Avrutin (2004, 2006) among others. The claim is that, in early stages of development, children rely on discourse licensing for the interpretation of missing obligatory morpho-syntactic material.

Based on previous studies about the effects of discourse-pragmatics on argument realisation (see Allen 2000 and Skarabela and Allen 2004 a. o.), Serratrice et al. (2004) tested the hypothesis that information in early utterances tends to be linguistically encoded for aspects of an event that are not highly accessible to the hearer. According to Allen (2000), children tend to omit arguments when the referents are maximally clear from the discourse and situational context, but not when the referent of an argument is in doubt. Children are thus claimed to be highly sensitive to the dynamics of information flow in discourse. In conversations, potential uncertainty regarding the referents they are talking about tends to be

reduced. Allen mentions a set of binary features that have been shown to influence argument representation in previous works by Clancy (1993, 1997) and Greenfield and Smith (1976), among others. These *informativeness features* are claimed to establish “how informative the speaker should be in representing in speech the referent at hand” (Allen 2000: 487). Informativeness features are binary. They have a ‘high’ informative value, which makes the identity of the referent less certain, and a ‘low’ informative value, which makes it more certain.

Serratrice et al. (2004) explored the influence of Informativeness features on the patterns of omissions of referential subjects and objects in the spontaneous speech of six monolingual Italian, six monolingual English and one Italian-English bilingual child. The data about the Italian monolingual children were collected using the Calabrone corpus in the CHILDES database. The features taken into consideration were Absence (referent absent from physical context), Contrast (contrast emphasized between potential referents), Differentiation (two or more potential referents in the preceding discourse), Query (referent subject of or answer to query) and Activation (referent associated with a completely new referent or with topic shift). According to their results, object omission follows a predictable pattern, where null objects are almost exclusively associated with uninformative features. They claim that there is “competition... between the syntactic requirement of an overt object and the pragmatic principle of Informativeness that allows null arguments when their informative status is low” (p. 200). They conclude that “a discourse pragmatics approach is necessary for a comprehensive understanding of the phenomenon of argument omission in child language” (p. 199). Their results show that children drop objects in uninformative contexts. This finding confirms the hypothesis that children rely on discourse when they omit objects. Reliance on the linguistic or extra-linguistic discourse allows the recovery of missing information.

Based on Serratrice et al.’s results, it is possible to establish a link between clitic omission and the pragmatically uninformative contexts in which clitics are used in adult language. In fact, from a discourse point of view, clitic pronouns have highly accessible antecedents (Ariel 1990). Italian object clitics can only refer to a prominent antecedent in the discourse (Cardinaletti and Starke 1999, 2000), and they are associated with [-focus] interpretation (Serratrice, Sorace and Paoli 2004). Thus, the use of Italian clitic pronouns is associated with referent maintenance. The fact that clitic pronouns refer to a prominent antecedent favours the hypothesis that when a clitic is omitted, discourse can in principle provide the required information. Moreover, since conversations between young children and adults tend to focus on referents that are physically present and perceptually available to both interlocutors (Van Kampen 2004, Van Kampen and Pinto this volume; Serratrice in press), the referent of a clitic is often available in the extra-linguistic context. The examples in (1) and (2) are clear cases of omissions in a context where the referent of the clitic is present in the situation. Based on Serratrice et al.’s findings, I propose an investigation of object drop in different pragmatic contexts. In section 4, I will present an elicited production task aimed at investigating how pre-school children encode the distribution of lexical object, object clitics and object omissions in different discourse conditions.

2.2 Weak syntax

According to Avrutin (2004, 2006), the optional omission of functional elements (telegraphic speech) can be accounted for by the fact that the missing information is encoded through the ‘context’. Information is encoded through a “non-syntactic channel,” instead of being encoded through syntax. Avrutin’s approach is modular in nature, in the sense that syntactic computations are encapsulated with respect to meaning. Avrutin calls this narrow syntax. Narrow syntax is in charge of combining lexical items in an order allowed by a given language through symbolic operations, and the output of such operation must be interpretable. Avrutin (2006) makes a distinction between linguistic discourse (also known as Conceptual-Intentional interface, or Information structure) and the context. Linguistic discourse only includes purely linguistic operations. It is the level where topic, focus, specificity and pronominal anaphora are encoded. The context is a non-linguistic system of thought that can be modified by both linguistic and non-linguistic means. Narrow syntax and discourse communicate through the syntax-discourse interface.

Avrutin proposes a discourse model to explain how the units of the narrow syntax are translated into discourse units. Linguistic discourse operates on units consisting of a *frame*, introduced by functional projections from the narrow syntax, and a *heading*, introduced by lexical projections from the narrow syntax. Frames separate the information units from each other, and headings provide the information necessary for interpretation. In general, the units of linguistic discourse must contain both a frame and a heading in order to be interpretable. However, according to Avrutin, certain contextual conditions can take over the function of functional categories, specifically to introduce a frame. The author underlines the fact that telegraphic speech is typical of two populations, children and aphasics, characterized by reduced processing capacities. A connection is hypothesized between the phenomenon of telegraphic speech in the two populations and their lack of processing resources. This hypothesis is not based on the idea of a pragmatic deficit, as a lack of pragmatic knowledge could hardly explain the data of adult aphasics. The poor performance of these two populations is thus explained as a processing deficit. Avrutin claims that in populations with lower than normal brain activation, such as children (Kolk 2001), syntax is weakened (syntactic operations consume more resources). By contrast, reliance on the context is a less or equally expensive option for encoding information. The occasional use of the syntactic or non-syntactic “channel” explains the optionality of omissions.

Despite some noticeable differences, both Serratrice et al.’s and Avrutin’s proposal predict that children should omit obligatory syntactic material in cases in which information is recoverable from the linguistic and extra-linguistic discourse (the ‘context’, according to Avrutin). In the case of object drop/clitic omission, both accounts predict omissions should be associated with a high level of salience and accessibility of the referents in the discourse. In the following sections I suggest a way to test these predictions in an elicited production task, based on studies

investigating the distribution of lexical, pronominal and null subjects in the production of English pre-school children.

2. Experiment: Referring expressions in early Italian

This section contains the results of an experiment investigating pre-school children's sensitivity to discourse cues in their use of referring expressions. More specifically, the discourse cue under consideration was the type of question asked by the experimenter. The language investigated is Italian.

A number of recent studies have investigated the relative contribution of perceptual and discourse cues to children's choice of subject referring expressions in pre-school and older children (see Matthews et al. 2006; Serratrice submitted, a. o.), showing that some sensitivity to the discourse appropriateness of referential expressions begins to emerge around the age of two. In the above mentioned studies one discourse manipulation consisted in asking (i) *general questions* like "What happened?", and (ii) *specific questions* like "What did the clown do?". In response to the discourse manipulation, English pre-school children produced significantly more nouns in response to general questions than in response to specific ones. On the other side, specific questions were associated with the use of subject pronouns or with null subjects. These findings suggest that also for objects there should be a correlation between the type of question asked and the choice of a referring expression (lexical object, object pronoun, object omission).

In the present experiment, children of different ages and adult controls answered general questions (such as *what happens?*) and specific questions (such as *what is X doing to Y?*) about a set of pictures. Responses were coded for the use of object referring expressions, namely lexical objects, object pronouns and object omissions. A different use of lexical objects and object pronouns was expected depending on the type of question asked by the experimenter, i. e. depending on whether the referents had been mentioned (specific questions) or not (general questions) in the immediately preceding discourse. When the referents were not introduced in the preceding discourse, subjects were expected to produce more lexical objects than pronouns. By contrast, when the antecedent had been mentioned in the immediately preceding discourse, a less informative expression (usually a clitic pronoun) was expected. An investigation of the contexts of object omission was instantiated in order to address the question whether object drop (an ungrammatical option in adult Italian) can be accounted for as a syntax-discourse interface phenomenon. The contexts in which children dropped objects were thus expected to be syntactically ungrammatical but pragmatically acceptable.

3.1 Participants

Eleven children (aged 2;6 – 6;5) and six adults (age 27 – 31) took part in the experiment. Children were divided into groups depending on age, as illustrated in table 1 below.

Group 1 (N = 3)	Age: 2;6 – 2;11 (m. a. 2;9)
Group 2 (N = 5)	Age: 3;6 – 4;1 (m. a. 3;9)
Group 3 (N = 3)	Age: 5;5 – 6;5 (m. a. 5;11)
Group 4 (N = 6)	Age: 27 - 31

Table 1: Participants divided according to age

Originally, five more children participated in the experiments, but they were excluded because they did not understand the task or because their answers were unintelligible. All the children who were excluded belonged to the group of two-year olds, suggesting that the task is particularly demanding for younger children. Children were tested in one or more sessions of 10 - 15 minutes at the day-care centre “San Giuseppe” in Corporeno, Italy. Adults were tested at their homes or at the experimenter’s home.

3.2 Procedure

A few days before starting the test, the experimenter spent some time at the day-care centre in order to familiarize with the children. A hand puppet called “Lumachina” was introduced to the children. The toy was presented as a shy, silly animal that wanted to talk with children, but not with adults. During the experimental sessions, children were presented with pictures and they were asked to describe them to the puppet by answering a generic question (such as “What happens in this picture?”) or a specific question (such as “What is X doing to Y?”) about each picture. Only positive reinforcement was given during the task. Data were recorded on a CF portable recorder. They were transcribed and coded by the experimenter. For details about the coding of data, see section 3.4 below.

3.3 Materials and methods

Each child was presented with 48 pictures. 32 of these were experimental items (16 required answering a generic question, the remaining 16 required answering a specific question) and 16 were fillers. The pictures depicted a man or a woman performing an action on one or two other characters (the agent and the patient/s always differed in gender). Each picture depicted new characters. The actions performed were ‘pettinare’ (to comb), ‘lavare’ (to wash), ‘salutare’ (to greet) and ‘toccare’ (to touch). Pictures were presented in a randomized order (the same order for all participants). All direct object clitics (singular - plural, masculine - feminine) were targeted.

Figure 1 gives an example of the pictures used for the test. In this particular picture, a man is combing a girl.

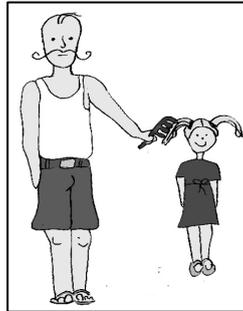


Fig. 1: Example of the pictures used for the experiment

Examples of general and specific questions are given below, together with their expected answers, in (3) – (4).

(3) Answering a general question

Question: Cosa succede in questo disegno?
“What happens in this picture?”
Answer: Un/il papà pettina una/la bimba
“A/the dad combs a/the child”

(4) Answering a specific question

Question: Cosa fa il papà alla bimba?
“What is the dad doing to the child?”
Answer: la pettina
“(He) cl-her combs”

To summarize, the following results were expected: 1) in response to generic questions, in which there was no mention of the referents in the preceding discourse, an “informative” referential expression (a full DP) was expected; 2) in response to specific questions, in which the referents were introduced in the immediately preceding discourse, a “less informative” referential expression was expected (an object clitic).

Moreover, since we know from previous literature that Italian children omit objects/clitics in early stages of development, the following results were expected for dropped objects: 1) low rates of omissions in response to generic questions; 2) higher rates of omissions in response to specific questions than in response to generic questions. These predictions reflect Serratrice et al.’s claim that arguments are dropped when the informative status of the referents is low (when they are salient and highly accessible).

3.4 Coding the data

The data were calculated on the base of the number of transitive and ditransitive verbs produced by each subject. Therefore, answers with intransitive verbs or with no verb were excluded. In total, 10% (55/544) of the answers were missing or discarded as irrelevant. The percentage of discarded answers was higher in the groups of two- and three-year olds (21% and 16% respectively). As mentioned above, sometimes the participants answered with ditransitive verbs. This especially occurred with pictures depicting the actions of touching and kissing (to touch someone's shoulder, to give a kiss to someone). Answers with verbs other than the expected ones were included if they were relevant for the description of the pictures. "Out of the blue" answers, i. e. cases when children described a picture without waiting for the question, were included in the "general question" condition since the referents were not introduced in the immediately preceding discourse ("out of the blue" answers amounted to 1%). Sentences with topicalized objects and clitic left dislocations were counted as lexical objects (such cases amounted to 2% of the answers). In general, the first answer given was counted, except when the child corrected herself. In this case, the correct answer was counted. Unclear cases were discussed with a linguist, a native speaker of Italian.

3.5 Results and discussion

The number of lexical objects (Full DPs), direct/indirect object clitics and null objects was counted. The results are presented in tables (2) – (5) below.

Gen. quest. Full DP	Gen. quest. Obj. pron.	Gen. quest. Omission	Spec. ques. Full DP	Spec. ques. Obj. pron.	Spec. ques. Omission
92/92	0/92	0/92	7/91	78/91	6/91

Table 2: Adult use of referring expressions in response to general and specific questions

As shown in Table 2 above, adults used more full DPs in response to generic questions than in response to specific questions. Moreover, they mostly used pronouns (object clitics) when answering specific questions. The different distribution of lexical objects and clitic pronouns in response to general and specific questions reached statistical significance ($\chi^2 = 147.239$, $df = 1$, $p < 0.001$). A low number of omissions (6.6%) were found in the adults' answers to specific questions.

Table (3) below illustrates the choice of referring expressions in the group of two-year olds.

Gen. quest. Full DP	Gen. quest. Obj. pron.	Gen. quest. Omission	Spec. ques. Full DP	Spec. ques. Obj. pron.	Spec. ques. Omission
14/38	16/38	8/38	1/38	29/38	8/38

Table 3: Use of referring expressions in response to general and specific questions in the group of two-year olds

The results indicated that the group of two-year olds showed sensitivity to discourse requirements in their choice of referring expressions in response to generic and specific questions. More lexical objects were used in response to generic questions than in response to specific ones, and more pronouns were used in response to specific questions than in response to generic ones. The difference reached statistical significance ($\chi^2 = 14.106$, $df = 1$, $p < 0.001$). Children dropped objects to the same extent in both conditions (21%). A partially different result was found for the group of three-year olds.

Gen. quest. Full DP	Gen. quest. Obj. pron.	Gen. quest. Omission	Spec. ques. Full DP	Spec. ques. Obj. pron.	Spec. ques. Omission
34/72	30/72	8/72	12/63	29/63	22/63

Table 4: Use of referring expressions in response to general and specific questions in the group of three-year olds

As shown in Table 4 above, the distribution of lexical objects and pronouns varied significantly depending on the type of question asked by the experimenter ($\chi^2 = 4.849$, $df = 1$, $p = 0.028$). However, although children used more lexical objects in response to generic questions than in response to specific questions, almost the same amount of pronouns was used to answer generic and specific questions. With respect to the proportion of omissions, a noticeable difference was found. Children omitted 11% objects when answering generic questions and 35% objects in their answers to specific questions. However, the difference did not reach statistical significance (Wilcoxon Signed Ranks $z = -1.826$, $p = 0.068$) The different use of pronouns and null objects in the group of three-year olds with respect to the group of two-year olds might reflect a different strategy in answering questions in the two groups. While two-year olds tend to encode specific information by using clitic pronouns, three-year olds tend to omit more objects when answering a specific question.

A developmental trend is noticeable when comparing the two-and three-year olds with the five-year olds. The results for the group of five-year olds are given in Table 5 below.

Gen. quest. Full DP	Gen. quest. Obj. pron.	Gen. quest. Omission	Spec. ques. Full DP	Spec. ques. Obj. pron.	Spec. ques. Omission
47/47	0/47	0/47	8/48	37/48	3/48

Table 5: Use of referring expressions in response to general and specific questions in the group of five-year olds

The results show that children in the group of five-year olds showed sensitivity to discourse in their use of referring expressions. They used more lexical objects in response to generic questions than in response to specific ones. Moreover, they did not use pronouns in response to generic questions, but only in their answers to specific questions. In this respect, their behavior is adult-like. Like the adult controls, and contrary to two- and three-year olds, they used pronouns only when the referents had been introduced in the immediately preceding discourse.

The results presented above indicate that all groups were sensitive to discourse requirements in their choice of referring expressions. More lexical objects were used in response to generic questions than in response to specific questions. More pronouns/null objects were used in response to specific questions than in response to generic questions. Predictions were thus born out. A developmental trend was found in the three groups of children: while two- and three year olds used a high proportion of pronouns in response to generic questions, five-year olds did not use a single pronoun in when answering questions in which the referents had not been introduced. The use of pronouns in response to generic questions, a pragmatically inappropriate option in adult language, might have been triggered by the fact that the referents were visually introduced in the physical context.² The influence of the extra-

² The use of clitic pronouns without a linguistic antecedent has already been documented for early French by Belzil, Pirvulescu and Roberge (2007) and by Van Kampen (2004), among others, and for early French and Italian by van Kampen & Pinto (this volume). Based on previous work by Cornish (1999), Belzil et al. define this phenomenon as exophoric use of clitics (contextual anaphora). They observe that in children's early production endophoric (i. e. with a linguistic antecedent) and exophoric anaphora coexist, although the use of endophoric anaphora increases with age. By contrast, Van Kampen and Pinto assume that all clitics produced in early stages of development are not instantiations of discourse anaphora but situation bound pronouns, as their use is always associated with the presence of a gesture-sustainable antecedent. In the beginning, children only have a situation bound device for encoding topic shift: each utterance is considered as a separate entity associated with its own topic. Anaphoric clitics appear in a later step, when the child has acquired the ability of reference-tracking in discourse. Clearly, the identification of exophoric and endophoric (adopting Belzil et al.'s terminology) uses of clitics in children's utterances is a complicated issue. Even in the presence of linguistic antecedents, the referents are very often physically present and perceptually available in the situation. However, an endophoric use of clitics in the present experiment cannot be excluded a priori. If clitic use was only situation-bound, one would expect a random choice of lexical objects and clitics in the two conditions. On the contrary, the fact that two-year olds used more clitic pronouns in response to specific questions than in response to generic questions indicates that children preferred the use of a pronoun when the referent had already been introduced in the linguistic discourse rather than

linguistic context on argument realization was not directly tested in the present experiment. It is anyway arguable that the physical presence of the referents in the context can be a sufficient requirement for younger children to assume that the referents are accessible to the hearer. In fact, in the experiment the referents were in the simultaneous physical focus of attention of both speaker and hearer. In such a context, joint attention between speaker and hearer is commonly assumed (Serratrice 2005). As illustrated in section 2.1, above, ‘absence’ can be considered as feature of ‘Informativeness.’ This feature has been found to be a good predictor of argument realization. When the referents are present in the physical contexts, objects are dropped more than when the referents are absent. It is thus possible that the presence of the referents in the pictures had a strong influence on children’s responses, making the referents salient and accessible enough to use a less informative referring expression where adults would have used a more informative one. Following this line of reasoning, the results suggest that both linguistic discourse and extra-linguistic discourse (the ‘context’ in Avrutin’s terms) can influence children’s use of referring expressions in early stages of development. Differently from two- and three-year olds, older children and adults instead rely on linguistic discourse rather than on the extra-linguistic context.

With respect to object omission, predictions were partly born out. With exception of the two-year olds, all groups omitted more objects in response to specific questions than in response to generic ones. In particular, five-year olds and adult controls never omitted an object when answering generic questions. In both groups, omissions only occurred in response to specific questions, and at very low rates (around 6%). The group of three-year olds omitted in response to both types of questions, but children omitted more objects when answering specific questions. However, the difference was not statistically significant. The group of two-year olds did not show any sensitivity to the discourse cue under consideration in their rates of omissions. Children in this group omitted to the same extent in both conditions. However, two-year-olds showed sensitivity to the discourse cue under consideration by using more clitic pronouns in response to specific questions than in response to generic ones. Surprisingly, the children in the youngest group made large use of clitic pronouns, and they omitted at lower rates than expected on the base of previous studies (21% of omissions in the present experiment vs. 64% in Schaeffer 2000). A possible explanation for this finding is presented in the next section.

3.6 Further considerations about object omission

The remainder of this chapter will focus on the data concerning object omission. The results are compared with previous findings about clitic omission in early Italian. First of all, the data show a lower percentage of omissions than found in Schaeffer’s results in the group of two-year olds. One reason for the discrepancy between

when the referent was new in discourse. Hence, at least to some extent, clitics were instantiations of discourse anaphora. The effects of “presence vs. absence” of the referents in the physical context will hopefully be the object of further research.

Schaeffer's findings and the results of the present experiment could be the difference in the age of the tested groups. The mean age is respectively 2;5 and 2;9 for the two-year olds of Schaeffer's study and the present one. Moreover, the age range differs (2;1 – 2;11 vs. 2;6 – 2;11). The phenomenon of object/clitic omission seems to decrease dramatically with age, and 'late' two-year olds appear to be already closer to adult performance. The hypothesis that a difference of a few months can make the difference is confirmed also by longitudinal studies on the acquisition of Italian morpho-syntax (see Cipriani et al. 1993; Antelmi 1997; Guasti 1993/94). In these studies, the omission rates decrease noticeably during the second year of life.

A second difference between Schaeffer's study and the present one concerns the methodology used. In the present study we made use of pictures depicting transitive actions involving two characters, an agent and a patient. The participants were asked to answer a question about each picture. Schaeffer's experiment was more elaborate. It consisted in a mixed truth - value judgment/elicited production task. Children interacted with Raja, a silly puppet that makes mistakes. Children were encouraged to correct her whenever she said something wrong.

An example of Schaeffer's scenarios (English translation) is given in (5) below.

- (5) - *Experimenter*: "Look, here we have a rabbit, a puppet with pink hair, and a comb. Look, the puppet has long hair, and it's a bit of a mess. Therefore, the rabbit is combing the puppet".
- *Raja*: "The rabbit is washing the puppet!"
- *Child*: "No!"
- *Raja*: "Why? Isn't the rabbit washing the puppet?"
- *Experimenter*: So, say it to Raja: What is the rabbit doing to the puppet?
- *Child*: (He)'s combing her (clitic her-ACC combs)

Schaeffer also reports some examples of children's answers (English translation):

- *Raja*: The rabbit is washing the puppet!
- *Child*: No, (she) is combing!

To explain the different results in the two elicited production studies, I would like to suggest that 1) in a specific question of the type "what is X doing to Y", the referents are introduced in the discourse and the agent and the patient roles are already established. The focus is on the action and not on the characters involved; 2) in Schaeffer's scenario the child is required to correct the puppet. Such an experimental setting implies a contrast between the verb used by the puppet and the one used by the child. I suggest that the use of specific questions, and even more the presence of a contrast, triggered a number of elliptical responses in Schaeffer's experiment.

To summarize, object omission in Italian seems to be more restricted than previously claimed, in both time and quantity.

3. Conclusion

In this paper, I presented the results of a study aimed at investigating the use of different referring expressions in early Italian. In particular, I tested the influence of the discourse cue “general vs. specific question” on children’s use of lexical objects, object clitics and object omissions. Predictions were based on the assumption that the integration of syntactic and discourse requirements in early stages of development would affect children’s production. The results of the experiment favoured the hypothesis that children are sensitive to both linguistic discourse and the context in early stages of development, as indicated by the fact that even for the youngest children the use of referring expressions differed depending on the type of question asked by the experimenter. All subjects used more pronouns/omissions when the referents had been introduced in the linguistic discourse than when the referents were new. By contrast, they used more lexical objects when the referents were new in the linguistic discourse than when they had been previously introduced. However, with exception of the five-year olds, children often used clitic pronouns to refer to an antecedent that had not been mentioned in the immediately preceding discourse. In this respect, the influence of extra-linguistic discourse (the ‘context’) on children’s performance was addressed as a possible cause of the non adult-like behaviour of two- and three-year olds. The different distribution of omissions found in response to generic and specific questions is in line with theories that view syntactic and discourse requirement as competing factors in early grammar, claiming that children rely on discourse for the interpretation of missing syntactic material.

The results of the present experiment were compared to data from previous studies on the phenomenon of clitic omission in early Italian. In particular, I discussed Schaeffer’s proposal that omission follows from the lack of a pragmatic principle. I observed that the findings of the present experiment could not be predicted by Schaeffer’s approach. In particular, the fact three-year olds and older children omitted more objects in response to specific questions than in response to general questions, and the fact that children used different referential expressions depending on the discourse status of the referents (including two-year olds). Whether this phenomenon can be accounted for by hypothesizing an overuse of pragmatic abilities, as suggested by Serratrice et al. (2004), or if processing capacities also come into play (as proposed by Avrutin 2006), remains an open question. Moreover, the question arises whether the phenomenon should be addressed as clitic omission, or if, following Serratrice et al.’s, the focus should be shifted to argument realization and argument drop. Further data are needed, in order to know more about the ways and the extent to which two- and three year olds show sensitivity to different discourse factors. Moreover, further research should be aimed at defining the status of missing objects from a grammatical point of view.

References

- Allen, S. (2000) 'A discourse-pragmatic explanation for argument representation in child Inuktitut'. *Linguistics* 38, 483-521.
- Antelmi, D. (1997) *La Prima Grammatica dell'Italiano. Indagine longitudinale sull'acquisizione della morfosintassi italiana*. Bologna: Il Mulino.
- Ariel, M. (1990) *Accessing noun-phrase antecedents*. London: Croom Helm.
- Avrutin, S. (1999) *Development of the Syntax-Discourse Interface*. Dordrecht: Kluwer Academic Publishers.
- Avrutin, S. (2004) 'Optionality in child and aphasic speech'. *Lingue e Linguaggio* 1, 67-89.
- Avrutin, S. (2006) 'Weak syntax', in: Y. Grodzinsky and L. Amunds (eds.) *Broca's Region*. Oxford: Oxford Press.
- Belzil, I., M. Pirvulescu and Y. Roberge (2007) 'Object clitic production and omission in early French and the role of the extralinguistic context', Paper presented at the Generative Approaches to Language Acquisition (GALA), Barcelona, September 7.
- Cardinaletti, A, and M. Starke (1999) 'The typology of structural deficiency: A case study of the three classes of pronouns, in: H. van Riemsdijk (ed.) *Clitics in the Languages of Europe*. Berlin/New York: Mouton de Gruyter, 145-233.
- Cardinaletti, A, and M. Starke (2000) 'Overview: The grammar (and acquisition) of clitics', in: S. M. Powers and C. Hamann (eds.) *The Acquisition of Scrambling and Cliticization*. Dordrecht: Kluwer Academic Publishers, 165-186.
- Cipriani, P., A. Chilosi, P. Bottari and L. Pfanner (1993) *L'Acquisizione Della Morfosintassi: Fasi e Processi*. Padova: Unipress.
- Cipriani, P., P. Pfanner, A.M. Chilosi, L. Cittadoni, A. Ciuti, A. Maccari, N. Pantano, L. Pfanner, P. Poli, S. Sarno, P. Bottari, G. Cappelli, C. Colombo and E. Veneziano (1989) *Protocolli Diagnostici e Terapeutici nello Sviluppo e nella Patologia del Linguaggio* (1/84 Ministry of Health). Pisa: Stella Maris Foundation.
- Clancy, P. (1993) 'Preferred argument structure in Korean acquisition', in: E. Clark (ed.) *Proceedings of the Twenty-fifth Annual Child Language Research Forum*. Stanford: CSLI, 307-314.
- Clancy, P. (1997) 'Discourse motivations for referential choice in Korean acquisition', in: H. Sohn and J. Haig (eds.) *Japanese/Korean Linguistics* 6. Stanford, CA: CSLI Publications, 639-659.
- Cornish, F. (1999) *Anaphora, Discourse, and Understanding. Evidence from English and French*. Oxford: Clarendon Press (Oxford University Press).
- Greenfield, P. and J.H. Smith (1976) *The Structure of Communication in Early Language Development*. New York: Academic Press.
- Guasti, M.T. (1993/94) 'Verb syntax in Italian child grammar: finite and nonfinite verbs'. *Language acquisition* 3, 1-40.
- Hamann, C., L. Rizzi and U.H. Frauenfelder (1996) 'On the acquisition of subject and object clitics in French', in: H. Clahsen (ed.) *Generative perspectives on Language Acquisition*. Amsterdam/Philadelphia: John Benjamins, 309-334.
- Jakubowicz C., L. Nash, C. Rigaut and C. Gerard (1998) 'Determiners and clitic pronouns in French-speaking children with SLI'. *Language Acquisition* 7, 113-160.
- Kampen, J. van (2004) 'Learnability order in the French pronominal system', in: R. Bok-Bennema, B. Hollebrandse, B. Kampers-Manhe and P. Sleeman (eds.) *Selected Papers from Going Romance 2002*. Amsterdam/Philadelphia: John Benjamins, 163-183.

- Kampen, J. van and M. Pinto (this volume) 'Germanic and Romance discourse devices in acquisition'.
- Kolk, H (2001) 'Does agrammatic speech constitute a regression to child language? A three-way comparison between agrammatic, child, and normal ellipsis'. *Brain and language* 77, 340-350
- MacWhinney, B. (2000) *The CHILDES project: Tools for analyzing talk* (3 ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Matthews, D., Lieven, E., Theakston, A. and M. Tomasello (2006) 'The effect of perceptual availability and prior discourse on young children's use of referring expressions'. *Applied Psycholinguistics* 27, 403-422.
- Rizzi, L. (2002) 'On the grammatical basis of language development', ms. University of Siena.
- Roeper, t. (1999) 'Universal bilingualism'. *Bilingualism: Language and Cognition*, 169-186.
- Schaeffer, J. (2000) *The acquisition of direct object scrambling and clitic placement*. Amsterdam/Philadelphia: John Benjamins.
- Serratrice (2005) 'The role of discourse pragmatics in the acquisition of subjects in Italian'. *Applied Psycholinguistics* 26, 437-462.
- Serratrice, L. (in press) 'The role of discourse and perceptual cues in the choice of referential expressions in English pre-schoolers, school-age children and adults'. *Language Learning and Development*.
- Serratrice, L. (this volume) 'Null and overt subjects at the syntax-discourse interface: Evidence from monolingual and bilingual acquisition'.
- Serratrice, L., A. Sorace and S. Paoli (2004) 'Crosslinguistic influence at the syntax-pragmatics interface: Subjects and objects in English-Italian bilingual and monolingual acquisition'. *Bilingualism and Cognition* 7, 183-205.
- Skarabela, B. and S.E.M. Allen (2004) 'The role of joint attention and argument realization in child Inuktitut', in: B. Skarabela, S. Fish, A. H. J-Do (eds.) *BUCLD Proceedings*. Somerville, MA: Cascadilla Press, 620-630.
- Wexler, K., A. Gavarró and V. Torrens (2004) 'Feature checking and object clitic omission in child Catalan and Spanish', in: R. Bok-Bennema, B. Hollebrandse, B. Kampers-Manhe and P. Sleeman (eds.) *Romance Languages and Linguistic Theory 2002*. Amsterdam/Philadelphia: John Benjamins, 253-268.