The development of regular and irregular verb inflection in Spanish child language*

HARALD CLAHSEN, FRAIBET AVELEDO
AND IGGY ROCA

Department of Linguistics, University of Essex

ABSTRACT

We present morphological analyses of verb inflections produced by 15 Spanish-speaking children (age range: 1;7 to 4;7) taken from longitudinal and cross-sectional samples of spontaneous speech and narratives. Our main observation is the existence of a dissociation between regular and irregular processes in the distribution of errors: regular suffixes and unmarked (non-alternating) stems are over-extended to irregulars in children's inflection errors, but not *vice versa*. We also found that overregularization errors at all ages are only a small minority of the children's irregular verbs, that the period of overregularization is preceded by a stage without errors, and that the onset of overregularizations is connected to the emergence of obligatory finiteness markings. These findings are explained in terms of the dual-mechanism model of inflection.

INTRODUCTION

The question of how inflected word forms are mentally represented, processed and acquired has been the subject of a debate in which – broadly speaking – two competing classes of theories can be distinguished. On one side is the so-called dual mechanism account according to which regular and irregular inflection are dissociated in children's grammars in basically the

same way as is claimed for the adult grammar, involving two distinct representational systems, a set of lexical entries that are (associatively) listed in memory, and a set of symbolic operations or rules to form larger linguistic expressions (see Clahsen, 1999; Pinker, 1999, for review). For example, past-tense overregularization errors such as *bring-ed* are due to the child applying a regular (-*ed*) affixation rule in cases in which the lexical entry for the irregular word form *brought* is not available, and they disappear once the child can reliably retrieve the correct irregular word form. An alternative theoretical viewpoint is represented by (different kinds of) single-mechanism models which hold that regular and irregular word forms employ the same representational and processing mechanisms and that generalizations in children's use of inflected word forms follow from the formation of patterns between existing word forms (Bybee, 1995; Elman, Bates, Johnson, Karmiloff-Smith, Parisi & Plunkett, 1996; Langacker, 2000). According to these accounts, children overregularize -*ed*, not because -*ed* is rule-based, but because children have heard it used with so many different English verbs. From the perspective of connectionist single-mechanism models (see e.g. Plunkett & Marchman, 1996), for example, the effect of the high frequency of -*ed* forms is that they are reinforced more and yield a strong pattern with a relatively high level of resting activation compared to irregulars, making them more accessible than irregulars and hence more likely to appear in overregularizations. Several studies have investigated child language acquisition data from the perspective of the two competing models but still, the debate is far from settled (Clahsen, Rothweiler, Woest & Marcus, 1992; Marcus, Pinker, Ullman, Hollander, Rosen & Xu, 1992; Clahsen & Rothweiler, 1993; Plunkett & Marchman, 1996; Marchman, Plunkett & Goodman, 1997; Orsolini, Fanari & Bowles, 1998; Ragnarsdóttir, Simonsen & Plunkett, 1999; Matcovich, 2000; Say & Clahsen, 2001, among others).

Despite the considerable number of studies, in particular on English and German, there are many reasons to conduct research on single versus dual-mechanism accounts of the acquisition of inflection over a wider range of languages. In two studies on Italian child language, for example, Orsolini *et al*. (1998) and Matcovich (2000) claimed that the dual-mechanism model of inflection does not generalize to Romance languages such as Italian. Orsolini *et al*. found that in Italian children between the ages of 4;4 and 10;5, most errors were overapplications of 2nd conjugation patterns, while 1st conjugation forms were rarely overapplied. They assume an interpretation of the dual-mechanism model for Italian whereby the 2nd conjugation, being non-productive, has inflected verb forms which are all listed, while 1st and 3rd conjugations, being productive, have rule-based inflected forms. Thus, it looks as if non-productive (listed) forms rather than rule-based (default) forms are overapplied by Italian children. This, they claim, shows that the dual-mechanism model does not hold for Italian; see Say & Clahsen (2001)

592

for a critique of Orsolini *et al.*'s account.[1] Matcovich (2000) found frequency effects in Italian children's past participle formations of existing verbs and similarity effects in adults' past participle productions of nonce verbs; both effects were found for regular 1st conjugation verbs (as well as for irregular 2nd conjugation verbs), and this is claimed to provide evidence against a dual-mechanism account of Italian verb morphology.[2]

In the present study, our main aim is to add to previous findings by investigating data from Spanish child language with respect to the question of whether the dual-mechanism model generalizes to Romance languages. It will be argued that by distinguishing between stem formation and the domain of inflection proper, the dual-mechanism model is able to account for the development of the subtleties of Spanish verb formation and its development.

Verb inflection in Spanish

In Spanish verbs, inflectional affixes are typically combined with stems, which in turn are combinations of theme vowels (TVs) and roots. This is illustrated in (1) for a finite form of the verb *cantar* 'to sing'.

(1) *cant*   *á*   *ba*   *mos*
     'sing' TV   past imperfect indicative 1st person plural

Non-finite forms also include a TV, followed by a specific marker for either the infinitive (*-r*), the gerund (*-nd-*) or the past participle (*-d-*), in turn followed by the so-called desinence in gerunds and past participles, and a

---

[1] The problem in Orsolini *et al.*'s treatment of Italian inflection is that it confuses the properties of conjugation membership with those of inflectional processes, even though in neither case is there a direct mapping between the two sets of properties in Italian. It is true that only the 1st conjugation shows unrestricted productivity in Italian and can accommodate any type of root, but verbs belonging to 2nd or 3rd conjugation classes may still be subject to rule-based inflectional processes in the same way as 1st conjugation verbs. Thus, contrary to Orsolini *et al.*'s claims, an inflectional error in which a regular 2nd conjugation past-tense form is produced instead of the required irregular one (e.g. **cadè* instead of *cadde* from *cadere* 'to fall') is a straightforward overregularization and does not represent an overextension of a non-productive form. Say & Clahsen (2001) reanalysed Orsolini *et al.*'s child language data and found that almost all of the errors were overgeneralizations of regular stems and regular affixation to irregular verb forms, indicating that their data are in fact compatible with the dual-mechanism model.

[2] Note, however, that the effects reported in Matcovich's study could be due to the verb roots used in the experiments, i.e. to the *root* frequencies of existing verbs in the child experiment and the similarity of a nonce verb to an existing verb *root*. Verb roots of all conjugations are said to be stored in the permanent lexicon in both dual and single-mechanism models. Hence low-frequency verb roots are more difficult to retrieve than high-frequency ones, irrespective of the inflectional suffixes that need to be added, and nonce verbs that rhyme with existing verb roots of a particular class are likely to be assigned to that class, irrespective of whether the inflected verb form to be produced is regular or irregular. Thus, Matcovich's findings can be explained in terms of stored verb roots and do not directly bear on the nature of (regular vs. irregular) inflectional processes.

593

plural suffix (preceded by a desinence sometimes analysed as an epenthetic vowel) in plural forms of nominalized infinitives:

(2) a. *cant* *a* *r* = 'to sing'
         'sing' TV infinitive

     b. *cant* *a* *nd* *o* = 'singing'
         'sing' TV gerund desinence

     c. *cant* *a* *d* *o* = 'sung'
         'sing' TV participle desinence

     d. *cant* *a* *r* *e* *s* = 'songs' (N)
         'sing' TV infinitive desinence plural

The root together with the TV is traditionally referred to as the stem. Most inflected forms are stem-based (e.g. *cant-a-steis* 'You-plural sang'), but some are root-based, at least in the surface, e.g. *cant-o* 'I sing'.

Stem formation

Most verb forms in Spanish have (non-alternating) regular stems which can be further divided into three conjugations, identified by the TV, e.g. in the infinitive. 1st conjugation verbs have the TV -*a*-, e.g. *cant-a-r* 'to sing', 2nd conjugation verbs have the TV -*e*-, e.g. *com-e-r* 'to eat', and 3rd conjugation verbs have the TV -*i*-, e.g. *viv-i-r* 'to live'. The 1st conjugation is by far the largest (in terms of verb types), and almost all of the 1st conjugation verbs have regular stem forms and regular inflectional suffixes. A count of the verb (type) frequencies in the digital version of the dictionary of the Spanish language (Real Academia Española, 1995) confirmed this.[3] In this corpus, we found 9706 1st conjugation, 712 2nd conjugation and 740 3rd conjugation verbs. The 1st conjugation is also the open class *par excellence*. For example, English *to stress* 'to cause stress' becomes Spanish *estresar*, English *to film* Spanish *filmar*, and so on. Moreover, many 2nd and 3rd conjugation verbs have irregular forms. The 3rd conjugation has slightly more members than the 2nd conjugation, but they both have much fewer members than the 1st conjugation.

In addition to regular stem/root forms, there are approximately 900 verbs that have inflected forms with stem or root changes. About 30 of these are highly irregular, including verbs such as *estar* 'to be', *caber* 'to fit', *querer* 'to want', *poner* 'to put', *tener* 'to have', *ir* 'to go', *venir* 'to come'. This is illustrated in (3) for the past indicative and past participle forms of irregular *poner* 'to put'.

[3] These verb (type) frequencies were obtained by producing a list of all Spanish verbs available from the digital version of the *Diccionario de la lengua española*. Then, the verbs from the list were classified (by hand) according to their conjugation class.

594

(3)   puse   'I put'   puesto   'put-past participle'
      pusiste   'You put'
      puso   'she/he/it put'
      pusimos   'we put'
      pusisteis   'you put'
      pusieron   'they put'

The irregularity of this paradigm evidently affects the root, *pus-* in the past and *pues-* in the participle. Irregular stems may lack a TV, as for example in *puse*, *puso*, *puesto*. In these cases, the 1st and 3rd singular have the irregular inflectional endings -*e* and -*o*, instead of the regular ones -*í*, -*ió*, and the past participle ends in -*to*, instead of the regular -*ido*, which includes the TV. Less obviously from the orthography, irregular 1st singular and 3rd singular past forms and past participle forms bear stress on the root (whence the traditional label 'strong past'), rather than on the TV, as is the case in regularly inflected forms. The remaining forms of the paradigm in (3) combine the irregular root *pus-* with regular endings. The 2nd singular past indicative form *pus-i-ste*, for example, has the same theme vowel -*i*- and the same inflectional ending -*ste* as the corresponding form of a regular 2nd conjugation verb such as *comer* 'to eat' (see Table 1 below).

TABLE 1. *Inflectional paradigms of present and past tenses and imperatives*

Conjugation	Person and number	Present tense, indicative	Indefinite past tense, indicative	Imperative
1st conjugation: *cantar* 'to sing'	1st singular	canto	canté	
	2nd singular	cantas	cantaste	canta
	3rd singular	canta	cantó	
	1st plural	cantamos	cantamos	
	2nd plural	cantáis	cantasteis	cantad
	3rd plural	cantan	cantaron	
2nd conjugation: *comer* 'to eat'	1st singular	como	comí	
	2nd singular	comes	comiste	come
	3rd singular	come	comió	
	1st plural	comemos	comimos	
	2nd plural	coméis	comisteis	comed
	3rd plural	comen	comieron	
3rd conjugation: *vivir* 'to live'	1st singular	vivo	viví	
	2nd singular	vives	viviste	vive
	3rd singular	vive	vivió	
	1st plural	vivimos	vivimos	
	2nd plural	vivís	vivisteis	vivid
	3rd plural	viven	vivieron	

595

The status of diphthongized stem forms, e.g. *entender-entiendo* 'to understand – I understand', *rogar-ruego* 'to beg – I beg', is controversial. Diphthongization is common in Spanish, not only in verbs, and it has been argued that the diphthongized form can be derived by general phonological rules (Harris, 1985). On the other hand, such stem/root alternations are LEXICALLY determined in the sense that verbs with diphthongized stems form an arbitrary class: there are many verbs that have (non-low) vowels which do not undergo diphthongization, e.g. *notár-nóto* 'to note – I note'. Moreover, a nonce-probe task administered to Spanish-speaking adults (Bybee & Pardo, 1981) revealed that in many cases speakers were reluctant to productively generalize diphthongized stems to nonce verbs. As an alternative to deriving alternations such as *entender-entiendo* from general phonological rules, one may therefore argue that these diphthongized stems are opaque forms stored in the mental lexicon of a native speaker of Spanish.

In the present study, we focus our analysis on stems such as those in (3), which are uncontroversially regarded as irregular. Diphthongized stems will be analysed separately.

Inflectional suffixes

Inflectional paradigms of the regular present tense, past tense, and imperative of the three conjugations are shown in Table 1. In these paradigms, the 2nd and 3rd conjugations merge in all forms except in the 1st and 2nd plural present and in the prescriptive plural imperative (which is not used in Latin America and is colloquially replaced by the infinitive in Spain). These two conjugations differ from the 1st conjugation in the TV and in the ending of the 1st singular and 3rd singular (indefinite) past.

While most of the irregularities of Spanish verb forms affect the stem, there are also irregular inflectional suffixes. For instance, the 1st person singular past tense form *quise* 'I wanted' from *querer* 'to want' illustrates suffix irregularity (in addition to the root irregularity already mentioned): the regularly suffixed form would be **quisí*. Similarly, the 3rd singular past tense suffix in *pus-o* 's/he/it put' of the verb *poner* 'to put' is irregular; the regularly suffixed form for a 2nd conjugation verb would be **pus-ió*; (the completely regular form with respect to both stem formation and inflectional suffixation would be **ponió*).

Summarizing, Spanish verb forms exhibit a regular/irregular distinction for both stem formation and inflectional suffixation. Consequently, if the dual-mechanism model extends to Spanish, we would expect Spanish children to show a regular/irregular asymmetry with respect to both stem formation and inflectional affixation. That is, regular patterns, whether they be stems or inflectional affixes, should overgeneralize to irregular items, whereas generalizations of irregular patterns to regular verbs should be rare or non-existent.

596

Previous studies on the acquisition of Spanish verb inflection

Overregularization errors in Spanish child language have been observed by many authors (Hernández-Pina, 1984; Pérez-Pereira & Singer, 1984; Fernández, 1994; López-Ornat, 1994; Johnson, 1995; Radford & Ploennig-Pacheco, 1995; Mueller Gathercole, Sebastian & Soto, 1999; Serrat & Aparici, 1999), but a systematic and detailed study of the development of overregularization, comparable for example to Marcus *et al.*'s (1992) treatment of the English past tense, is not yet available. Here we will briefly summarize some relevant observations from previous studies.

Mueller Gathercole *et al.* (1999) analysed longitudinal data from two children between ages 0;9 and 2;6. They observed that Spanish children overapplied regular inflectional patterns to verbs that are irregular in adult Spanish. Interestingly they found that overregularization errors were absent from the earliest recordings and only occurred after the children started to use the regular rule productively and contrastively.

Johnson (1995) studied 1-hour speech samples of 42 monolingual Spanish-speaking children from 2;0 to 4;0. She found a total of 100 inflectional errors. There were cases in which children incorrectly used 1st conjugation forms instead of 2nd or 3rd conjugation forms, e.g. **rompó* instead of *rompió* 's/he/it broke', **caíba* instead of *caía* 'I fell'. The second type of error were cases in which children used a regular 2nd or 3rd conjugation form for a verb that requires an irregular 2nd or 3rd conjugation form, e.g. **poní* instead of *puse* 'I put'-past. There were also cases in which the child picked the wrong root for an irregular verb. On the other hand, Johnson did not include any irregularization errors, i.e. over-applications of irregular 2nd or 3rd conjugation forms to regular verbs, but this may be because she presented just a selected (rather than an exhaustive) list of all errors.

Radford & Ploennig-Pacheco (1995) analysed the morphosyntax of subjects and verbs in the grammar of a Mexican child acquiring Spanish as her first language between ages 2;2 and 2;8. They identified three kinds of verb inflection errors produced by this child: (i) morphological errors in which the child produced a regular affix and/or a regular stem form for a verb that requires an irregular form (e.g. **pusí* instead of *puse* 'I put-past'); (ii) conjugation class errors in which the child incorrectly inflected a 2nd or 3rd conjugation verb according to the 1st conjugation (e.g. **abré* instead of *abrí* 'I opened'); (iii) overapplications of 3rd singular forms in contexts that require 1st, 2nd singular or plural forms (e.g. **¿Tú presta tus monedas?* 'Can you let me have your coins?'). Radford & Ploennig-Pacheco (1995) argue that what is common to these error types is that the child replaces specific or irregular forms with default forms. These results are compatible with those of the other studies on Spanish child language mentioned above. Unfortunately, however, Radford & Ploennig-Pacheco's error analysis is example-

597

based rather than exhaustive. Irregularization errors are not reported, but this does not necessarily mean that such errors did not occur in the data. As in the previous studies on Spanish child language, questions also remain with respect to quantitative aspects of overregularizations and their development over time.

Serrat & Aparici (1999) studied longitudinal data from 5 monolingual Catalan-speaking children, 2 monolingual Spanish-speaking and 3 bilingual (Catalan/Spanish) children in the age range of 1;7 to 3;0. They found an interesting developmental difference between what they call agreement errors and overregularizations. While the former frequently occur even in the earliest data sets, overregularization errors emerge later in development. They also note that in most agreement errors a 3rd singular present form is used instead of some other grammatical person and/or number, similarly to what Radford & Ploennig-Pacheco (1995) observed; see their finding (iii) above. Note, however, that while Serrat & Aparici treat these incorrect 3rd singular present forms as commission (agreement) errors, Radford & Ploennig-Pacheco interpret them as default forms, i.e. as forms which are underspecified for finiteness features; we will return to this issue.

Given the results currently available it is hard to decide whether or not the dual-mechanism model extends to Spanish child language. To address this question, longitudinal data of children at relevant ages are required to shed light on the DEVELOPMENT of overregularizations. We also need large samples of speech for calculating meaningful overregularization rates, and we need to undertake a detailed analysis of the children's errors to determine whether they can be explained in terms of the distinction between rule-based and memory-based processes.

METHOD

Subjects

The data we have investigated consist of 64 samples of spontaneous speech or narratives from 15 children covering the age period of 1;7 to 4;7. Appendix A shows a detailed breakdown of the data. The total sample size is 48 hours (not including the Idaira data for which sample size was not provided; see appendix). There are longitudinal data from 4 children (María, Koki, Idaira and Pablo) in the relevant age range, supplemented by cross-sectional samples from 11 children. Most of the data are available from the 1995 and 2000 versions of the CHILDES database (MacWhinney, 1995, 2000). The data from Pablo, Rosmary and Solange were taken from other sources as shown in appendix A. The longitudinal data are based on spontaneous speech samples. María was recorded every 2 weeks from the age of 1;7 to 4;0, and her data represent the largest corpus. From Pablo we have a list of the inflectional errors he produced, but we do not have the correctly

598

inflected verb forms he produced.[4] We have therefore included Pablo's data in the analyses of error types, but not in the analyses of error rates and of how they change over time. The cross-sectional samples are based on semi-spontaneous speech in which the children were asked to talk about past events. The samples from Rosmary and Solange were recorded for 20 minutes each, while each of the 9 children from Díez-Itza & Pérez-Toral (1996) were recorded for 45 minutes.

The samples are representative of several varieties of Spanish. The relevant difference concerns the verb form used to refer to past-time events. In particular, María's Madrid variety favours a compound tense consisting of an auxiliary (the conjugated verb *haber*) followed by a past participle, e.g. *he comido* 'I have eaten'. In contrast, the other varieties prefer the simple preterite, e.g. *comí* 'I ate' (Díez-Itza & Pérez-Toral's (1996) children are from Oviedo, in Northern Spain; Idaira from the Canary Islands; Koki from Mexico; and Rosmary, Solange and Pablo from Venezuela).

Procedure

We extracted BY HAND all correct irregular verb forms (including one-word utterances) used by the children.[5] Incorrect verb forms were also determined by hand. Regular verbs were extracted both by hand and with the help of the FREQ programme of the CLAN software package (MacWhinney, 1995, 2000), which was run on all samples. Ambiguous forms were discarded. Such cases include reduced forms the children often produced in the early samples. For example, María produced the form *pi* which could refer to the action she was performing (*pintar* 'to draw') or it might refer to the object she was using (*lápiz* 'pencil'). In the former case, *pi* could be regarded as a verb form, in the latter one as a noun. We have not included such ambiguous cases. Utterances such as *venga* 'come', *mira* 'look', *oye* 'listen', which are typically used as discourse markers and are likely to be formulaic, were also discarded.

We counted as 'irregulars' verb forms that require unpredictable stem root changes (see e.g. (3) above) and/or inflectional suffixes that are different from those of the regular paradigms illustrated in Table 1. 'Regulars' are verb forms without stem changes and with inflectional endings taken from the paradigms in Table 1. In the data under study, there were 3446 correct irregular tokens, 2071 correct regular tokens, and 174 verb tokens that were incorrect. In the analyses to be reported below, we will focus on these cases. In particular, as mentioned above, verb forms that require diphthongized stems, such as *dormir–duermo*, were not included as 'regulars' or 'irregulars',

---

[4] Unfortunately, Posada (1997) did not publish Pablo's longitudinal data. However, she does provide a list of all overregularization errors and a description of them.
[5] Parts of the data analyses to be presented in the following originally stem from unpublished work by Aveledo (1999).

599

since their status is controversial in the linguistic literature; we have analysed these cases separately.

RESULTS

In what follows, we first present an analysis of the types of verb inflection errors produced by the children. Secondly, we determine error rates and how they change throughout development. Finally, we examine how verb inflection errors are related to other aspects of language development, particularly to changes in the children's vocabulary and to syntactic development.

Error types

Table 2 presents a breakdown of all errors across children. These data show a clear regular/irregular difference. There were 168 verb tokens in which a stem formation and/or an inflectional suffixation error occurred in verb forms that are IRREGULAR in the adult language (against 3446 correctly used irregular forms), yielding an overall error rate of 4·6%. These errors occurred in present tense and imperative forms ($= 77$), in past tense forms ($= 83$), in participles ($= 7$), and in 1 incorrect imperfect subjunctive form. On the other hand, there were only 2 cases in which errors occurred in verb forms that are regular in adult Spanish (against 2071 correctly used regular verb forms), i.e. a rate of 0·001%. Thus, errors occurred mainly for irregular verb forms, and only rarely for regular ones. In addition, there were 4 incorrect imperfect indicative forms in which a 1st conjugation form was used for verbs that belong to the 2nd conjugation; these were classified as 'other errors' in Table 2.

To further analyse potential regular/irregular distinctions in the data, we distinguish between errors of inflectional suffixation and stem/root errors. The former involve the incorrect substitution of an inflectional suffix, while the latter refer to cases of incorrect root or stem forms.[6] These two basic error types are further divided into three subtypes. The first error type is *overregularizations*; see A-I and B-I in (4). In such cases an irregular stem or affix is replaced by the regular one; we also include here overapplications of the (default) 1st conjugation to 2nd conjugation or 3rd conjugation verbs of adult Spanish. The second type is IRREGULARIZATIONS, i.e. cases in which children overapply irregular stem and/or affixes to verb forms that are regular in the adult language (AII and BII in (4)). The third type (OTHER

[6] A JCL reviewer wondered whether the distinction between stem formation and inflectional processes is already operative in young children. We will return to this question in the *Discussion* section, where some evidence is provided to show that stem formation and inflectional suffixation can indeed be dissociated in Spanish children (at least after they start to produce verb inflection errors).

600

TABLE 2. *Incorrect verb forms produced by each child*

	Present tense and imperatives		Past tense		Participles		Subjunctive	
	Irregular incorrect	Regular incorrect	Irregular incorrect	Regular incorrect	Irregular incorrect	Regular incorrect	Irregular incorrect	Other errors
María	24		1		1			
Koki	18		7	1				
Idaira	5		7		1			2
Pablo	24		50	1				1
Solange			1					
Rosmary	1		2		1			
Child 1	4		1		1		1	
Child 5			6		3			
Child 7			4					1
Child 10	1							
Child 11			1					
Child 12			2					
Child 13			1					
Total	77		83	2	7		1	4

ERRORS) represents cases in which the child replaces either a correct regular form by some other (incorrect) regular form or a correct irregular verb form by some other irregular but incorrect form. The error taxonomy is shown in (4), and the distribution of these errors in Table 3. Appendix B presents a list of all types of stem and/or suffixation errors found in the data, shown separately for each child.

TABLE 3. *Distribution of error types*

A. Stem errors		B. Suffixation errors	
I. overregularizations	116	I. overregularizations (n = 132)	
		a. 1st conjugation overapplications	8
		b. conjugation-internal regularizations	124
II. irregularizations	1	II. irregularizations	0
III. other errors	3	III. other errors	1
Totals	120	*Totals*	133

(4) A. *Stem/root errors*
  I. *Overregularizations*:
    sabo 'I know' correct form: *sé* (María, 2;1)
  II. *Irregularizations*:
    cayí 'I fell' correct form: *caí* (Koki 2;1)
  III. *Other stem errors*:
    punieron 'They put-past' correct: *pusieron* (Rosmary 3;10)
  B. *Suffixation errors*
  I. *Overregularizations*:
    a. *Overapplications of 1st conjugation forms*: incorrect use of a 1st conjugation suffix instead of a 2nd or 3rd conjugation one: queriba 'I /s/he/it wanted' correct: *quería* (Idaira 3;7)
    b. *Conjugation-internal regularizations*: incorrect use of a regular 2nd or 3rd conjugation suffix instead of the required irregular one: pusí 'I put-past' correct form: *puse* (Koki 2;4)
  II. *Irregularizations*: there were no errors of this kind in the data.
  III. *Other suffixation errors*:
    ensucí 'I got dirty' correct form: *ensucié* (Pablo 2;6)

Table 3 shows that among the 174 incorrect verb forms, there are 120 stem errors and 133 errors of inflectional suffixation; that there are more errors (n = 253) than verb forms (n = 174) results from the fact that many verb forms contain both a stem and a suffixation error. Both among the stem errors and

602

among the suffixation errors, overapplications of regular patterns, i.e. error types A-I and B-I account for the vast majority of errors, whereas over-extensions of irregular patterns are practically non-existent. There are 116 stem overregularization errors (type A-I) compared to just 1 irregularization error, and 132 suffix overregularizations but no single suffix irregularization. Moreover, 'other errors' are also extremely infrequent; there are 3 stem errors (see A-III) and one case (B-III) in which the required regular suffix is replaced by a corresponding regular suffix from a different conjugation.

It is important to note that among the stem errors, there are no conjugation class errors. Children produced, for example, *romp-i-da* instead of *rota* 'broken', but not *romp-a-da*, i.e. they combined the regular root *romp-* with the 2nd conjugation theme vowel *-i-*, treating *romper* as a regular 2nd conjugation verb, rather than an irregular one. Similarly, among the 133 suffixation errors, there are only 9 cases with errors in conjugation class, and in 8 of these a 1st conjugation suffix was used instead of a 2nd or 3rd conjugation one. All other 124 suffixation errors were truly inflectional errors in which a regular form is used instead of an irregular one, while the verbs' conjugation class is as in the adult language. These observations show that errors occur in irregular roots and affixes, while conjugation class assignment is most often correct.

The suffixation errors can be further classified as in (5):

(5) a. *Overapplications of 1st conjugation endings*: the 1st singular past *-é* and the 1st/3rd singular imperfect indicative *-ba* are overregularized to 2nd conjugation verbs such as *querer* 'to want', *tener* 'to have', *traer* 'to bring', yielding errors such as *queriba*, *teniba*, and *trajé*.

   b. *Overapplications of regular 2nd/3rd conjugation endings*: the 1st singular past *-í*, the 3rd singular past *-ió*, and the imperative *-e* are overregularized to irregular verbs such as *poner* 'to put', *tener* 'to have', *hacer* 'to do', yielding overregularization errors such as *punió, *pusí, *tení, *hace.

   c. *Overapplications of other regular endings*: the 1st singular present tense *-o* and the participle suffix *-d-* (both of which apply to all regular verbs irrespective of conjugation) are overregularized to irregular verbs such as *saber* 'to know' and *romper* to break', producing errors such as *sabo* and *rompida*.

The distribution of the suffixation errors shows that regular inflectional affixes are overapplied to verbs that require irregular forms. On the other hand, there was no single case in which an irregular ending was overapplied to a verb that required a regular one.

We also examined combinations of affixation errors with particular stem types. Table 4 presents frequencies of these for each child. In 85 % of cases

603

TABLE 4. *Suffixation errors and stem types*

| | Suffixation errors | | | |
	Combined with correct irregular stem	Combined with incorrect irregular stem	Combined with regular stem	Totals
María			14	14
Koki	3	1	10	14
Idaira			12	12
Pablo	9		63	72
Rosmary			1	1
Child 1			6	6
Child 5	4		2	6
Child 7			3	3
Child 10			1	1
Child 11	1			1
Child 12		2		2
Child 13			1	1
Totals	17	3	113	133

(= 113 out of 133), a suffixation error co-occurs with a regular stem for verbs that require irregular stems in Spanish. These are cases such as *rompida and *tení instead of *rota* and *tuve*, in which the regular root (*romp-er*, *ten-er*) was combined with a regular inflectional affix in verb forms that require both an irregular root and an irregular affix. Thus, in most of the errors the child treats an irregular verb as if it were entirely regular.[7] In addition to these cases, there are 17 suffixation errors in verb forms with the correct irregular stem/root forms, e.g. *tuví instead of *tuve* 'I had' (Child 5, 3;7), and 3 cases in which a suffixation error co-occurs with an incorrect irregular root, for example *punió instead of *puso* 'S/he/it put-past' (Child 12, 3;7).

Finally, we searched the longitudinal data (María, Koki, and Idaira) for verb forms that require diphthongized stems in the adult language, e.g. *dormir–duermo* 'to sleep–I sleep', as well as for incorrectly used diphthongized stem forms of verbs, e.g. *cuem-e instead of *com-e* 's/he/it eats'. We found a total of 603 tokens of correctly used diphthongized stem forms and 107 errors; a breakdown of these cases is shown in Table 5. All of

[7] It is true that Pablo produced more overregularization errors than the other children. However, the distribution of the error types in Pablo's data is similar to the other children. For Pablo, 87% of the incorrectly used inflectional suffixes combine with regular verb stems, for the other children this was the case in 82%. Pablo produced 12·5% suffixation errors with a correct irregular stem; for the other children the rate was 13·1%. Pablo did not produce any suffixation errors with incorrect irregular stems, while the other children did so in just 4·9% of cases. These comparisons show that even though the absolute number of errors is higher with Pablo than with the other children, Pablo does not show a different pattern from the rest of the children.

604

TABLE 5. *Verb forms that require diphthongized stem vowels*

		Present tense and imperatives	
	Type of alternation	Correct (diphthongized) forms	Incorrect (non-diphthongized) forms
María	o ∼ ue	110	14
	u ∼ ue	2	13
	e ∼ ie	250	40
Koki	o ∼ ue	41	12
	u ∼ ue	1	
	e ∼ ie	148	21
Idaira	o ∼ ue	16	
	u ∼ ue		
	e ∼ ie	35	7
Totals		603	107

the errors have the same form: the children produced a regular (non-diphthongized) stem form in a context that required a diphthongized one, e.g. **juga* instead of *juega* 's/he/it plays' (María 2;2). In the data under study, there were no cases (such as **cuem-e*) in which a regular stem form was replaced by a diphthongized one. Thus in terms of error types, diphthongized stem forms exhibit the same behaviour as the irregular stem forms discussed above. They are not overapplied to other verb forms, and when they are required, they are sometimes replaced by the regular stem.

Summarizing, we found a striking asymmetry in the distribution of errors for both stem formation and inflectional suffixation: Spanish children generalize regular patterns, i.e. the unmarked roots and regular inflectional affixes, to irregular items (and to diphthongized stem forms), but irregular (and diphthongized) stems as well as irregular affixes do not extend to regular forms (and to non-diphthongized stems). The different generalization patterns indicate a sharp regular/irregular distinction in Spanish children's inflectional errors.

Overregularization rates

To examine overregularization rates in Spanish child language, we calculated error scores in present tense (including imperatives), past tense and participle forms, separately for the longitudinal data and the cross-sectional samples. Overregularizations were defined in terms of the error typology in (4) above. In the analysis of overregularizations, we have only included verb forms that are irregular in adult Spanish. Verb forms that require diphthongized stems were analysed separately. We counted error types A-I and B-I as over-regularizations. Each incorrect verb form was only counted once, even if

605

it contained a stem/root error and an inflectional error. Overregularization rates were calculated as in Marcus *et al.* (1992), as the proportion of tokens of irregular forms that were overregularizations.

We found that in the longitudinal samples, the overregularization rates range from 1·5 % to 3·3 % with a mean of 2 %, whereas in the cross-sectional samples the overregularization rates are higher, with a mean of 12 %, a rate that is in line with estimates from semi-elicited speech (Marcus *et al.*, 1992). Appendix C-1 shows the overregularization rates for the different children. We also note that the overregularization rates are higher in the past tense than in the present tense. In the data from María, Koki and Idaira, the mean past-tense overregularization rate is 14 % compared to 1·5 % for the present tense (see appendix C-2). This difference may result from the fact that the past tense has more irregular forms, i.e. more opportunities of producing errors, than the present tense. We also determined error rates for verb forms that require diphthongized stems. Recall that in such cases children sometimes replace the diphthongized stem by the corresponding regular one, as for example in *juga instead of *juega* 'I play'. The mean rate for this kind of error is 18 %; appendix C-3 shows the error rates for individual children. This is similar to the 14 % mean overregularization rate for irregular past tense forms, suggesting that diphthongized stem forms behave in similar ways in Spanish child language as irregular verb forms, not only in terms of error types, but also in terms of error rates.

To assess the effect of frequency on overregularization rates, we adopted an analysis from Maratsos (2000). Verbs were grouped according to their sample frequencies: verbs sampled 1–9 times, verbs sampled 10–49 times, verbs sampled 50–99 times, and verbs sampled 100 times or more. Mean overregularization rates were calculated for the individual verbs of each of these categories. These are shown in Table 6.

TABLE 6. *Overregularization rates for verbs of different sample frequencies*

	Sample = 1–9		Sample = 10–49		Sample = 50–99		Sample = 100+	
	Number*	Rate	Number*	Rate	Number*	Rate	Number*	Rate
Idaira	31	0·15	5	0·02	0	–	0	–
Koki	22	0·24	8	0·11	0	–	0	–
María	24	0·11	9	0·003	4	0	6	0·02

* Number of different irregular verb forms (types) in the relevant sample frequency class.

Table 6 shows that there are some individual differences, with lower overregularization rates in sample-frequency bands 1–9 and 10–49 for María than for Idaira and Koki. Clearly, however, overregularization rates are

606

higher for verbs with low sample frequencies than for verbs that the children use more frequently, and this holds for the three children studied here, thus replicating Marcus *et al.*'s (1992) and Maratsos' (2000) findings for English. Yet, even for verbs with low sample frequencies, overregularization rates are relatively low, with a mean rate of 0·17 for verbs that were sampled one to nine times.

The development of overregularizations

We examined the development of overregularizations in the longitudinal data from María and Koki. The third child, Idaira, was only recorded from age 2;7 onwards, which (as will be seen) is too late for determining the onset of overregularizations. Overregularization errors were determined in the same way as above, i.e. error types A-I and B-I (see (4)) in present tense (including imperatives), past tense and participle forms. For each sample of the longitudinal data, overregularization rates were calculated as proportions of irregular forms (tokens) that were overregularizations (see Marcus *et al.*, 1992: 34ff.; Maratsos 2000: 191f.). The results are shown in Figures 1 and
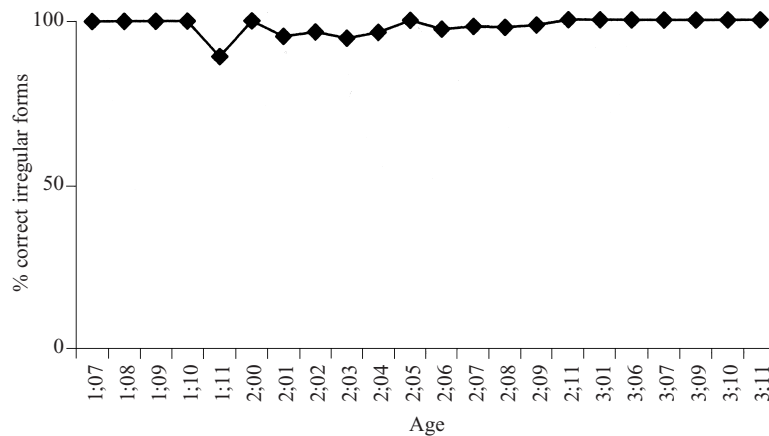

Fig. 1. Percentage of María's irregular verb forms (tokens) that are correct (100 % minus the overregularization rate).

2. Following Marcus *et al.* (1992) these graphs show the proportion of irregular forms that are correct, i.e. 100 % minus the overregularization rate.

The first overregularizations, at age 1;11 (for María) and 2;1 (for Koki), are preceded by a phase without any errors. Moreover, even though overregularizations do not completely disappear, the overregularization rates decrease when the children get older, i.e. from 2;6 onwards.
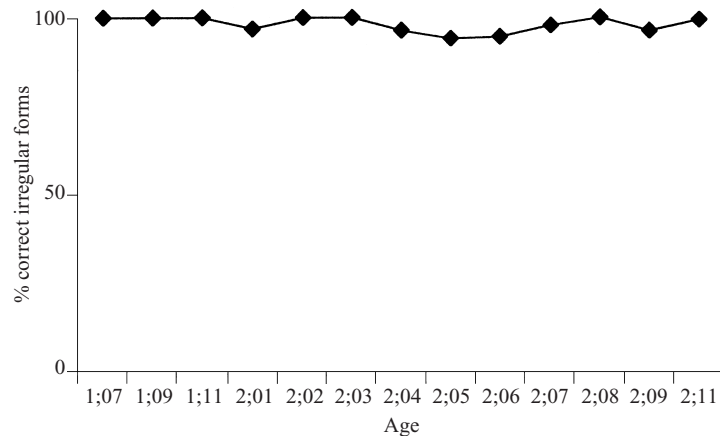
607

Fig. 2. Percentage of Koki's irregular verb forms (tokens) that are correct (100% minus the overregularization rate).

Note, however, that developmental changes in longitudinal data might be more apparent than real in that they could be due to sampling errors. Clearly, the longitudinal data sets available to us represent only a limited part of a child's speech, and it is possible that a child produced overregularization errors at an earlier age than the first one occurring in the recorded data. We therefore tested developmental changes in the longitudinal data by applying the probability measure from Marcus *et al.* (1992: 41). The idea is that if the overregularization rates were not significantly different at different ages, then there should be a high probability of obtaining the consecutive correct number of irregulars for the period preceding the onset of overregularizations from the period when the child starts to produce overregularizations. Table 7 presents the ages of the children when they first produced over-regularization errors, the corresponding overregularization rates, the number of consecutive correct irregulars in preceding samples, and the probability of obtaining the latter on the basis of the individual overregularization rate.[8]

The figures in the last column of Table 7 show that the probability of predicting the number of observed correct irregulars prior to the first overregularization is very small, if the overregularization rate is assumed not to have changed when the children start to produce such errors. This shows that the latter assumption is incorrect and that the overregularization rate is unlikely to be constant over time. Instead, it is more likely that children

[8] The analysis adopted here from Marcus *et al.* (1992) does not presuppose that once a child starts overregularizing, s/he does it consistently in every sample. Instead, the analysis compares two different time periods, i.e. the period preceding the onset of over-regularization and the period during which the child produces such errors. It is possible that within the latter period overregularization errors appear sporadically.

608

TABLE 7. *Test of U-shaped development*

	First overregularization		Consecutive correct in preceding samples	Overregulation rate	Probability
	Age	Sample			
María	1;11	5	119	0·015	0·16
Koki	2;1	4	212	0·023	0·007
Idaira	3;3	3	96	0·033	0·04

display correct performance for irregulars before they start to produce overregularization errors.

Summarizing the results presented in this section, we found that children start to overregularize after a period of correct performance, and that overregularization errors decrease when the children get older. Thus, overregularization errors in Spanish-speaking children seem to follow a U-shaped developmental curve.

Overregularizations and verb frequencies

One controversial issue in studies of the acquisition of inflection concerns the role of (type) frequency for the development of overregularizations. Whereas Marcus *et al.* (1992) claimed that overregularization errors are not directly related to the children's vocabulary, other researchers have argued that the onset of overregularizations can be explained in terms of the (type) frequency distribution of regular and irregular verb forms in the child's vocabulary (see e.g. Bybee, 1995). To test potential links between verb frequencies and the onset of overregularizations in child Spanish, we have analysed the use of regular and irregular verbal stem types in comparison with the onset of stem overregularization errors. Included were all present and past tense stem types except direct imitations that occurred in the two large longitudinal corpora from María and Koki. Idaira was not analysed for this purpose, because the corpus was too small. Even such extensive spontaneous speech corpora, however, can only account for the actual words the children used in a given sample, and the possibility that the children know more verb stems than they produce cannot be excluded. In order to minimize the danger of under-estimating the size of the child's verb vocabulary, we calculated cumulative (type) frequencies, adopting an analysis from Maratsos (2000: 205) for English child language (see Figures 7 to 9 in Maratsos, 2000). If a verbal stem is used in a given sample, it is credited to the child's vocabulary from then on. Note that this measurement is very generous to low-frequency stems. We counted stem types as explained in the *Method* section: 'irregulars' are stems that require irregular root changes (e.g. *pus-*, the irregular past tense stem of

609

poner 'to put'), 'regulars' do not involve any root changes. The cumulative frequencies for regular stems are counted separately for the three conjugations. The results are shown in Figures 3 and 4.

Figures 3 and 4 show that below the age of 2;4 for Koki and below 2;0 for María, regular and irregular stems have similar (type) frequencies in the children's vocabulary. It is only at a later age that regular forms (of the 1st conjugation) begin to clearly outnumber irregular ones. Thus, if differences between the children's use of regular and irregular stems were to follow from the frequency distribution in their vocabularies, we would not expect to find any such differences before 2;4 for Koki and 2;0 for María. This, however, is not the case. Both children produce stem overregularizations before that age, María from 1;11 and Koki from 2;1 onwards; this is indicated by the arrows in Figures 3 and 4. At the same time, none of the children produce overapplications of irregular stems. Thus, the regular/irregular distinction in child Spanish (as evidenced by children's overregularizations and the lack of irregularization errors) cannot be attributed to the (type) frequency distribution of regular and irregular stems in the children's vocabulary.

Finiteness marking and the development of overregularizations

If, as shown in the previous section, the onset of overregularizations is not directly tied to the child's developing vocabulary, one wonders what causes children to produce such errors. An alternative possibility is that over-regularization errors are caused by changes in the children's grammar, more specifically by independent SYNTACTIC developments. Studies on syntactic development in different languages have led to the identification of a developmental stage, sometimes dubbed the optional-infinitive stage (Wexler, 1994), in which children alternate between using finite (i.e. tense and/or agreement-marked) verb forms and non-finite verb forms in contexts in which adults require finite verb forms. How to account for this phenomenon is a controversial matter; see e.g. Rizzi (1993); Wexler (1994); Clahsen, Eisenbeiss & Penke (1996) for discussion. What is common to these accounts is that the children's syntactic representations in the optional infinitive stage are said to be underspecified (compared to corresponding adult representations) with respect to tense and agreement features. If these features are fully specified, the child will produce a correct finite verb form, just like an adult. If, however, tense and/or agreement features are left unspecified, the child will resort to an unmarked (bare or infinitive) form, hence the optional use of finiteness markers.

Studies on the acquisition of Italian, Catalan and Spanish (Guasti, 1993; Torrens, 1995) have shown that children sometimes produce root infinitives, but that in two to three-year-old children they are much less frequent than in German, Dutch and Swedish children at similar ages. Guasti (1993) found, for example, that in Italian children root infinitives drop to less than
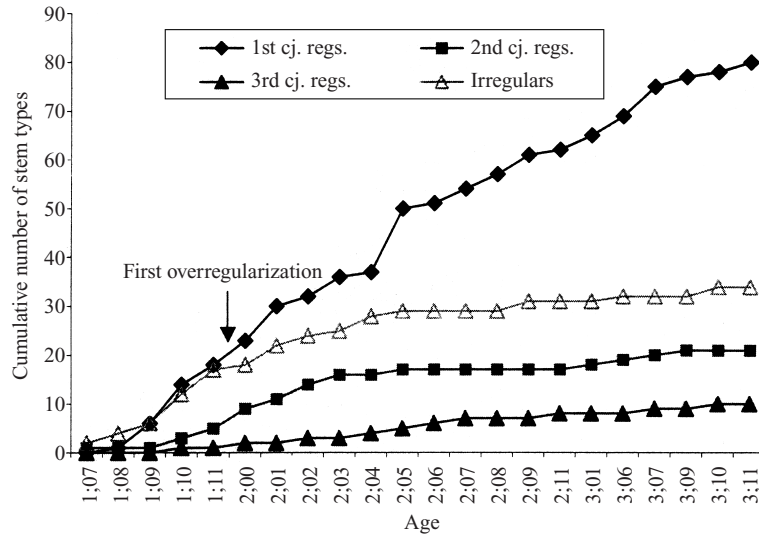
610

Fig. 3. María's cumulative regular and irregular stem types for present and past tense.
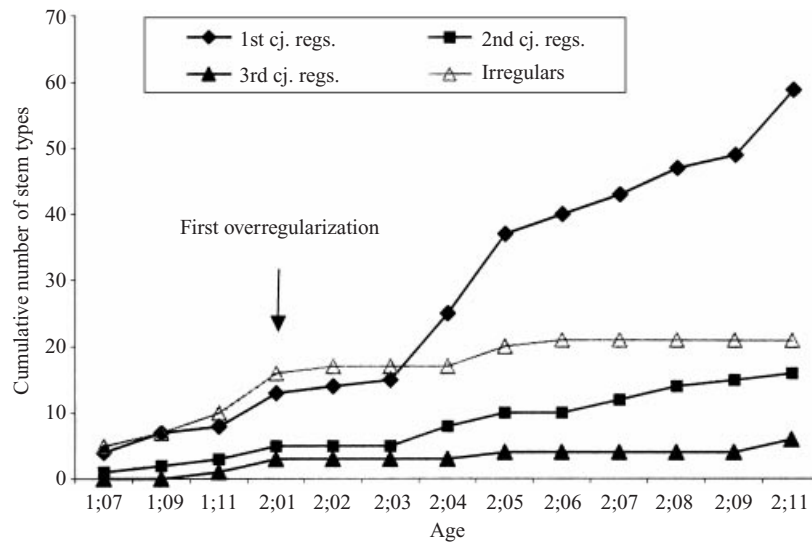

Fig. 4. Koki's cumulative regular and irregular stem types for present and past tense.

5 % by around age 2;0. In addition, it should be noted that Spanish (as well as other Romance languages) have another form which children might use as an underspecified form, and this is the 3rd singular of the present indicative. Recall that this form, e.g. *cant-a* 's/he/it sings' consists of the base stem, i.e. the lexical root plus the theme vowel without an overt tense or agreement
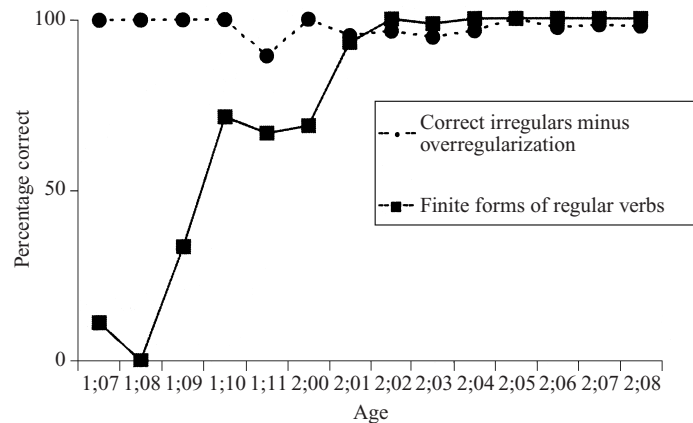
611

Fig. 5. Proportion of María's finite forms of regular verbs in obligatory contexts and overregularizations.

ending, and that this is the base form from which most other forms in the paradigm are derived, e.g. the 2nd singular *canta-s*, the 1st plural *canta-mos*, the 2nd plural *cantá-is* or the 3rd plural *canta-n*. Studies of child language development have indeed found that children learning Italian, Catalan, Spanish, and other languages with stem-based inflectional paradigms produce 3rd singular forms in cases in which other forms are required (Hernández-Pina, 1984; Guasti, 1993; Serrat & Aparici, 1999). For Spanish, Radford & Ploennig-Pacheco (1995) found, for example, that 20 % of the 3rd singular forms produced by the Mexican child they studied appeared in contexts that required 1st and 2nd person forms, while on the other hand all the 1st and 2nd forms used by the child showed correct 1st and 2nd person agreement. Consequently, Radford & Ploennig-Pacheco (pp. 52f.) claim that 3rd singular present tense forms may serve as a kind of default form in child Spanish.

We tested whether the onset of overregularizations is linked to the development of finiteness marking in the María corpus, the largest longitudinal data set available to us. We analysed the longitudinal data up to sample 14 (age: 2;8), since after that age overregularization errors have practically disappeared from María's data. To make sure that the two measures are independent, we calculated overregularization rates based on irregular verb tokens, and finiteness marking based on regular verb tokens. Included in the analysis of finiteness marking were all obligatory contexts, i.e. sentences in which adult Spanish requires the use of a finite form of a regular verb. Ambiguous contexts, for example answers to questions in which an infinitive form would be possible in spoken adult Spanish, were left out. For example, María (2;1) saying (*le*)*vantar* in response to her father's question *lo vas a levantar, ¿no?* 'Are you going to pick it up?' were not

612

included. We counted all finite verb forms María produced in contexts that unambiguously required a finite verb form and compared this with how often she produced underspecified forms in such contexts. As 'underspecified' forms we counted infinitives in sentences that did not contain a finite verb form and 3rd singular present tense forms used in contexts that required some other person, number or tense form. An example is shown in (6).

(6) Mother: ¿Qué estás haciendo María, mi vida?
              'What are you doing Maria, darling?'
     María: a timpá e culito [= a limpiar el culito]   M (1;11)
              'to clean the bum' ('I'm cleaning the bum')

The results can be summarized as follows. Before the first over-regularization error appears, i.e. during the period of 1;7 to 1;10, there were 42 obligatory contexts for finite forms of regular verbs, and in 23 (= 55%) of these, María produced an underspecified form; in 20 cases an infinitive and in 3 cases an incorrect 3rd singular present tense form. This is different for the later period. Between 1;11 and 2;8, i.e. during the period in which María produced overregularization errors, there were 253 obligatory contexts for finite forms of regular verbs, and only in 11 (= 4·3%) of these did María produce an underspecified form (8 infinitives and 3 incorrect 3rd singular present tense forms). This suggests that finiteness marking has become obligatory during this later period. To see the relationship between over-regularizations and finiteness markings in more detail, we plotted in Figure 5 the proportion of finiteness markings of regular verbs in obligatory contexts against overregularization rates in finite verb forms; both counts are based on token frequencies. Note that in contrast to Figure 1, we have excluded overregularizations in non-finite verb forms (i.e. in participles) from the present analysis, since these are unlikely to be related to the emergence of finiteness markings. The dotted line in Figure 5 represents the proportion of irregular present tense, imperatives, and past tense forms that are correct, i.e. 100% minus the percentage of overregularizations.

Figure 5 shows that overregularizations start to emerge once finiteness markings are produced in more than 70% of the obligatory contexts. The graph for regular verbs illustrates the gradual increase of sentences with fully specified finite verb forms. The finding that after age 2:0, root infinitives practically disappear from María's speech replicates earlier results on child Spanish, Italian and Catalan and indicates that finiteness features are now required to be specified under the same circumstances as in the adult grammar. At the same time, however, the child's still developing lexicon does not always provide the required finite form of irregular verbs or the child might be unable to retrieve it. It is in such cases of mismatch between syntactic requirements and retrieval failures that overregularization errors arise. When the syntactic requirement (to fully specify tense and agreement

613

features) is not yet in place, i.e. before 1;10 for María, the child can alternate between underspecified and (fully specified) finite verb forms. Over-regularization errors do not occur at that stage even in cases in which lexical retrieval of the correct form fails, because in such cases the child is free to use a verb form that is underspecified for finiteness. We conclude that the (delayed) onset of overregularizations appears to be caused by a syntactic requirement (= to specify finiteness) paired with lexical gaps and/or lexical retrieval failures for the correct finite form of irregular verbs.

DISCUSSION

The main empirical finding from the previous sections is that regular and irregular processes seem to be dissociated in Spanish child language, both in terms of the distribution of the errors and in terms of their developmental patterns. The dual mechanism model of inflection postulates sharp regular/irregular dissociations, and in the following discussion we will explore whether it can account for the findings reported in the present study.

Generalization properties of regular and irregular forms in child Spanish

We examined stem formation and inflectional suffixation separately in the Spanish data and found that in both domains children overapplied regular forms to irregulars, whereas irregular forms do not generalize to regulars. We took this as support for the regular/irregular distinction posited in the dual-mechanism model. One might, however, question the appropriateness of separating stem formation from inflectional processes for analysing early child language, even if such a distinction is needed for describing the corresponding adult language. It is conceivable, for example, that at least initially when they start producing verbs, children store all inflected verb forms as wholes without distinguishing between stems and inflectional suffixes.

Note, however, that while in the current data set children produce verbs from the youngest ages onwards, errors in inflected verb forms do not appear until at a later age (see Figures 3 and 4). For our error analysis and interpretation of the data, we therefore need to determine whether at this later age Spanish children distinguish between stem formation and inflectional processes. There is evidence from different sources that this is indeed the case.

Consider, for example, mixed errors such as those in (7) with correct irregular stems but incorrect inflectional endings:

(7)   a. vinió   's/he/it came'   (correct: vin-o)   Pablo (2;3, 2;6)
      b. pusí   'I put-past'   (correct: pus-e)   Koki (2;6), Pablo (2;11)

In such errors, marked (irregular) stems of the corresponding verbs (e.g. *vin-* and *pus-*) are combined with regular inflectional suffixes (instead of

614

irregular ones) indicating that children can manipulate stems and inflectional endings separately. There are 17 such cases in the data under study from different children (Pablo, Koki, Children 5, 11, 12). The data from Pablo and Koki indicate that such errors occur early in development; for Pablo, for example, (7a) represents the first overregularisation error documented in the data.

Another piece of evidence for the distinction between stem formation and inflectional suffixes comes from the use of theme vowels (TVs) in over-regularization errors:

(8) devolvido 'returned-participle' (correct: devuel-t-o) María (2;4)

Errors such as this one occur early in development, from 2;3 onwards for María and Pablo, and from 2;5 for Koki; they are also documented for most other children under study. In (8), both the root and the inflectional ending are incorrect; instead of the required irregular forms the child produced an unmarked regular root and a regular inflectional ending. Yet, the TV supplied by the child is correct; *devolv-* is a 2nd conjugation root for which the required TV is *-i-*. This does not only hold for the example in (8); indeed, as pointed out in relation to the error types in (4), there was no single case of a conjugation class error in the children's stem formations. For example, errors such as **devolv-a-da*, i.e. a regularized form of *devuel-t-o* with the 1st conjugation TV *-a-* were not represented in the data. Thus, even in cases in which inflectional suffixes are overregularized, stem formation (= i.e. the combination of roots and TVs) seems to function appropriately according to the root's conjugation class. These observations indicate that in the over-regularization errors in inflected verb forms that Spanish children produce, stem formation and inflectional processes can be dissociated.

We conclude that the linguistic distinction between stem formation and inflection proper, which is essential to the structure of morphologically complex words in adult Spanish (as well as in other Romance languages), can also be applied to an error analysis of inflected verb forms in child Spanish.

With respect to STEM FORMATION, we found that almost all errors involve regular (non-alternating) roots replacing irregular (or diphthongized) ones; there was only one instance of an irregular stem replacing a regular one. One way of accounting for this finding is in terms of default inheritance hierarchies (Corbett & Fraser, 1993; Wunderlich, 1996) according to which the different stem/root variants of a lexeme are stored as subnodes of hierarchically structured lexical entries. Thus, while the lexical entry of a regular verb such as *cantar* 'to sing' contains just one root *cant-* and does not have any subnodes, an irregular verb such as *poner* 'to put' has a regular root *pon-* and subentries with irregular roots (*pus-*, *puest-*, *pong-*, *pondr-*) which are restricted to particular inflected word forms. Stem/root errors arise when subnode information containing the correct pairing of phonological strings

615

and morphological feature values for irregular roots is not available or not accessible to children. In such cases, they fall back on the regular root producing errors such as *pon-i-ste instead of the correct *pusiste*. In this way, the regular root serves as a default form in circumstances in which the required specific forms are not retrieved.

What is common to the SUFFIXATION ERRORS is that forms from the regular paradigms are overapplied to verbs that require irregular forms, while irregular endings are not overapplied to verbs that require a regular form. For example, in errors such as *tení 'I had', the regular 1st singular past tense affix -í is overapplied to a verb that requires an irregular past-tense form (= *tuve*). Corresponding overapplications of irregular past-tense forms to regular verbs were not found. This difference is compatible with the different generalization processes postulated in the dual-mechanism model. Regular patterns are based on symbolic categories or rules, and may therefore generalize freely to any instance of such categories or rules. The -í rule, for example, may apply to any 2nd or 3rd conjugation verb (unless it is blocked by some irregular form). In cases such as *tení, the irregular word form that would block the application of the -í rule was not available and/or not retrieved, and thus led to an overregularization error.[9] The generalization of irregular patterns, on the other hand, depends upon the similarity and frequency of previously learned items. In Spanish, however, there are only about 30 verbs that take irregular suffixes. It is therefore conceivable that the system in which these irregularities are represented is too weak to foster associative generalizations, hence the lack of overapplications of irregular endings in the Spanish child data.

The role of phonological transparency in children's overregularization errors

According to the dual-mechanism model, the different generalization properties of regular and irregular forms are explained in morphological terms. Consider, alternatively, a PHONOLOGICAL account of children's over-regularization errors. Orsolini (1999) argued that Italian children overapply

[9] Note that while previous studies on overregularizations of inflection examined one regular affix per category in comparison with children's performance on irregulars, we have identified overregularizations of different inflectional affixes in child Spanish. As pointed out by a JCL reviewer, this raises the question of how children choose between different regular rules of inflection. Inflectional rules in Spanish encode morphosyntactic features (e.g. past vs. present tense), and these features determine which rule is chosen, in much the same way in which in English a plural feature on a noun triggers suffixation of -s and a past tense feature on a verb triggers suffixation of -ed. Moreover, inflectional suffixation may also depend on conjugation class, and in such cases it is the particular root's conjugation class that determines which regular suffix needs to be supplied. Thus, in addition to rules that apply to all conjugations, Spanish children also need to learn conjugation-specific inflectional rules, something which is not required for the acquisition of the English past tense or German participles.

616

inflectional patterns that are phonologically transparent, i.e. that do not involve any root changes, whereas inflectional patterns that involve root changes, e.g. the past definite and participle forms *prese–preso* of the verb *prendere* 'to take', have extremely low generalizability. In this way, the fact that Italian children generalize 1st conjugation and the no-root-change 2nd and 3rd conjugation forms is claimed to follow directly from the phonological form of the inflected words and without any reference to their morphological structure.

Are the overregularization errors produced by Spanish-speaking children explainable in these terms? It is true that in most of the stem errors, the children produced the unmarked regular stem instead of the required irregular one, as for example in **poniste* instead of the correct *pusiste* 'you put-past'. These errors are compatible with a phonological transparency account. The suffixation errors, however, cannot be explained in these terms. Consider cases such as **queriba* instead of *quería* 's/he/it wanted', **creiba* instead of *creía* 's/he/it think', and **teniba* instead of *tenía* 's/he/it had'. In these cases, the target forms and the forms the children produced do not differ with respect to phonological transparency, and yet they are incorrect. The error is a morphological one, an overapplication of the 1st conjugation imperfect indicative suffix *-ba* to 2nd conjugation verbs. Among the suffixation errors, we also found 20 cases in which the children produced an irregular stem/root form paired with a regular inflectional ending, e.g. **pusí* (corr.: *puse* 'I put-past'), **trajé* (corr.: *traje* 'brought'). Again, there is no difference between the incorrect and the target forms with respect to 'phonological transparency', and the errors can better be explained in morphological terms, i.e. as overapplications of regular inflectional endings.

We conclude that the phonological transparency hypothesis, which does not make reference to the morphological structure of the word forms involved, provides only a partial account of the generalization properties of inflected word forms in Spanish child language.

Developmental aspects

Another set of findings from the present study concerns the frequency, the distribution and the development of overregularization errors. We found a low and relatively steady mean overregularization rate of 2% in the spontaneous speech data, a rate which is in the same range as the 4% to 5% mean overregularization rates for the English past-tense *-ed* (Marcus *et al.*, 1992)[10] and for the German regular *-t* participle (Clahsen & Rothweiler, 1993).

---

[10] An exception seems to be the English-speaking child Abe (Kuczaj, 1977) who has been reported to overregularize *-ed* more frequently than other children. Marcus *et al.* (1992) argued that unusual pragmatic circumstances during Abe's recording sessions may have produced higher than usual overregularization rates.

617

According to the dual mechanism model, overregularizations are prevented by the existence of irregular forms that take precedence over (i.e. block) default forms. Hence overregularization errors are rare overall; they arise in exceptional circumstances only, i.e. when the child fails to access or retrieve an irregular form. We also found a frequency effect in children's errors: low-frequency irregular verb forms yielded more errors than high-frequency ones. This finding is compatible with the view that irregulars are stored in memory, since memory storage and retrieval are likely to be dependent on frequency of exposure. We also found a U-shaped development of over-regularizations; they are absent from the earliest recordings, occur during an intermediate period, and disappear at a later age. The distribution of regular and irregular (stem/root) types does not show any particular increase, when children start to overregularize or shortly before that point, suggesting that the development of overregularizations is not directly linked to the frequency distribution in the child's vocabulary. Instead, we argued that the onset of overregularizations is syntactically triggered, by the requirement to generate a fully specified finite verb form in every sentence, in conjunction with lexical gaps or retrieval failures for irregulars. Overregularizations gradually de-crease over time when children get older and memory traces for irregulars are becoming stronger and the children's ability to retrieve them is becoming more reliable.

CONCLUSION

Our main finding is a dissociation between regular and irregular processes in Spanish-speaking children's productions of inflected verb forms which is evident from both the distribution of errors and the development of overregularizations. These findings are compatible with the difference posited by the dual-mechanism model between rule-based (regular) and memory-based (irregular) representations for morphologically complex words.

REFERENCES

Aveledo, F. (1999). Overregularization in the acquisition of Spanish verb inflection. Un-published MA thesis, University of Essex.
Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes* **10**, 425–55.
Bybee, J. L. & Pardo, E. (1981). On lexical and morphological conditioning of alternations: a nonce-probe experiment with Spanish verbs. *Linguistics* **9**, 937–68.
Clahsen, H. (1999). Lexical entries and rules of language: a multi-disciplinary study of German inflection. *Behavioral and Brain Sciences* **22**, 991–1013.
Clahsen, H., Eisenbeiss, S. & Penke, M. (1996). Lexical learning in early syntactic development. In H. Clahsen (ed.), *Generative perspectives on language acquisition*. Amsterdam: Benjamins.

618

Clahsen, H. & Rothweiler, M. (1993). Inflectional rules in children's grammars: evidence from the development of participles in German. *Yearbook of Morphology* 1992, 1–34.

Clahsen, H., Rothweiler, M., Woest, A. & Marcus, G. (1992). Regular and irregular inflection in the acquisition of German noun plurals. *Cognition* **45**, 225–55.

Corbett, G. & Fraser, N. (1993). Network Morphology: a DATR account of Russian nominal inflection. *Journal of Linguistics* **29**, 113–42.

Díaz, M. (1996). La posición de la frase nominal sujeto respecto del verbo: un estudio infantil del habla caraqueña. Unpublished ms., Caracas: Universidad Central de Venezuela.

Díez-Itza, E. & Pérez-Toral, M. (1996). El desarrollo temprano de funciones discursivas. In M. Pérez-Perreira (ed.), *Estudios sobre la adquisición del castellano, catalán, euskera y gallego*. Santiago de Compostela: Universidad de Santiago de Compostela.

Elman, J., Bates, L., Johnson, M., Karmiloff-Smith, A., Parisi, D. & Plunkett, K. (1996). *Rethinking innateness*. Cambridge, MA: MIT Press.

Fernández, A. (1994). El aprendizaje de los morfemas verbales. Datos de un estudio longitudinal. In S. López-Ornat, A. Fernández, P. Gallo & S. Mariscal (eds), *La adquisición de la lengua española*. Madrid: Siglo XXI.

Guasti, M. T. (1993). Verb syntax in Italian child grammar: finite and non-finite verbs. *Language Acquisition* **3**, 1–40.

Harris, J. (1985). Spanish diphthongization and stress: a paradox resolved. *Phonology Yearbook* **2**, 31–45.

Hernández-Pina, R. (1984). *Teorías psicosociolingüísticas y su aplicación a la adquisición del español como lengua materna*. Madrid: Siglo XXI.

Johnson, C. (1995). Verb errors in the early acquisition of Mexican and Castilian Spanish. In E. Clark (ed.), *The Proceedings of the 27th Annual Child Language Research Forum*. Cambridge: CUP.

Kuczaj, S. A. (1977). The acquisition of regular and irregular past tense forms. *Journal of Verbal Learning and Verbal Behaviour* **16**, 589–600.

Langacker, R. (2000). *Grammar and conceptualization*. Berlin, New York: de Gruyter.

López-Ornat, S. (1994). La metodología de la investigación longitudinal. In S. López-Ornat, A. Fernández, P. Gallo & S. Mariscal (eds), *La adquisición de la lengua española*. Madrid: Siglo XXI.

MacWhinney, B. (1995, 2000). *The CHILDES project*. Hillsdale, NJ: Erlbaum.

Maratsos, M. (2000). More overregularizations after all: new data and discussion on Marcus, Pinker, Ullman, Hollander, Rosen & Xu. *Journal of Child Language* **27**, 183–212.

Marchman, V., Plunkett, K. & Goodman, J. (1997). Overregularization in English plural and past tense inflectional morphology. *Journal of Child Language* **24**, 767–79.

Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. & Xu, F. (1992). Overregularization in language acquisition. *Monographs of the Society for Research in Child Development*, Serial no. 228, Vol. 57.

Matcovich, P. F. (2000). Regular inflection in the mental lexicon: evidence from Italian. In H. G. Simonsen & R. T. Endresen (eds), *A cognitive approach to the verb: morphological and constructional perspectives*. Berlin, New York: de Gruyter.

Montes, R. (1992). Achieving understanding: repair mechanism in mother–child conversations. Unpublished doctoral dissertation: Georgetown University.

Mueller Gathercole, V., Sebastian, E. & Soto, P. (1999). The early acquisition of Spanish verbal morphology; across-the-board or piecemeal knowledge. *The International Journal of Bilingualism* **2/3**, 133–82.

Orsolini, M. (1999). On the cross-linguistic validity of the dual-mechanism model. *Behavioral and Brain Sciences* **22**, 1033–5.

Orsolini, M., Fanari, R. & Bowles, H. (1998). Acquiring regular and irregular inflection in a language with verb classes. *Language and Cognitive Processes* **13**, 425–64.

Pérez-Pereira, M. & Singer, D. (1984). Sobre la adquisición de morfemas del español. *Infancia y Aprendizaje* **27/28**, 205–21.

Pinker, S. (1999). *Words and rules: the ingredients of language*. New York: Basic Books.

Plunkett, K. & Marchman, V. (1996). Learning from a connectionist model of the English past tense. *Cognition* **61**, 299–308.

619

Posada, A. (1997). *Aspectos de la adquisición de la lengua española entre los 27 y los 38 meses de edad.* Unpublished ms., Universidad Simón Bolívar, Caracas.

Radford, A. & Ploennig-Pacheco, I. (1995). The morphosyntax of subjects and verbs in child Spanish: a case study. *Essex Research Reports in Linguistics* **5**, 23–67.

Ragnarsdóttir, H., Simonsen, H. G. & Plunkett, K. (1999). The acquisition of past tense morphology in Icelandic and Norwegian children. An experimental study. *Journal of Child Language* **26**, 577–618.

Real Academia de la Lengua Española. (1995). *Diccionario de la lengua española.* (21st edition). Madrid: Espasa Calpe.

Rizzi, L. (1993). Some notes on linguistic theory and language development: the case of root infinitives. *Language Acquisition* **3**, 371–93.

Say, T. & Clahsen, H. (2001). Words, rules and stems in the Italian mental lexicon. In S. Nooteboom, F. Weerman & F. Wijnen (eds), *Storage and computation in the language faculty.* Dordrecht: Kluwer.

Serrat, E. & Aparici, M. (1999). Morphological errors in early language acquisition: evidence from Catalan and Spanish. Unpublished ms., Universities of Girona and Barcelona.

Shiro, M. (1997). Un estudio de la narrativa oral en niños de edad preescolar. In A. Bolivar & P. Bentivoglio (eds), *Actas del Primer Coloquio Latinoamericano de Analistas del Discurso.* Caracas, Universidad Central de Venezuela.

Torrens, V. (1995). The acquisition of the functional category Inflection in Spanish and Catalan. *MIT Working Papers in Linguistics* **26**, 451–72.

Wexler, K. (1994). Optional infinitives, head movement and the economy of derivations in child grammar. In D. Lightfoot & N. Hornstein (eds), *Verb movement.* Cambridge: CUP.

Wunderlich, D. (1996). Minimalist Morphology: the role of paradigms. *Yearbook of Morphology* 1995, 93–114.

APPENDIX

A. OVERVIEW OF DATA

Child	Age	Source	No. of samples	Sampling frequency	Sample size
Longitudinal data					
María	1;7–4;0	Spain-Madrid, López Ornat (1994)	22	Every 2 weeks	22 hours
Koki	1;7–2;11	México, Montes (1992)	13	Monthly	6·40 hours
Pablo	2;3–3;2	Venezuela, Posada (1997)	12	Monthly	12 hours
Idaira	2;7–4;7	Canary Islands, Marrero & Albalá; see MacWhinney (1995)	6	Every 4–5 months	–
Cross-sectional samples					
Solange	4;2	Venezuela, Díaz (1996)	1	1 sample	20 (mts.)
Rosmary	3;10	Venezuela, Shiro (1997)	1	1 sample	20 (mts.)
Child 1	3;2				
Child 5	3;7				
Child 6	3;0				
Child 7	3;0	Spain – Oviedo, Díez-Itza & Pérez-Toral (1996)	9	1 sample (each)	6·45 hours (45 mts. per sample)
Child 10	3;4				
Child 11	3;6				
Child 12	3;7				
Child 13	3;9				
Child 15	3;10				

620

B. TYPES OF ERRORS PRODUCED BY EACH CHILD

(correct target forms are shown in brackets)

María:	teno (tengo), tieno (tengo), sabo (sé), salo (salgo), venió (vino), devolvido (devuelto), hace (haz), pono (pongo)
Idaira:	ponié (puse), poniste (pusiste), rompida (rota), ponió (puso), queriba (quería), sale (sal), teno (tengo), venió (vino), trayó (trajo), pone (pon)
Pablo:	vinió (vino), ponió (puso), pone (pon), ensucí (ensucié), ponó (puso), poní (puse), hicié (hice), pusió (puso), sabo (sé), hació (hizo), trajé (traje), trají (traje), hicí (hice), pusí (puse), teniba (tenía), poniste (pusiste)
Koki:	cayí (caí), sale (sal), sabo (sé), poní (puse), hace (haz), poniste (pusiste), oyo (oigo), haciste (hiciste), pusí (puse), ponimos (pusimos), parezo (parezco)
Solange:	ponieron (pusieron)
Rosmary:	conozo (conozco), punieron (pusieron), hacido (hecho)
Child 1:	sabo (sé), pone (pon), pompida (rota), poniéramos (pusiéramos), poní (puse)
Child 5:	tení (tuve), tuví (tuve), tuvió (tuvo), hació (hizo), rompido (roto)
Child 7:	trayó (trajo), ponió (puso), creiba (creía)
Child 10:	hace (haz)
Child 11:	vinió (vino)
Child 12:	punió (puso)
Child 13:	podió (pudo)

C. OVERREGULARIZATION RATES

C-1. Overregularization rates for individual children

	Irregular verbs (tokens)	Overregularizations (tokens)	Overregularization rate
Longitudinal data			
María	1720	26	0·015
Koki	1091	25	0·022
Idaira	384	13	0·033
Cross-sectional data			
Solange	22	1	0·043
Rosmary	72	4	0·053
Child 1	48	7	0·127
Child 5	29	9	0·237
Child 7	18	4	0·182
Child 10	3	1	0·250
Child 11	10	1	0·091
Child 12	21	2	0·087
Child 13	28	1	0·034

621

C-2. Overregularization rates in present and past tense in the longitudinal samples

	Present tense	Past tense
María	0·014	0·067
Koki	0·017	0·169
Idaira	0·015	0·175

C-3. Error rates for verb forms requiring diphthongized stems

	Type of alternation	Error rates
María	o ~ ue	0·113
	u ~ ue	0·867
	e ~ ie	0·138
Koki	o ~ ue	0·226
	u ~ ue	0·000
	e ~ ie	0·124
Idaira	o ~ ue	0·000
	u ~ ue	0·000
	e ~ ie	0·167

622