Verb Movement Phenomena in Spanish: "Mixed Languages" and Bilingualism

Dalila Ayoun
University of Arizona

1. Introduction

The verb movement parameter was initially proposed following early work by Emonds (1978) and Pollock (1989, 1997a, 1997b) in French, and Torrego (1984) as well as Suñer (1994) in Spanish.[1] It has been extensively studied in theoretical linguistics from a generative perspective to a minimalist perspective (Chomsky, 1995; Pollock, 1997a, b). And it has been investigated in applied linguistics in first language (L1) acquisition (Deprez, 1994; Deprez & Pierce, 1990, 1993; Meisel, 1992; Pierce, 1992; Verrips & Weissenborn, 1992; Weissenborn, 1988, 1992; Weissenborn, Verrips & Berman, 1989), as well as in second language (L2) acquisition (Ayoun, 2003, 1999a; Antes, Moritz & Roebuck, 1995; Downey-Vanover, 1994; Hawkins, Towell & Bazergui *al*., 1993; Herschensöhn, 1998; Hulk, 1991; Mandell, 1998; Trahey & White, 1993; White, 1991, 1992). However, most of the L2 studies have focused on Francophone learners of English or Anglophone learners of French. Few studies have investigated the case of L2 Spanish learners, the focus of the present empirical study. The studies by Antes *et al*. (1995) and Mandell (1998) are two notable exceptions, but both suffer from methodological shortcomings, as we will see below. This lack of studies with L2 Spanish learners is unfortunate because Anglophone learners of Spanish as an L2 present a particularly interesting case from a learnability perspective due to the status of Spanish as a "mixed language" with respect to the verb movement parameter (Ayoun, 1999b). A "mixed" or "hybrid" language is defined as a language which does not exhibit strictly mutually exclusive parameter settings. Instead, these languages exhibit partially co-existing settings in that they allow both parametric settings for the different structures subsumed under a given parameter. For instance, Hebrew licenses null subjects with the first and second persons (but not the third person), while requiring overt expletives, a property of non-null subject languages such as English (Berman, 1990; Shlonsky, 1997). It is in this sense that Muysken's (1998) definition of bilingualism as the simultaneous operation of two grammars, as opposed to a single grammar being active in "true" bilinguals, can be extended to the grammars of monolingual adult speakers of "mixed" languages.

I will first review the facts of the verb movement parameter in Spanish to argue that this language is best characterized as a mixed or hybrid language with regard to the verb movement parameter, and will consider the learnability implications for adult L2 learners. After a brief review of previous studies, the various computerized elicitation tasks and stimuli of the present study will be described along with the three hypotheses to be tested.

2. Verb movement in Spanish

Given an underlying structure in which the verb follows adverbs and negatives as shown in (1):

(1) $[_{IP}$ NP I ($[_{Neg}$ *not/no*]) $[_{VP}$ (Adv) V...]]

a surface structure in which the verb precedes adverbs or negation, or both, indicates that it has moved out of its initial position by raising to the left (Pollock, 1989, 1997).

It is generally assumed that it is the visibility of "rich" morphological features at LF that triggers overt verb movement to Infl before Spell-Out in order to check and erase these features which would otherwise violate the Full Interpretation Principle (Chomsky, 1995) as formalized in (2):

 ISB4: Proceedings of the 4th International Symposium on Bilingualism, ed. James Cohen, Kara T. McAlister, Kellie Rolstad, and Jeff MacSwan, 143-162. Somerville, MA: Cascadilla Press.

(2) A strong [+finite] verb moves to Infl before Spell-Out to Check and Erase its Infl features.

It follows that verb movement cannot apply in the case of an English verb because its verbal paradigm is morphologically poor, lacking features of person and number, contrary to morphologically rich languages such as French or Spanish. However, we will see below that English verbs do undergo some form of movement. We will also see how difficult it is to define rich morphology, let alone as a potential trigger to verb movement.

Following Pollock (1989) and others, I will assume that the verb movement parameter subsumes these syntactic properties: negation placement and adverb placement in both finite and nonfinite contexts, subject floating quantifiers and inverted questions. Let us examine each of these properties. We will see that although Spanish is traditionally classified as a verb movement language, it actually exhibits several properties without lexical verb movement, starting with negation placement.

2.1. Negation placement

The following examples show that Spanish and English lexical verbs and auxiliaries remain *in situ* in both finite and non–finite contexts since they follow the negation *not/no*:

(3) a. Martha does not buy the tapes.
    b. Marta no compra las casetas.
    c. *Martha buys not the tapes.
    d. *Marta compra no las casetas.
(4) a. To not play is strange.
    b. No jugar es extraño.
    c. *To play not is strange.
    d. *Jugar no, es extraño.
(5) a. Not to be afraid is rare.
    b. No tener miedo es raro.
    c. *To be not afraid is rare.
    d. (*)Tener no miedo es raro.

The example in (5d) is acceptable only if the negation has scope over the adjective instead of the auxiliary. Thus neither English or Spanish allow verb movement with respect to negation. It is consistently the case with other negation elements as in the following examples:

(6) a. Not to invite anyone is selfish.
    b. No invitar a nadie es egoista
    c. *To invite not anyone is selfish
    d. *Invitar no a nadie es egoista
(7) a. John does not see anyone / has not seen anyone.
    b. Juan no ve a nadie/ no ha visto a nadie.
(8) a. My friends do not read anything / have not read anything.
    b. Mis amigos no leen nada / no han leído nada.

Thus English and Spanish lexical verbs still remain *in situ* with respect to negation. However, this lack of verb movement in Spanish is only apparent: since NegP is the highest projection, it always dominates IP, thus even if lexical verbs systematically raise to Infl as it is often assumed (*e.g.*, Contreras, 1991), they always follow negation on surface structures (*e.g.*, Olarrea, 1997).

2.2. Adverb placement

In finite contexts, verb movement is optional in Spanish as illustrated in (9b, d).

(9) a. John always reads books.
b. Juan siempre lee libros.
a'. *John reads often novels.
b'. Juan lee siempre libros.

On the other hand, nonfinite verbs do move past adverbs in Spanish, as shown in (10b, d), contrary to theoretical predictions according to which the morphological features of these verbs are too weak to trigger movement.

(10) a. To often go out is fun.
b. Salir a menudo is divertido
c. ?To go out often is fun.
d. A menudo salir es divertido

However, nonfinite verb movement appears to depend on the adverb as illustrated by the example in (11b) where verb movement past the adverb *apenas*, 'barely', yields an ungrammatical structure:

(11) a. To barely sleep is strange.
b. Apenas dormir es extraño.
c.*To sleep barely is strange.
d. *Dormir apenas es extraño.

Spanish thus appears to allow both structures with and without verb raising in finite and nonfinite contexts.

2.3. Floating quantifiers

Spanish partly differs from English with regard to subject floating quantifiers as exemplified by the following sentences :

(12) a. All my friends love the beach.
b. Todos mis amigos quieren la playa.
(13) a. My friends all love the beach.
b. Mis amigos todos quieren la playa.
(14) a. *My friends love all the beach.
b. Mis amigos quieren todos la playa.
(15) a. *My friends love the beach all.
b. Mis amigos quieren la playa todos.

The examples in (12a, b), grammatical in both languages, show a non-floating quantifier modifying the subject, while the remaining examples illustrate a floating subject quantifier. All the instances of floating quantifiers are acceptable in Spanish while only (13a) is allowed in English. The ungrammaticality of (14a) is explained as follows, according to Pollock (1989) based on Sportiche's (1988) analysis. Let us consider the structure in (16):

(16) $[_{Infl}$ $[_{V}$ [all + my friends]$[_{V}[_{V}$ love] the beach]]]

As seen above, lexical verbs such as *love* are morphologically poor and do not raise to Infl. Consequently, the quantifier all ends up preceding instead of following the verb, while the corresponding Spanish structure in (14b) is ungrammatical. Note that an alternative analysis suggests

that it is the subject that moves out of its initial position, leaving the floating quantifier stranded (Olarrea, 2001, personal communication). Whatever the case may be, Spanish allows both types of structures, with and without floating quantifiers, whereas English does not.

2.4. Inverted questions

Let us examine the following examples which illustrate inversions with verb and subject pronouns as well as verbs and nominal subjects:

(16) a. Does she want this book?
b. *Wants she this book?
c. ¿Quiere (ella) este libro?
(17) a. Isn't she /is she sick?
b. ¿No está enferma?
(18 a. Does Mary want this book?
b. ¿Quiere María este libro?
c. *Want Mary this book?
d. ¿María quiere este libro?

English requires *do*-support to form inverted questions with a lexical verb-subject order (16a) which are grammatical in Spanish (16b), but *have/be* raising is perfectly acceptable (17a)(see Roberts, 1998, for a detailed account) The example in (17b) shows that the Spanish auxiliaries do not raise past the negation in inverted questions. And finally, the examples in (18) indicate that Spanish allows inverted questions with a nominal subject, while English remains consistent in requiring *do* support. *Have/be* raising also occurs with adverb, negation and floating quantifiers as in (19):

(19) a. My parents are often taking trips.
b. Quentin has not spoken to her.
c. My friends have all read books.

To sum up, it is clear that the interlinguistic differences depend on whether overt verb movement to Infl is allowed or disallowed. However, verb movement does not systematically apply to [+finite] lexical verbs with strong morphological features: both English and Spanish allow structures both with and without verb movement, only in the case of [+finite] contexts for English auxiliaries, but in both [+finite] and [–finite] contexts with respect to adverb placement for Spanish lexical verbs.[2]

English does permit two types of verb movement: first, auxiliary movement known as *have/be* raising (*e.g.*, Roberts, 1998), or V-to-I raising with a more limited scope (Battye & Roberts, 1995); second, short verb movement in so-called quotative inversion (Collins & Branigan, 1997) as illustrated in (20a, b):

(20) a. "Keep quiet", urged her attorney.
b. "Let's all go together", suggested Lisa.

These sentences are defined as "sentences with direct speech complements ('quote') to verbs of saying, thinking and writing" which permit an inversion of the subject and the verb. Or technically speaking: "the subject DP remains in Spec-V [...] and the verb raises past to Agr_o" (Collins & Branigan, 1997, p. 1, 2). However, this type of inversion does not change the fact that English lexical verbs have weak features as noted by Collins & Branigan (1997): "The verb has the same features in quotative inversion as it has in any other construction – its V-features are all weak, as are the V-features of Agr_O. The verb raises solely because it must check the Case features of a Spec-Agr_o trace, so that later movement of the subject will be allowed" (p. 38).

Putting aside the issue of *have/be* raising, we can conclude that verb movement phenomena in Spanish does not fall neatly into the category of [+mvt] or [–mvt] as summarized in Table 1.

Table 1: Verb Movement Properties in English and Spanish

	English	Spanish
Negation [–finite]/[+finite]	[–mvt]	[–mvt]
Adverb [–finite]/[+finite]	[–mvt]	[–mvt]/[+mvt]
Floating Quantifier	[–mvt]	[–mvt]/[+mvt]
Pronominal Inversion	[–mvt]	[–mvt]/[+mvt]

English is indeed a [–mvt] language since it strictly follows the principle in (2) replicated here in (21):

(21) A strong [+finite] verb moves to Infl before Spell-Out to Check and Erase its Infl features.

English verbs are morphologically weak or poor and therefore do not raise. However, in spite of their morphological strength or richness, Spanish verbs do not necessarily raise out of their initial position. As indicated by the "[–mvt]/[+mvt]" notation in Table 2, verb movement with respect to adverb placement in (non)finite contexts, pronominal inversion and floating quantifiers, is only optional. Furthermore, verb movement is obscured with respect to negation: whether the verb moves to Infl or not, it will not raise past negation resulting in an apparently [–mvt] surface structure, as indicated in Table 1. Spanish is thus more accurately characterized as a mixed language than as a strictly verb movement language, in the sense that it instantiates surface structures both with and without movement. What are the learnability implications of the verb movement parameter for adult Anglophone learners of Spanish as an L2?

3. Learnability implications

Let us consider what these learners will need to do from a parameter (re)-setting perspective as presented in Table 2.

Table 2: Parameter (re)-settings from L1 English to L2 Spanish

Properties	L1 English to L2 Spanish
Negation [–finite]/[+finite]	no apparent re-setting: L1 and L2 are [–mvt]
Adverb [–finite]/[+finite]	re-setting from [–mvt] → [+mvt] but [–mvt] structures are also possible
Floating Quantifier	re-setting from [–mvt] → [+mvt] but [–mvt] structures are also possible
Pronominal Inversion	re-setting from [–mvt] → [+mvt] but [–mvt] structures are also possible

Table 2 indicates that L2 Spanish learners face an interesting task: parameter re-setting is not simply a matter of re-setting the verb movement parameter from an English [–mvt] setting to the Spanish [+mvt] setting because Spanish instantiates structures representative of both settings for three out of the four properties. In other words, Spanish is a mixed language, and English and Spanish are in a subset/superset relation. What does that mean for L2 Spanish learners?

First, experimental evidence shows that it is not necessarily easier to go from an L1 as a subset to an L2 as a superset, especially since the Subset Principle may or may not be operative in adult L2 learning (*e.g.*, Ayoun, 1996; Berent, 1994; Fodor, 1994; MacLaughin, 1995; Yuang, 1995).

Second, it is generally accepted that a match in parameter setting between the L1 to L2 may facilitate L2 acquisition, whereas a mismatch may hinder it and/or result in a delay. Here we have a mismatch, as well as an apparent overlap of values. The mismatch comes from the fact that English is a [–mvt] language, while Spanish is primarily a [+mvt] language. The overlap comes from the co-existence of surface structures with movement along with surface structures without verb movement, since verb movement is optional for floating quantifiers, pronominal inversion, and certain adverbs in nonfinite contexts. Moreover, although verbs do raise, they do not raise past negation. Consequently,

re-setting does not appear to be required for any of the properties subsumed under the verb movement parameter.

These factors lead us to this crucial question: If the placement of tensed thematic verbs with respect to negation functions as a trigger for verb movement (Ayoun, 1999b; Deprez, 1994; Pierce, 1992), how will L2 learners know that Spanish is predominantly a verb movement language since the verb raises but not past the highest projection, negation? This is at best an extremely weak or 'ambiguous' trigger (Fodor, 1998, 1999) that does not appear to be viable for Spanish. Are L2 learners able to fall back on morphology as an alternative trigger since Spanish verbal inflectional endings are particularly rich? That is possible, even if it has been difficult to come up with a viable cross-linguistic account *(*see *e.g.,* Rohrbacher, 1994, 1999; Vikner, 1995, 1997).

To sum up, the acquisition of verb movement phenomena by English L2 Spanish learners will be a difficult task due to: 1) ambiguous input; 2) questionable triggering input; and 3) Spanish as a mixed, superset language with a mismatch of values between the L1 and the L2.

Consequently, it would be unreasonable to expect the traditional immediate clustering of all the properties subsumed under the verb movement parameter as evidence of successful parameter-(re)setting in adult L2 acquisition. Instead, and following others (Ayoun, 1999a; Boe, 1996; Bolotin, 1996a, b, c; Flynn, 1989; Herschensohn, 2000; Uziel, 1993), I will argue that parameter re-setting in adult L2 acquisition is a gradual and progressive process as the learners work out the deductive consequences of the L2 setting, or settings, as in the case of a mixed language like Spanish. The mismatch and overlap in the L1 and L2 settings are bound to slow down L2 acquisition, resulting in a delay of the manifestation of parametric effects. A gradual and partial clustering of properties, as opposed to an immediate and complete clustering of properties, is therefore to be expected.

Thus the following three predictions will be tested: i) properties will partially cluster; ii) learners will initially entertain both settings; iii) the learners' performance will vary from task to task.

The first prediction assumes that adult L2 acquisition may rely on parameter-(re)setting as a deductive process. Evidence of clustering will be taken to be a non-significant difference in the learners' performance on the various properties subsumed under the verb movement parameter. The second prediction takes into account the ambiguity of the triggering input of Spanish as a mixed language, and the mismatch and overlap of parameter settings in English as an L1 and Spanish as an L2. Earlier findings with L2 French learners revealed that learners initially entertain both settings of a parameter (Trahey, 1992; White, 1991a, 1991b, 1992). In the case of Spanish as an L2, it is exactly what learners should do: adopt the [+mvt] setting in some cases but not in others. The last prediction is based upon earlier findings with L2 French learners (Ayoun, 1999a, 2000) which showed that the lower the level of proficiency, the stronger the task effect. Since the classroom learners who participated in this study were only enrolled in second or third year courses, a similar task effect is likely to be found.

4. Previous studies

Antes *et al.* (1995) compared Spanish L2 learners with French L2 learners. Following the results of a cloze test which was used as a pre-test, participants were placed in three proficiency levels. They were given what the researchers refer to as a Sentence Acceptability Task, but it may be more appropriately qualified as a preference task because it required the participants to select sentences presented in pairs following a lead (contextualizing) sentence. This study is limited by at least two methodological shortcomings: the small number of stimuli (4 to 6 pairs of sentences), and the fact that only one elicitation task was used which is highly problematic in the light of recent results with both French and Spanish L2 learners that clearly revealed a strong task effect (Ayoun, 1999b, 2000). Moreover, because only two properties were tested in French (adverb and negation placement), and three in Spanish (nominal subject inversion, adverb and negation placement), no conclusion can be reached with regard to the re-setting of the entire parameter.

The French learners performed extremely well on negation (from 96% to 100%) but only moderately well on adverb placement (from 60% to 78%). L2 French learners performed significantly better than L2 Spanish learners on adverb (from 46% to 67%) and negation (from 56% to 66%) placement, accepting more instances of verb movement. The L2 Spanish learners performed poorly on

pronominal inversion, barely reaching the 50% mark, but so did the native speakers. Antes *et al.* attribute this surprising performance to pragmatics. However, the limited number of stimuli does not allow for a definite conclusion. Antes *et al.* also suggest that more readily available and less ambiguous positive evidence can account for the fact that the L2 French learners outperformed the L2 Spanish learners.

Mandell (1998) tested the acquisition of three verb-movement properties: adverb placement, inversion in yes/no questions, thematic *wh*-question without inversion as exemplified in (21):

(21) a. ¿Por qué Maria tiene más tiempo que nosotros?
Why does Maria have more time than us?
b. *¿Qué Marina compra?
What does Marina buy?

The participants were English native speakers enrolled in a Spanish college-level class. They were in second-semester (n = 48), fourth-semester (n = 24), sixth-semester (n = 19), or fifth year and above (n = 15) classes. The elicitation tasks consisted in a grammaticality judgment task (GJT), delivered with a pre-recorded, timed audio-tape, and a dehydrated sentence task which consisted in asking the participants to combine given elements to form an acceptable sentence. For the GJT, the participants were simply asked to indicate whether a sentence was "possible" or "impossible", and to provide a correction for the sentences rejected as "impossible". The results showed a gradual improvement with the proficiency level, and they also indicated that, in general, the participants performed better on the yes/no questions and *wh*-questions than on adverb placement. The results of the dehydrated sentence task confirmed this trend, but with much lower means on adverb placement. For instance, the highest level of proficiency (fifth year and above) which obtained a mean of 8.33 on the GJT, scored only 2.33 on the dehydrated sentence task. However, as discussed above, verb movement is optional in finite contexts with respect to adverbs. All the test items were given in the order corresponding to the [+mvt] setting. In other words, all the adverbs preceded the verbs. The participants may have thought that the task could not be as simple as reproducing the given elements. In any case, it is unlikely that they had time to do anything but rewrite these elements in the mere fifteen seconds allotted per sentence. The GJT was administered under similar time constraints, in addition to the audio mode which may have added some unnecessary difficulty for the learners in the lower levels of proficiency. Finally, since this study did not test negation placement or floating quantifiers, no conclusions regarding parameter (re)-setting as a whole can be drawn.

5. The study

The present study differs from the two studies reviewed above in that it uses a wide variety of computerized elicitation tasks, and in that it tests all the syntactic properties subsumed under the verb movement parameter.

5.1. Participants

There were 11 females and 4 males who averaged 21.8 year old (range = 18 to 38, with the majority of the participants between the ages of 20 and 22).[3] They were majoring in a variety of disciplines in a North American university where they were enrolled in second or third year Spanish classes. Most of the participants (n = 12) started learning Spanish between the ages of 14 and 16; the graduate student was 32, and only two of the participants started learning Spanish as children (age 5), because one of their parents spoke it at home.

The students received extra credit on homework assignments in their respective classes in exchange for their participation. They went to a computer laboratory three times during the same week for three sessions. During the first session, they filled out a background information questionnaire to provide the usual information such as age, gender, native language, second language, and self-assessed language proficiency level. They also took the pre-test (Task 1), a scalar grammaticality judgment task which included 44 items to test a variety of properties or structures: the *ser/estar* and

por/para distinctions, and the properties subsumed under the Oblique-Case parameter (Kayne 1981, 1984; Ayoun 1995). This pre-test was designed to obtain an independent measure of the participants' proficiency level. Then the participants performed the first two experimental tasks during the second session, and the last two tasks during the third and last session. The researcher was present to answer questions participants may have.

5.2. Elicitation tasks

The participants performed the elicitation tasks in the same order in which they are presented here. They consisted in a scaled grammaticality judgment task (S-GJT), a preference/grammaticality task (PrefG), a production task (ProdT) and a magnitude estimation acceptability judgment task (ME-AJT).

All the elicitation tasks were computerized. They were designed with Claris Homepage® to be uploaded on a server so that they could be accessed on any computer equipped with a Web browser. The procedure was identical for each task: the participants entered their first and last names (which were later coded to preserve their anonymity), completed the tasks one at a time and clicked on SUBMIT at the bottom of the form to send the data to their individual folders on the server.

Computerized data collection presents several distinct advantages over more traditional non-computerized tasks. For instance, there are less to no human error in collecting and recording the data which can be more easily organized and analyzed. There is no variation or difference in the way the participants are exposed to the stimuli or the treatment (Hulstijn, 1997). It avoids the common and frustrating pitfall of missing data because the tasks can be designed to require the participants to enter a response before submitting the form. In the present study, if participants attempted to submit an incomplete form, they automatically received a message asking them to click on 'Back' on their browser to return and complete the task by providing the missing information. When the task was properly completed, participants received a positive confirmation message. Computerized data collection thus permits a better control of experimental variables in general. The data are then more likely to better inform L2 acquisition processes.

5.2.1 Preference/grammaticality task (PrefG,Task 2)

The participants were presented with pairs of sentences. Below the stimuli were two pop-up menus: the first one asked the participants to indicate which of the two sentences, A or B, they preferred; while the second pop-up menu permitted the participants to specify whether the other sentence, *i.e.*, the one they had not selected, was grammatical or ungrammatical. This latter feature is unique to the present study and makes this preference task more informative than traditional preference tasks which typically ask participants to indicate which sentence they prefer without asking them to indicate the (un)grammaticality of the other sentence.

5.2.2 Scaled grammaticality judgment/correction task (S-GJT,pre-test (Task 1)and Task 3)

The participants read and judged individual sentences according to the following categories: 'completely ungrammatical', 'ungrammatical', 'I don't know', 'grammatical', and 'completely grammatical'. The categories were presented in pop-up menus and were coded from 1 to 5 respectively for statistical analysis. The codings were invisible to the participants. The likert scale was chosen over a numeric scale because numbers can be perceived differently or assigned different degrees of (un)grammaticality by participants. In addition, participants were asked to provide a correction to the sentences they judged to be either 'completely ungrammatical' or 'ungrammatical' to ensure that the proper syntactic criterion motivated the rejection of these stimuli.

5.2.3 Production task (Prod, Task 4)

Participants were asked to translate sentences from their L1 (English) into Spanish. They typed in their answers in a text field below the stimulus sentence. All text fields were specified as required to

avoid missing data. Participants were awarded one point for each correctly produced sentence: only the main syntactic criterion was considered. Minor stylistic and agreement mistakes for instance were ignored.

5.2.4 Magnitude estimation acceptability judgment task (ME-AJT,Task 5)

This elicitation task is the most innovative both in terms of its design and for L2 acquisition studies. It was inspired by Sorace (1996) and Bard, Robertson & Sorace (1996) who applied magnitude estimation from the field of psychophysics to linguistic judgments. Magnitude estimation consists in asking participants to judge or rate stimuli in relation to one another. For example, participants may be "presented with a series of stimuli of unequal magnitudes, one at a time in random order, and are asked to assign a number (the modulus) to the perceived magnitude of the first stimulus (the standard), and then successive numbers to the perceived magnitude of stimuli in proportion to the modulus (Sorace, 1996, p. 401). In the present study, the stimuli consisted of the individual sentences that the participants were asked to read in order to make two decisions: first, whether the sentence was grammatical or ungrammatical; second, how much more or less (un)grammatical a stimulus was compared to the last stimulus judged to also be (un)grammatical.

Task 5 appeared as follows: there were five sentences per screen with a long blue rectangular box above each stimulus with a cursor in the middle; the portion to the right of the cursor was devoted to "grammaticality" and the portion to the left of the cursor to "ungrammaticality", as indicated above the box. The two decisions that the participants made were carried out by moving the cursor within the box with the mouse. Moving the cursor blackened the area covered so that participants could clearly see how (un)grammatical they were indicating each stimulus to be. Invisible to the participants was a scale from +1 to +100 for grammaticality, and –1 to –100 for ungrammaticality. The cursor could be moved back and forth as desired, and the numerical value was recorded when the cursor was released.

Stimuli

The experimental tasks (Task 2 through Task 5) tested the following properties: negation placement in finite contexts (NegFin) and nonfinite contexts (NegInf); adverb placement in finite contexts (AdvFin) and nonfinite contexts (AdvInf); inverted questions (ProInv); and floating quantifier (FQ). All stimuli were controlled for length and simplicity of vocabulary. No distractors were deemed necessary due to the variety of the syntactic properties subsumed under the verb movement parameter.

Task 3, task 4 and task 5 tested all properties with 24, 26 and 21 sentences, respectively. Task 2 was comprised of 20 pairs of sentences, with one grammatical sentence and its ungrammatical counterpart. Both sentences were grammatical for the properties which allow two different structures such as FQ, ProInv, AdvFin. Tasks 2, 3 and 5 thus do not have ungrammatical stimuli for these properties. Furthermore, since movement past adverbs in nonfinite contexts depends on the type of adverb, there were both grammatical and ungrammatical instances of AdvInf. No attempt was made to perfectly balance the number of grammatical and ungrammatical stimuli to avoid a disproportionately greater number of tokens with certain structures. These structures would have been more salient than the others , thereby introducing a distributional bias.

6. Results

6.1. Analysis of data

The raw data were computed in the following manner. In the S-GJT (task 3) and ME-AJT (task 5), participants were awarded one point per correct answer: namely one point for each correctly accepted sentence (*i.e.*, a grammatical sentence accepted as 'grammatical') and one point for each correctly rejected sentence (*i.e.*, an ungrammatical sentence rejected as 'ungrammatical'). The number of correct answers for grammatical sentences was divided by the total number of grammatical sentences for each participant and multiplied by 100 to obtain a percentage accuracy score (*e.g.*, 10 correctly accepted sentences out of 15 grammatical sentences = 66.67% accuracy score). The process was repeated for ungrammatical sentences, thus yielding two accuracy percentage scores for each participant. These two accuracy scores were averaged to obtain an overall accuracy score for each participant (*e.g.*, 66.67% for grammatical sentences + 82.42% for ungrammatical sentences = overall accuracy score of 74.54%).

The scores for the PrefG (task 2) reflect the participants' ability to make two different decisions correctly: first, choosing the correct sentence from a pair of stimuli and second, correctly accepting (if grammatical), or correctly rejecting (if ungrammatical) the other sentence. A point was given for each correct answer. The points were totalled and divided by the total number of sentences to obtain an overall accuracy percentage score. More details are given below with the results for this particular task.

Finally, participants were awarded one point per sentence for correctly producing the relevant structure in the ProdT. Vocabulary errors or gender agreement mistakes were disregarded because they may be due to performance error, not competence, and are found even in native speaker elicited data or spontaneous productions. The total number of points was divided by the total number of sentences to obtain an overall accuracy percentage score (*e.g.*, 20 correctly produced sentences out of 35 yields a score of 57.14%).

6.2 Overall results

The results of the pre-test (task 1) show that the participants averaged 74.7% accuracy on a scaled GJT. They may thus be characterized as being at a low-intermediate level of proficiency. Then the data obtained from the experimental tasks 2 through 5 were analyzed with repeated measures ANOVAs run with SPSS. The results are presented in Table 3.

Table 3: ANOVA on Property and Task

	SS	*df*	*MS*	*F*	Pr >*F*
Property	23054.35	5	4610.87	13.54	.001
Error (property)	23837.05	70	340.53		
Task	3558.24	2	1779.12	4.42	.02
Error (task)	11277.91	28	402.78		
Property by Task	16519.97	10	1652.00	3.55	.001
Error (prop by task)	65121.20	140	465.15		

Table 3 reveals the following results: first, there is a significant property effect (F (5,70) = 14.70, p > .001), which indicates that the participants' performance varied with the type of property. There is also a significant task effect (F (2,28) = 6.96, p = .004). Finally, there is a significant property-by-task effect, which means that the differences among the tasks depend on specific properties (F (10,140) = 5.08, p > .001).

The next table, Table 4, presents the accuracy means for each property within each one of the four tasks, along with the standard deviations, as well as the minimum and maximum accuracy means obtained in each case. A brief overview allows us to observe that the learners' performance noticeably varied from property to property as well as from task to task.

As a group and across all tasks, the learners obtained mixed results. They performed well on NegFin (average = 82.58%), NegInf (average = 81.16%) and ProInv (average = 88.29%), but not so well on AdvFin (average = 77.00%), and rather poorly on FQ, although the ProdT average (task 4) of 93.33% is much higher than the overall average of 61.95%. The learners' worst performance is on AdvInf, the placement of adverbs in nonfinite contexts (average = 56.66%) across all tasks, since their best performance of 66.67% is barely above chance level, indicating that they might as well have been guessing.

We may then initially conclude that, as a group, our L2 learners have acquired just half of these syntactic properties: the placement of negation in finite and nonfinite contexts, the placement of adverbs in finite contexts but not nonfinite contexts, and pronominal inversion. However, the standard deviations, as well as the minimum and maximum means, reveal large gaps, and therefore important individual differences, which will be examined below.

Table 4: Accuracy means by Property and Task

Property	Mean	*SD*	Minimum	Maximum
NegFin	82.58%			
Task 2	68.33%	24.03	25.0%	100.0%
Task 3	83.33%	20.41	25.0%	100.0%
Task 4	98.67%	5.16	80.0%	100.0%
Task 5	80.00%	19.36	50.0%	100.0%
NegInf	81.16%			
Task 2	75.67%	24.12	40.0%	100.0%
Task 3	95.00%	10.35	75.0%	100.0%
Task 4	70.67%	32.83	20.0%	100.0%
Task 5	83.33%	18.09	50.0%	100.0%
AdvFin	77.00%			
Task 3	69.33%	16.68	40.0%	100.0%
Task 4	81.67%	17.59	50.0%	100.0%
Task 5	80.00%	25.35	25.0%	100.0%
AdvInf	56.66%			
Task 2	41.67%	20.41	0.0%	75.0%
Task 3	53.33%	20.85	25.0%	100.0%
Task 4	66.67%	37.80	0.0%	100.0%
Task 5	65.00%	29.58	0.0%	100.0%
FQ	61.95%			
Task 2	42.84%	28.04	0.0%	85.7%
Task 3	65.33%	27.74	20.0%	100.0%
Task 4	93.33%	14.84	50.0%	100.0%
Task 5	46.33%	26.82	0.0%	100.0%
ProInv	89.29%			
Task 3	75.55%	27.37	33.3%	100.0%
Task 4	96.00%	8.28	80.0%	100.0%
Task 5	96.33%	25.82	0.0%	100.0%

6.3 Results by properties

It was hypothesized that the properties subsumed under the verb movement parameter would partially cluster, depending on the consistency of the application of verb movement to Infl. This hypothesis is supported as indicated by the results of a post hoc Tukey analysis (p = .05), displayed in Table 5.

Table 5: Tukey groupings by Property and Task

Subset 1	FQ - T5 46.3% NegInf - T4 70.6%	AdvInf - T3 53.3%	AdvInf - T5 65.0%	FQ - T3 65.3%	AdvInf - T4 66.7%	AdvFin - T3 69.3%
Subset 2	AdvInf - T3 53.3% ProInv - T3 75.6%	AdvInf - T5 65.0% NegInf - T5 80.0%	FQ - T3 63.3% AdvFin - T5 80.0%	AdvInf-T4 66.7% AdvFin - T4 81.7%	AdvFin-T3 69.3%	NegInf - T4 70.7%
Subset 3	AdvInf - T5 65.0% NegFin - T5 80.0% ProInv - T5 93.3%	FQ - T3 65.3% AdvFin - T5 80.0%	AdvInf-T4 66.7% AdvFin - T4 81.7%	AdvFin-T3 69.3% NegFin - T3 83.3%	NegInf - T4 70.7% NegInf - T5 83.3%	ProInv - T3 75.6% FQ - T4 93.3%
Subset 4	AdvInf -T4 66.7% AdvFin - T4 81.7%	AdvFin-T3 69.3% NegFin - T3 83.3%	NegInf - T4 70.7% NegInf - T5 83.3%	ProInv - T3 75.6% FQ - T4 93.3%	NegFin - T5 80.0% ProInv - T5 93.3%	AdvFin - T5 80.0% NegInf - T3 95.0%
Subset 5	AdvFin -T3 69.3% NegFin - T3 83.3%	NegInf - T4 70.7% NegInf - T5 83.3%	ProInv - T3 75.6% FQ - T4 93.3%	NegFin - T5 80.0% ProInv - T5 93.3%	AdvFin - T5 80.0% NegInf - T3 95.0%	AdvFin - T4 81.7% ProInv - T4 96.0%
Subset 6	NegInf - T4 70.7% NegInf - T5 83.3%	ProInv - T3 75.6% FQ - T4 93.3%	NegFin - T5 80.0% ProInv - T5 93.3%	AdvFin - T5 80.0% NegInf - T3 95.0%	AdvFin - T4 81.7% ProInv - T4 96.0%	NegFin - T3 83.3% NegFin - T4 98.7%

The analysis was run with the means for each property within each task because the learners' performance depended on the type of elicitation task, as shown in Table 4. Thus Table 6 reads as follows: for instance, in Subset 1, AdvFin - T3 refers to the group accuracy means on the placement of adverbs in finite contexts in task 3, the scaled GJT. The analysis included all the properties and all the tasks except for task 2 (PrefG), because it did not test all the properties.

Each subset is homogeneous which means that the highest and lowest means are not significantly different. The relatively large number of subsets is due to the large number of properties tested combined with the number of elicitation tasks. However, most of the properties are included in each one of the subsets with a few exceptions: NegFin and ProInv are not part of Subset 1 (the subset with the lowest means); NegFin is not part of Subset 2; and AdvInf is excluded of Subset 5 and Subset 6 (the subsets with the highest means). In other words, we do find partial clustering of the properties tested as hypothesized. Evidence of clustering is taken to be a non-significant difference in the learners' performance on the various properties. There is no significant difference between the properties within each subset. And again, almost all of the properties are represented in the 6 subsets with accuracy means above chance level and as high as 93.3% (ProInv in the ME-AJT) or 98.7% (NegFin in the ProdT).

Post hoc Tukey analyses were also run for each property. These results are presented in Tables 6 through 11.

Table 6: Tukey groupings for AdvFin

Subset 1	AdvFin - T3 69.3%	AdvFin - T5 80.0%	AdvFin - T4 81.7%

Table 7: Tukey groupings for Advinf

Subset 1	AdvInf -T2 41.7%	AdvInf -T3 53.3%	AdvInf -T5 65.0%	AdvInf -T4 66.7%

Only one subset was found for AdvFin and AdvInf, indicating that the participants' performance was not affected by the type of elicitation task.

Table 8: Tukey groupings for NegInf

Subset 1	NegInf - T4 70.7%	NegInf - T2 75.7%	NegInf - T5 83.3%
Subset 2	NegInf - T2 75.7%	NegInf - T5 83.3%	NegInf - T3 95.0%

Table 9: Tukey groupings for ProInv

Subset 1	ProInv - T3 75.6%	ProInv - T5 93.3%
Subset 2	ProInv - T5 93.3%	ProInv - T4 96.0%

There was a slightly stronger task effect for NegInf and ProInv where two subsets were found: in both cases the learners' performance on task 3 (scaled GJT) was significantly different from their performance on task 4 (ProdT). They were better able to correctly produce structures with ProInv (average = 96.0%), than to judge them (average = 75.6%). However, the opposite is true for NegInf: they correctly judged 95.0% of the stimuli but correctly produced them only 70.7% of the time.

Table 10: Tukey groupings for NegFin

Subset 1	NegFin - T2 68.3%	NegFin - T5 80.0%
Subset 2	NegFin - T5 80.0%	NegFin - T3 83.3%
Subset 3	NegFin - T4 98.7%	

Table 11: Tukey groupings for FQ

Subset 1	FQ - T2 42.8%	FQ - T5 46.3%
Subset 2	FQ - T5 46.3%	FQ - T3 65.3%
Subset 3	FQ - T4 93.3%	

Finally, three subsets were found for NegFin and FQ. For both these properties, the participants' performance was much better on the ProdT task (98.7% and 93.3%, respectively), than on the PrefT (68.3% and 42.8%, respectively). It is interesting to note that learners seemed to be affected by the elicitation tasks regardless of whether or not they have acquired a given property. [4]

6.4 Results by tasks

It was hypothesized that low proficiency L2 classroom learners would exhibit little consistency on different elicitation tasks which solicit various metalinguistic skills. This hypothesis is confirmed by the overall task effect significance displayed above in Table 3 (F (28,2) = 4.42, p = .02). Then, as shown by the results displayed in Table 12, there is a significant task effect for three out of six properties (NegFin, NegInf, FQ).

Table 12: Oneway ANOVA for Task Effect for each Property

	df	*SS*	*MS*	*F*	Pr >*F*
NegFin					
Between groups	3	7034.58	2344.86	10.88	.001
Within groups	42	9046.67	215.40		
NegInf					
Between groups	3	5048.33	1682.78	3.74	.017
Within groups	42	18876.67	449.44		
AdvFin					
Between groups	2	1343.33	671.67	1.67	.205
Within groups	28	11240.00	401.43		
AdvInf					
Between groups	3	6083.33	2027.78	2.53	.070
Within groups	42	33610.42	800.25		
FQ					
Between groups	3	24079.57	8026.52	15.33	.001
Within groups	42	21985.74	523.47		
ProInv					
Between groups	2	3706.53	1853.26	4.15	.026
Within groups	28	12503.40	446.55		

As expected, we found a significant difference among all the properties tested on each one of the four tasks:

- on task 2, the preference/grammaticality task (F (3, 42) = 6.74, p < .001);
- on task 3, the scaled grammaticality judgment task (F (5, 70) = 8.25, p < .001);
- on task 4, the production task (F (5, 70) = 6.38, p < .001);
- on task 5, the magnitude estimation acceptability judgment task (F (5, 70) = 4.73, p < .001).

Again, the participants' performance clearly varied from task to task. The variation is such that it is difficult to distinguish any trends although it appears that the participants performed best on the ProdT (with the exception of AdvInf), then either on the S-GJT or the ME-AJT, and their worst performance was on the PrefG task.

The following tables, Tables 13, 14, 15, 16, display the results of the post hoc analyses which were conducted for each task. The asterix indicates a significant difference (adjusted significance level used in Bonferroni is .05/6 = .00083 for task 2, and .05/15=.0033 for tasks 3, 4 and 5).

Table 13: Task 2 effect

	NegFin	NegInf	AdvInf
NegInf	.386		
AdvInf	.003*	.000*	
FQ	.063	.008*	.209

Table 14: Task 3 effect

	NegFin	NegInf	AdvFin	AdvInf	FQ
NegInf	.094				
AdvFin	.045	.000*			
AdvInf	.000*	.000*	.023		
FQ	.151	.002*	.559	.005	
ProInv	.261	.006	.368	002*	.751

Table 15: Task 4 effect

	NegFin	NegInf	AdvFin	AdvInf	FQ
NegInf	.000*				
AdvFin	.023	.136			
AdvInf	.000*	.585	.043		
FQ	.104	.031	.495	.008	
ProInv	.716	.001	.053	.000*	.205

Table 16: Task 5 effect

	NegFin	NegInf	AdvFin	AdvInf	FQ
NegInf	.690				
AdvFin	1.00	.690			
AdvInf	.076	.031	.076		
FQ	.011	.004*	.011	.425	
ProInv	.113	.233	.113	.001*	.000*

It is particularly interesting to note that task 3 and task 5, the two tasks which directly elicited grammaticality judgments, yield different task effects on properties. There is little difference between the overall task average: 75.0% for task 3, the S-GJT, and 77.1% for task 5, the ME-AJT (displayed in Table 4). However, there are drastic differences between the two tasks if we consider individual properties. For instance, for the placement of floating quantifiers, the participants averaged 73.4% on task 3 but only 58.3% on task 5. And if their performance on floating quantifiers is significantly different from their performance on NegInf in both tasks, it is also significantly different from ProInv only on task 5.

6.5 Individual results

The participants did well as a group on NegFin (average 82.6%). Indeed, two learners performed at a 100% accuracy and four learners performed at 93.7% accuracy, across all four tasks. All learners performed at a 100% accuracy on task 4, the production task. The problem lies with task 2, the PrefG task, for 53% of the participants who were clearly at a loss or simply guessing as indicated by 1 accuracy score of 25% and 7 scores of 50%.

Again a small group of learners did very well on NegInf across all tasks: one performed at 100% accuracy while three others obtained averages between 93.7% and 96.0%. All participants performed extremely well on the grammaticality/acceptability judgment tasks: 80% scored at 100% accuracy on task 5 (the remaining learners scored at 75% accuracy); and slightly less than half of the participants obtained perfect scores on task 3.

The learners' performance presents more variability on AdvFin, the placement of adverbs in finite contexts. Again, there is a handful of learners (n = 3) who have acquired this property with an overall average of 93.3% and a perfect score on the ProdT and the ME-AJT; another small group of learners (n = 3) obtained the same perfect score on the same properties but performed at only 60% accuracy on task 3, to obtain an overall average of 86.7%.

The placement of adverbs in nonfinite contexts shows puzzling results. None of the four learners who obtained accuracy scores on 100% on task 5 were able to perform consistently on the other tasks

where the scores range from 0% to 70% accuracy. Apparently some learners can judge the (un)grammaticality of these stimuli but cannot produce them and vice-versa.

Individual results do not reveal any interesting trends for FQ. As indicated by the standard deviations displayed in Table 6 above, there is a lot of individual variation. The learners' best performance is on the production task, but mostly with non-floating quantifiers, which are also grammatical in English, their L1.

Almost half of the participants (n = 6) obtained perfect accuracy scores on the last property, pronominal inversion, on all four different elicitation tasks. For almost as many learners (n = 5), the culprit was task 3 on which they were clearly guessing as indicated by an accuracy score of 50% in contrast with a perfect score of 100% accuracy on both the production task and most importantly the ME-AJT (overall average of 93.3%). Again, it is very interesting that L2 learners perform so differently on elicitation tasks that require essentially the same metalinguistic skill but that have a different presentation and format.

To sum up, the individual results suggest a better performance of the part of some learners while confirming the variability across tasks and properties that characterized the group results. Some learners performed extremely well if we consider accuracy scores for properties: one learner averaged 90%-100% accuracy on 4 properties, while four learners performed within the same 90%-100% range on 3 properties. If we lower the accuracy range to 80%-100%, we find that it applies to five learners on 4 properties, two learners on 2 properties, and one learner on 3 properties. The learners' performance was less consistent on the elicitation tasks: only three learners obtained 80%-100% accuracy on 2 tasks, and three learners performed at the same accuracy level on 3 tasks. None of the learners were consistent across all four different tasks, but their best performance was on the production task. These results, which indicate that some learners have acquired the appropriate L2 values of the verb movement parameter, also stress that at this overall low level of proficiency, the L2 learners' performance is plagued with inconsistency. This inconsistency may reflect an instability in the principle that is responsible for the structures being tested (Kanno, 1998), or it may be a reflection of the "fuzziness" (Sorace, 1996) of their grammar at this stage of their L2 acquisition.

7. Discussion and conclusion

It was argued that Spanish should be considered as a mixed language with respect to the verb movement parameter because it instantiates both the [+mvt] setting and the [–mvt] setting: lexical verbs do not necessarily raise to Infl. They raise past adverbs in nonfinite contexts and subject floating quantifiers. However, verb movement is optional past adverbs in finite contexts and for pronominal inversion, while it never occurs for negation in finite or nonfinite contexts. On the other hand, English is a [–mvt] language, with the sole exception of *have/be* raising.

It was further argued that the mismatch of parametric values from English as an L1 to Spanish as an L2, in addition to the overlap of Spanish values, would result in a delay of the manifestation of parametric values. However, assuming the availability of the process of parameter (re)setting, it was predicted that the properties would partially cluster. This prediction was supported since the results by properties show homogeneous subsets in which the lowest and highest means are not significantly different. Our L2 learners have acquired the L2 value for negation in both finite and nonfinite contexts, adverbs in finite contexts and pronominal inversion. They have demonstrated that they could accept, reject and produce structures which are not part of their native grammar, thus eliminating the possibility of positive transfer from the L1.

It is difficult to determine whether the second prediction which stated that parameter (re)setting would depend on the consistency of the principle governing verb movement is supported or not, since the participants performed well on pronominal inversion and adverb placement in finite contexts, two properties for which verb movement is optional. On the other hand, they did very well on negation but poorly on adverbs in nonfinite contexts, two properties for which there is consistency: there is no verb movement past negation, whereas the verb raises past adverbs. The relative rarity of structures with adverbs in nonfinite contexts may partly explain the learners' performance in this case. The input may not contain sufficient examples of adverbs in nonfinite contexts to allow the L2 learners to determine which setting of the verb movement parameter is being used for this property. In addition, since verb

movement past adverbs in finite contexts is optional, learners may initially assume that this optionality is also allowed in nonfinite contexts.

These results tend to support the second prediction which stated that learners would initially entertain both settings. They do so correctly for some properties but not others. Again, this is to be expected with verb movement phenomena in Spanish which is best characterized as a mixed language. It is expected that parametric effects be only partially evidenced in the L2 learners' interlanguage as they work out the deductive consequences of the new parameter settings. It is also possible that cognitively mature learners try to apply inductive rules, affecting grammar restructuring. Future research with L2 learners at higher levels of proficiency would allow us to observe whether the parameter (re)setting process is eventually fully completed.

Finally, it was found that the learners' performance varied from task to task, as stated in the third and last prediction. The use of a variety of web-based elicitation tasks clearly produced a rich set of data which is more informative and dependable than a study limited to a single elicitation task, particularly a scaled-GJT. It is usually agreed that traditional scalar GJTs are plagued with so many pitfalls and caveats (*e.g.*, Bard, Robertson & Sorace, 1996; Birdsong, 1989, 1992; Cowart, 1997; Gass, 1994; Schütze, 1996; Sorace, 1996) that they constitute a poor reflection of the learners' competence, particularly when they are used alone. In the case of the present study the single use of a scaled-GJT would have led to inaccurate conclusions since it was found that overall, the learners' performance was influenced by the type of judgment task they were asked to perform. The participants obtained better accuracy scores on the magnitude estimation-AJT than on the scaled-GJT for 3 out of the 6 properties tested (AdvFin, 69.3% vs. 80.0%; AdvInf 53.3% vs. 65.0%; and ProInv, 75.6% vs. 96.3%).

Overall, the learners' best performance was on the production tasks and their worst performance was on the preference/grammaticality task. Let us be more specific about the various tasks. It appears that the participants had the most success on the tasks which were the least artificial and present fewer constraints. The production task, which required the participants to translate sentences from English into Spanish, is an exercise that the learners are very likely to have experienced in their respective foreign language classes. The accuracy means are well above chance level for most of the properties (NegFin, 98.7%; NegInf, 70.7%; AdvFin, 81.7%; AdvInf, 66.7%; FQ, 86.7%; ProInv, 96.0%), indicating that the properties have been acquired and the parameter (re)set since it is implausible that learners could repeatedly "fake" a structure that is not part of their interlanguage. Translating from the L1 to the L2 calls upon the learners to directly compare the two languages. The comparison either yields a corresponding structure in their interlanguage, or it fails to do so, and learners are forced to "retreat" to the L1 structure. However, the drawback of a production task is that it does not exclude the possibility that L2 learners may not know which structures are ungrammatical. This is where the G/AJTs and the PrefG tasks came into play.

The G/AJTs (task 3 and task 5, respectively) are more artificial than the production task. They require complex metalinguistic skills that are unlikely to be developed in a formal instructional setting, or in a naturalistic setting for that matter, which may explain why native speakers rarely perform at a 100% accuracy (although some of our learners did), in addition to explanations which appeal to dialectal variations, individual variations or the "fuzziness" of the structures being tested (Sorace, 1996). Furthermore, "serial order, repeated presentation, deliberate judgment strategies, modality, register, preparation, and judgment speed are all features of the elicitation task that might contribute systematically to variation in judgment", as argued by Schütze (1996, p. 169).

The PrefG task proved to be particularly difficult, and informative, for our participants. It showed that, in spite of their ability to correctly produce a given structure, they were not always able to indicate their preference between two sentences as well as correctly determine whether the other sentence was grammatical or ungrammatical. Thus they consistently performed better on the production task, obtaining a group accuracy means of 98.7% on NegFin for instance, than on the PrefG task on which they scored only 68.3% on the same property. NegInf is the sole exception: they performed better on the PrefG task (75.7%) than on the production task (70.7%).

To conclude, the variety of web-based elicitation tasks provided a more complete insight into the learners' interlanguage than a single traditional GJT (Chapelle, 1998) by yielding a more informative set of data. The mixed language status of Spanish with respect to verb movement phenomena complicated and delayed the process of parameter (re)setting which may nonetheless to be a part of adult L2 acquisition.

Endnotes

1 The verb movement parameter has been alternatively referred to as the V-to-I parameter (Deprez, 1994), the (strength of) AGR parameter (Williams, 1994) and the V-Raising parameter (Culicover, 1997). This study assumes Pollock's (1989, 1997) account while acknowledging that it is not without problems. For a more critical account see Ayoun (1999) who summarizes the theoretical issues raised by for instance Baker, 1991; Bouchard, 1995; Iatridou, 1990; Roberts, 1998; Williams, 1994) for French and English. Moreover, alternative, non-parametric accounts have been suggested (*e.g.*, Baker, 1991; Rohrbacher, 1999).

2 In forthcoming work, it is suggested that these facts can be accounted for within a parameter-and-principle approach with the two following proposals: 1) the two settings of the verb movement parameter are not mutually exclusive but co-existing within the same language for different properties. Thus French and Spanish use both settings of the verb movement parameter: the [+mvt] setting and the [–mvt] setting but for different structures; and 2) languages select a dominant or primary parameter setting along with a minor or secondary parameter setting. Thus French and Spanish use predominantly a "[+mvt] setting".

3 The relatively small number of participants is due to the fact that the study reported here are part of a larger study with two other groups of learners. However, this small number allows for a report of individual results, which would not be manageable with a larger group.

4 Ayoun (1999), who tested the acquisition of the verb movement parameter by Anglophone learners of French as an L2, also found much better accuracy scores on the production tasks than on the scaled GJT.

References

Antes, T., Moritz, C. & Roebuck, R. (1995). Input and parameter resetting in second language acquisition. Cornell Working Papers in Linguistics, 13, 1-23.

Ayoun, D. (1999a). Verb movement in French L2 acquisition. Bilingualism: Language & Cognition, 2, 103-125.

Ayoun, D. (1999b). Bi-lingualism and the "mixed language" phenomenon. Bilingualism: Language & Cognition, 2, 187-189.

Ayoun, D. (2000). Computer-based and web-based elicitation tasks in second language acquisition. Language Learning and Technology, 3, 78-98.

Ayoun, D. (2003). Parameter setting in language acquisition. Cambridge, UK: Continuum Press.

Bard, E., Robertson, B. & Sorace, A. (1996). Magnitude estimation of linguistic acceptability. Language, 72, 32-68.

Benmamoun, E. (1991). Negation and verb movement. NELS, 21, 17-31.

Birdsong, D. (1989). Metalinguistic performance and interlinguistic competence. Berlin: Springer -Verlag.

Birdsong, D. (1992). Ultimate attainment in second language acquisition. Language, 68, 706-755.

Boe, D. (1996). Parameter resetting in adult second language acquisition: inflectional richness and the null subject parameter. Unpublished doctoral dissertation, Indiana University.

Bolotin, N. (1996a). Arabic speakers' resetting of parameters. In M. Eid (Ed.), Perspectives on Arabic linguistics VIII (pp. 135-155). Amsterdam: John Benjamins.

Bolotin, N. (1996b). Resetting parameters in acquiring Arabic. In M. Eid & D. Parkinson (Eds.), Perspectives on Arabic Linguistics IX (pp. 167-178).Amsterdam: John Benjamins.

Bolotin, N. (1996c). Arabic speakers and parasitic gaps. In M. Eid & R. Ratcliffe (Eds.), Perspectives on Arabic linguistics X (pp. 273-282). Amsterdam: John Benjamins.

Chapelle, C. (1998). Construct definition and validity inquiry in SLA research. In L. Bachman & A. Cohen (eds.), Interfaces between second language acquisition and language testing research (pp. 32-70). Cambridge: Cambridge University Press.

Contreras, H. (1991). On the position of subjects. In S. Rothstein (Ed.), Perspectives on phrase structure: heads and licensing. Syntax and semantics 25 (pp. 63-79). New York: Academic Press.

Cowart, W. (1997). Experimental syntax. Applying objective methods to sentence judgments. Thousand Oaks: Sage Publications.

Culicover, P. (1997). Principles and Parameters. An introduction to syntactic theory. New York: Oxford University Press.

Deprez, V. (1994). Under-specification, functional projections, and parameter setting. In B. Lust, M. Suñer & J. Whitman (Eds.), Syntactic theory and first language acquisition: cross-linguistic perspectives. Heads, projections and learnability. Volume 1 (pp. 249-271). Hillsdale: Lawrence Erlbaum.
Downey-Vanover, J. (1994). Adult access to Universal Grammar in second language acquisition. Unpublished doctoral dissertation. University of Maryland.
Emonds, J. (1978). The verbal complex V'–V in French. Linguistic Inquiry, 9, 151-175.
Falk, C. (1993). Non-referential subjects and agreement in the history of Swedish. Lingua, 89, 143-180.
Fodor, J. D. (1994). How to obey the Subset Principle: Binding and locality. In B. Lust, G. Hermon & J. Kornfilt (eds.), Syntactic theory and first language acquisition: volume 2 (pp. 429-451). Hillsdale, NJ: Lawrence Erlbaum.
Fodor, J.D. (1998). Unambiguous triggers. Linguistic Inquiry, 29, 1-36.
Fodor, J.D. (1999). Learnability theory. Triggers for parsing with. In E. Klein & G. Martohardjono (Eds.), The development of second language grammars. A generative approach (pp. 363-406). Amsterdam: John Benjamins.
Gass, S. (1994). The reliability in second language grammaticality judgments. In E. Tarone, S. Gass & A. Cohen (eds), Research methodology in second language acquisition (pp. 303-322). Hillsdale: Erlbaum.
Hawkins, R., Towel, R. & Bazergui, N. (1993). Universal Grammar and the acquisition of French Verb Movement by native speakers of English. Second Language Research, 9, 189-233.
Herschensöhn, J. (1998). Minimally raising the verb issue. In A. Greenhill, M. Hughes, H. Littlefield & H. Walsh (Eds.), Proceedings of the 22nd Annual Boston University Conference on Language Development (pp. 325-336). Somerville: Cascadilla Press.
Herschensöhn, J. (2000). The second time around. Minimalism and L2 acquisition. Amsterdam/ Philadelphia: John Benjamins.
Hulk, A. (1991). Parameter setting and the acquisition of word order in L2 French. Second Language Research, 7, 1-34.
Hulstijn, H. (1997). Second language acquisition research in the laboratory. Possibilities and limitation. Studies in Second Language Acquisition, 19, 131-43.
Kanno, K. (1998). The stability of UG principles in second language acquisition: evidence from Japanese. Linguistics, 36, 1125-1146.
Kayne, R. (1981). On certain differences between French and English. Linguistic Inquiry, 12, 349-371.
Kayne, R. (1984). Connected and binary branching. Dordrecht: Foris.
MacLaughlin, D. (1995). Language acquisition and the Subset Principle. The Linguistic Review, 12, 143-191.
Mandell, P. (1998). The V-movement parameter: syntactic properties and adult L2learners of Spanish. Spanish Applied Linguistics, 2, 169-197.
Muysken, P. (1998). Code-switching processes: alternation insertion, congruent lexicalization. In M. Putz (Ed.), Language choices, constraints and consequences (pp. 361-380). Amsterdam: John Benjamins.
Olarrea, A. (1997). Pre and postverbal subject positions in Spanish: a Minimalist account. Unpublished doctoral dissertation, University of Washington.
Ouhalla, J. (1990). Sentential negation, relativized minimality and the aspectual status of auxiliaries. The Linguistic Review, 7, 183-231.
Ouhalla, J. (1991a). Functional categories and parametric variation. London/New York: Routledge.
Platzack, C. & Holmberg, A. (1989). The role of AGR and finiteness in Germanic VO languages. Working Papers in Scandinavian Syntax, 43, 51-76.
Pollock, J.-Y. (1989). Verb movement, Universal Grammar, and the structure of IP. Linguistic Inquiry, 20, 365-424.
Pollock, J.-Y. (1997a). Langage et cognition. Introduction au programme minimaliste de la grammaire générative. Paris: Presses Universitaires de France.
Pollock, J.-Y. (1997b). Notes on clause structure. In L. Haegeman (Ed.), Elements of grammar: handbook in generative syntax (pp. 237-279). Dordrecht: Kluwer.
Roberge, Y. (1990). The syntactic recoverability of null arguments. Montreal: McGill-Queen's University Press.
Roberts, I. (1993). Verbs and diachronic syntax: a comparative history of English and French. Dordrecht: Kluwer Academic Publishers.
Rohrbacher, B. (1994). The Germanic VO languages and the full paradigm. Unpublished doctoral dissertation, University of Massachusetts, Amherst.
Rohrbacher, B. (1999). Morphology-driven syntax. A theory of V to I raising and pro-drop. Amsterdam: John Benjamins.

Schütze, C. (1996). The empirical base of linguistics. Chicago: University of Chicago Press.
Sorace, A. (1996). The use of acceptability judgments in second language acquisition research. In W. Ritchie & T. Bhatia (Eds), Handbook of second language acquisition (pp. 375-409). San Diego: Academic Press.
Suñer, M. (1994). V-movement and the licensing of argumental wh-phrases in Spanish. Natural Language and Linguistic Theory, 12, 335-372.
Torrego, E. (1984). On inversion in Spanish and some of its effects. Linguistic Inquiry, 15, 103-129.
Trosterud, T. (1989). The null subject parameter and the new mainland Scandinavian word order: a possible counter-example from a Norwegian dialect. Papers from the 11th Scandinavian Conference of Linguistics, 87-100.
Vainikka, A. & Levy, Y. (1999). Empty subjects in Finnish and Hebrew. Natural Language & Linguistic Theory, 17, 613-671.
Vikner, S. (1995). V^o-to-I^o movement and inflection for person in all tenses. Working Papers in Scandinavian Syntax, 55, 1-27.
Vikner, S. (1997). V^o-to-I^o movement and inflection for person in all tenses. In L. Haegeman (Ed.), The new comparative syntax, (pp. 189-213). London/New York: Longman.
Weissenborn, J., Verrips, M. & Berman, S. (1989). Negation as a window to the structure of early child language. Ms. Max Planck Institute für Psycholinguistik, Nijmegen.
White, L. (1991). The verb-movement parameter in second language acquisition. Language Acquisition, 1, 337-60.
White, L. (1992). Long and short verb movement in second language acquisition. Canadian Journal of Linguistics, 37, 273-86.
Williams, E. (1994). A reinterpretation of evidence for verb movement in French. In D. Lightfoot & N. Hornstein (Eds.), Verb Movement (pp.189-205). Cambridge: Cambridge University Press.
Yuan, B. (1995). Acquisition of base-generated topics by English-speaking learners of Chinese. Language Learning, 45, 4, 567-603.

ISB4: Proceedings of the 4th International Symposium on Bilingualism

edited by James Cohen, Kara T. McAlister, Kellie Rolstad, and Jeff MacSwan

Cascadilla Press Somerville, MA 2005

Copyright information

ISBN 978-1-57473-210-8 CD-ROM
ISBN 978-1-57473-107-1 library binding (5-volume set)

Ordering information

To order a copy of the proceedings, contact:

Cascadilla Press
P.O. Box 440355
Somerville, MA 02144, USA

phone: 1-617-776-2370
fax: 1-617-776-2271
sales@cascadilla.com
www.cascadilla.com

Web access and citation information

This paper is available from www.cascadilla.com/isb4.html and is identical to the version published by Cascadilla Press on CD-ROM and in library binding.