Language Learning & Technology
http://llt.msu.edu/vol3num2/ayoun/index.html

January 2000, Volume 3, Number 2
pp. 82-105
(page numbers in PDF differ and should not be used for reference)

WEB-BASED ELICITATION TASKS IN SLA RESEARCH [1]

Dalila Ayoun
University of Arizona

ABSTRACT

This paper presents an experimental study in second language acquisition (SLA) designed with web-based elicitation tasks to obtain greater internal and external validity. The background information questionnaire and three experimental tasks--a scaled grammaticality judgment task, a preference/grammaticality task and a production task--were created with a web-based software, Claris Homepage®, while a fourth experimental task--a magnitude estimation acceptability judgment task--was designed with a multimedia software, Director®. The present study tested the acquisition of the properties subsumed under the verb movement parameter and the null subject parameter by English native speakers enrolled in French college classes. The results for the former parameter but not the latter support the hypothesis of progressive parametric manifestation in L2 learners' grammar. The preference/grammaticality and production tasks proved particularly informative.

Methodology in the field of second language acquisition (SLA) research has been surprisingly limited to a few elicitation tasks, such as scalar grammaticality judgment tasks (GJT), with simple media, such as pencil-and-paper, since the early 1980s. The problems and caveats of GJTs are well-documented (e.g., Bard, Robertson, & Sorace, 1996; Birdsong, 1989, 1992; Cowart, 1997; Gass, 1994; Schütze, 1996; Sorace, 1996). Researchers usually conclude that traditional scalar GJTs are a poor reflection of the learners' competence, particularly when they are used alone. However, they may be modified to be more informative, especially when combined with other types of elicitation tasks, such as computer- or Web-based tasks.

The versatility and potentially powerful applications of computer-based tasks remains largely unexplored. One rare exception is Doughty (1991) who used computer-assisted reading lessons to evaluate the effectiveness of instruction in the acquisition of English relative clauses by native speakers of various languages. The experimental part of the lessons was highly controlled, while the participants were given a certain amount of user flexibility whenever appropriate (e.g, to navigate through the lessons or to access lexical information).

This study intends to show how a rich and wide set of data can be elicited using a variety of innovative Web-based tasks which offers the following advantages:

- it eliminates or minimizes human error in collecting and recording the data;
- it facilitates the organization and analysis of data;
- it avoids missing data since the tasks may be designed to require the participants to enter a response;
- it ensures that all participants are exposed to the stimuli and/or the treatment in the same way (Hulstijn, 1997);
- it permits the accuracy of timed experiments; and
- it allows for a better control of experimental variables in general. The data are then more likely to better inform SLA processes.

Participants may also approach Web-based elicitation tasks differently than they approach traditional pencil-and-paper tasks. The mere novelty of the medium with its sophisticated presentation, ease of navigation, and speed may have all a positive impact on the participants. A 50 item GJT is spread over several pages on paper, whereas the length does not seem as daunting on a computer screen with scrolling arrows. It may also be speculated that performing a task on a computer screen may provide some isolation which participants lack in a classroom. That isolation may be conducive to better concentration, and therefore improved performance.

Being able to better inform SLA processes is especially critical for empirical studies which have focused on the parameter-setting approach to language acquisition. This line of research has produced mixed results and has been variously interpreted (e.g., Meisel, 1998; Schwartz, 1998; White, 1996) partly because it lacks sufficient convincing evidence such as converging data from a variety of tasks. Due to practical constraints, participants are often administered a single task. It is difficult to draw solid, generalizable conclusions based on a sole elicitation task. However, the parameter-setting approach remains actively pursued for it directly addresses the question of ultimate attainment in adult SLA which amounts to asking whether adult learners are able to (re)set their parameters and thus still access Universal Grammar (Ayoun, 1999; Bley-Vroman, 1998; Schwartz, 1998).

This paper reports on the innovative Web-based design of an empirical study which is part of a larger project investigating the acquisition of French and Spanish by English native speakers from a parameter (re)setting perspective (Ayoun, 2000). It focuses on the (re)setting of two of the most studied parameters of UG by English native speakers learning French, the verb movement parameter (Pollock, 1989, 1997) and the null subject parameter (Chomsky, 1981). Since the main emphasis of this paper is on the experimental design of the study, the linguistic and learnability issues will only be briefly mentioned (see e.g., Ayoun, 1999; Pollock, 1997, for a more detailed account).

LINGUISTIC FOCUS

The Verb Movement Parameter[2]

Following early work from Emonds (1978), Pollock (1989, 1997) observed that the distribution of four surface level syntactic properties was determined by whether finite lexical verbs raise out of the verb phrase (VP) as in French, or do not, as in English. These four properties are negation placement, adverb placement, floating quantifiers and inverted questions as illustrated in the following examples:

(1) a. Jean ne veut pas les livres.

b. *John wants not the books.

(2) a. Eric regarde toujours les actualités.

b. *Eric watches always the news.

(3) a. Mes amis vont tous à la plage.

b. *My friends go all to the beach.

(4) a. Veut il les livres?

b. *Wants he the books?

Assuming an underlying structure in which the verb follows adverbs and negatives as illustrated in (5),

(5) [$_{IP}$ NP I ([$_{Neg}$ not/pas]) [$_{VP}$ (Adv) V...]]

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

Within the Minimalist framework (Chomsky, 1995), the visibility of "rich" morphological features at LF triggers overt verb movement to Infl before Spell-Out to check and erase these features which would otherwise violate the Full Interpretation Principle. It follows that verb movement in English is limited to *have/be* raising, to the exclusion of lexical verbs because their verbal paradigm is morphologically poor, lacking features of person and number as shown in (6):

(6) a. Have you seen your brother?

b. Didn't you say he was sick?

Non-finite lexical and auxiliary verbs should not raise since their morphological features are assumed to be weak. However, Ayoun (1998a) found that native speakers allow nonfinite lexical verbs and auxiliaries to move past adverbs as in (7a), while rejecting the optional raising of nonfinite auxiliaries past negation as assumed in Pollock (1995, 1997) and Roberts (1998) as in (7b):

(7) a. Sortir souvent, c'est amusant.

"To often go out is fun"

b. Ne pas avoir peur, c'est rare / n'avoir pas peur, c'est rare

"To not be afraid is rare"

There is no principled explanation for this optionality of verb movement in relation to adverbs.

Thus French and English differ on whether overt verb movement is allowed or not. Verb movement does not systematically apply to [+finite] lexical verbs with strong morphological features: both languages allow structures with and without verb movement, only in [+finite] contexts for English auxiliaries, but in both [+finite] and [-finite] contexts with respect to adverb placement for French lexical verbs as well as with respect to negation for auxiliaries in the grammar of older speakers (see Pollock, 1995).

The Null Subject Parameter

English and French are considered to be [-null subject] languages which require overt subject pronouns as in (8) and expletives as in (9), while disallowing *that*-trace effect as in (10) and nominal inversion as illustrated in (11):

(8) a. Elle sort trop souvent / *sort trop souvent.

b. She goes out too often / *goes out too often.

(9) a. Il a plu toute la journée / *a plu toute la journée.

b. It rained all day / *rained all day.

(10) a. *Qui a-t-il dit qu'est venu?

b. *Who did he say that came?

(11) a. *Vient Marie cet après-midi.

b. *Comes Mary this afternoon.

In spite of the extensive research it has generated since its initial formulation by Perlmutter (1971), the licensing of null subjects parameter still remains somewhat of a mystery: first, not all richly inflected languages allow null subjects while some poorly inflected languages do, such as Japanese, Korean, and Chinese (e.g., Fukui, 1995; Hasegawa, 1985; Huang, 1989); second, uniformly inflected languages such as German and Icelandic show that the Morphological Uniformity Principle (Jaeggli & Safir, 1989) does

not hold. The same problem exists for the verb movement parameter: there seems to be a correlation between richly inflected verbs and verb raising, but not all languages with rich verbal inflection allow verb movement (e.g., Lardiere, 1998, Rohrbacher, 1999). These issues are not addressed here and will have to be in future research.

From a learnability perspective, English native speakers learning French will have to reset the verb movement parameter which differs from the L1 to the L2, but not the null subject parameter. Following others (e.g., Ayoun, 1999; Bolotin, 1996a, 1996b, 1996c; Uziel, 1993), it is argued that parameter resetting is evidenced by a partial clustering of the properties due to a progressive manifestation of parameter-setting properties (e.g., Ayoun, 1999; Platt, 1993; Schwartz & Tomaselli, 1990; Uziel, 1993). This partial clustering, as opposed to a simultaneous clustering is due to the following two factors:

> a) the complexity and systematic application of the abstract syntactic mechanism involved for each parameter: verb movement does not apply in a systematic and consistent fashion in French, while the null subject parameter does.
>
> b) the learners' level of proficiency: it is assumed that the longer learners are exposed to primary language data, the higher their proficiency level will be[3].

This learnability approach translates into the following research hypotheses to be tested statistically with a repeated measures ANOVA design followed by post hoc Tukey tests as appropriate.

RESEARCH HYPOTHESES

Hypothesis 1

> The syntactic properties subsumed under the verb movement parameter and the syntactic properties subsumed under the null subject parameter will be acquired as a cluster.

Hypothesis 2

> There will be a relationship between the learners' proficiency level, as measured by the pre-test, and their performance on the experimental tasks: the higher the proficiency level, the better the learners' performance across tasks should be.

Hypothesis 3

> There will be a relationship between the learners' performance on various syntactic properties and task types: the learners' performance should be consistent.

Hypothesis 1 predicts that there should not be significant differences between the properties subsumed under each parameter. Evidence of successful parameter resetting is taken to be a better-than-chance performance (which is usually assumed to be at or above 60%). Properties are not expected to cluster following any particular or specific order since they are governed by the same abstract principle of grammar. Hypothesis 2 predicts that there will be a positive correlation between the learners' score on the proficiency test and their performance across experimental tasks. And Hypothesis 3 predicts that that there should be an interaction between the learners' performance on various syntactic properties and task types.

STUDY DESIGN

Description of the software

Claris Homepage®Claris Homepage is software used to design World Wide Web (Web) pages. The opening untitled page resembles a classic word processing page and one may simply start typing new text, or import (by copying and pasting) text from an existing text file created by another application such as a word processor. Once created, Web pages are written in HTML (Hypertext Markup Language), the language used to format documents so that they may be "published" or uploaded on the Web. HTML uses formatting commands or "tags" that are embedded around the various parts of a Web document so that a browsing software such as Netscape® may display documents' links, text, and graphics properly. The tags do not appear on the browser screen, but translate the information for the browser. However, no knowledge of HTML encoding is required to use Claris Homepage® since it automatically encodes any text file in the appropriate HTML format.

Claris Homepage® is flexible and provides pull-down menus for a variety of options in text style, font, color, size, etc. Users may simply click on the small icons located on top of the page to take advantage of many functions to create links to other Web pages, to preview and test their pages in a Web browser, to change the alignment of the text, or to insert tables, images, or a variety of forms. The latter are of special interest for designing the types of elicitation tasks commonly used in SLA research and refer to the following:

- *Text fields* are rectangular boxes, the dimensions of which may be specified as needed. They may be used for a cloze test or to enter a single word of information such as 'user-name', 'ID#' or 'password'. Participants click in the text field to start typing and use either the tab key or the mouse to move from field to field.
- *Text areas* are larger boxes; their dimensions may be specified as well, and they are best used to enter more extensive text such as short narratives or longer answers to open-ended questions. They are bordered by vertical and horizontal arrows to scroll through the text.
- *Check boxes* and *radio buttons* are small squares and circles respectively. When participants click on them, an X or a check mark appears to indicate that they have been selected. They may be used to collect data such as participants' gender, college major or status. Researchers may allow or disallow multiple selections.
- *Pop-up menus* are used to list several options, for example a variety of languages under the category 'native language' or 'second language'. To use pop-up menus, the participants scan up and down the answers provided and select one by releasing the mouse on the desired option. Any given selection may be changed and multiple selections may be allowed.
- Finally, you may insert a form known as *password* to limit access to the Web page(s) to specific individuals if necessary.

Director Macromedia® Director® is a very sophisticated and flexible software whose purpose is to create multi-media, computer-based documents. Text, sound, and images may be combined to create interactive animations called "movies." The software provides ready-made scripts to make basic movies. These scripts may be used as provided or customized for specific purposes. The complexity of this software requires either a solid knowledge of its programming language or the assistance of a computer programmer. However, its extreme versatility and power make it worthwhile for creating multi-media documents. In the present study, a pre-existing movie script was modified to create an original magnitude estimation acceptability judgment task to be described below[4]

The next section details how Claris Homepage® and Director® were used to create innovative elicitation tasks to obtain a richer set of data in a parameter (re)settting study.

Tasks and Stimuli

The tasks were performed in the order in which they are presented here. Task 1 was a pre-test which included 41 stimuli exemplifying a variety of syntactic properties (dative alternation, dative passive, Exceptional-Case Marking, interrogative and declarative pied piping/preposition stranding, *il est/c'est*, and various tense/aspectual distinctions) to obtain an independent measure of the participants' proficiency level[5]. The experimental tasks (Task 2 through Task 5) tested the following properties:

- verb movement parameter: 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).
- null subject parameter: (null) subject pronouns in main clauses (ProSub) and embedded clauses (ProEmb); expletives (Explet); extraction for that-trace effect (Extr); nominal subjects (NomSub); and nominal subject inversion (NomInv).

All stimuli were controlled for length and simplicity of vocabulary. No distractors were included since the stimuli offered sufficient variety.

Background Information Questionnaire Participants provided the usual background information such as age, gender, native language, second language, language proficiency, and so forth. Almost all types of forms were put to use: check boxes, radio buttons, pop-up menus, text fields, and text areas. Pop-up menus were especially useful to indicate language proficiency: the learners were given several options from novice-low to native-like. They were coded into numbers (invisible to the participants) for statistical purposes. To avoid missing data, the fields were designed as required fields: participants had to enter a response for the submission of the form to be valid.

Scaled Grammaticality Judgment/Correction Task Task 1 (PreTest) and Task 3 (S-GJT)
Participants were asked to judge 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 not visible to the participants. It was deemed preferable to ask participants to make judgments according to the Likert scale rather than to choose numbers since numbers may be perceived differently or assigned different degrees of grammaticality or ungrammaticality by participants. The researcher was present to answer questions participants had the Likert scale or other parts of the experiment. 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.[6] Task 3 included 45 sentences to test all properties with the exception of adverb placement in nonfinite contexts.

Preference/grammaticality task (Task 2) (PrefGT) Participants were presented with pairs of sentences. Below the stimuli were two pop-up menus: the first allowed the participants to indicate which of the two sentences, A or B, they preferred; while the second permitted the participants to specify whether the other sentence, that is, 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. Participants clicked on the pop-up menus to see the relevant options and released the mouse on the desired selection. They had the option of going back to change a previous selection. Participants were instructed to complete all items before submitting the form. This task was composed of 34 pairs of sentences to exemplify all the properties except for adverb placement in infinitive clauses and subject pronoun inversion in finite

clauses. The first property was tested only in the production task, following earlier research (Ayoun, 1998a, 1998b) which showed its "ambivalence" or "fuzziness" (Sorace, 1996). The last property is directly tested in all other tasks and indirectly in several stimuli in Task 2.

Production task (Task 4) (ProdT) Participants were asked to translate sentences from their L1 into French. Participants typed their answers in a text field below the stimulus sentence. All text fields were specified as required to avoid missing data. Task 4 tested all properties with 35 sentences.

Magnitude estimation acceptability judgment task (Task 5) (ME-AJT) This task is the most innovative both in terms of its design and for SLA research. It was inspired by the work of Sorace and her colleagues (Bard et al., 1996; Sorace, 1996) who explored the possibility of applying magnitude estimation from psychophysics to linguistic judgments. It consists in asking participants to judge or rate stimuli in relation to one another. For instance,

> In numerical estimation, informants are 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 individual sentences that the participants were asked to read 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.

This is how Task 5 appeared: 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 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. To avoid missing data, the form could not be submitted until all stimuli presented in each screen had been judged. This task included all properties with the exception of adverb placement in infinitive clauses with a total of 50 sentences.

COMPUTERIZED DATA COLLECTION

Once the elicitation tasks designed with Claris Homepage® were completed, they were uploaded on a server before the beginning of the experiment so that participants could access them from any computer equipped with a Web browser. The participants entered their first and last names (which were later coded to preserve their anonymity), completed the tasks, and clicked on the button SUBMIT to send the data to a folder located on the server.

Computerized data collection is greatly facilitated by the possibility to organize it in any way the researcher may see fit. For example, the data may be sent to group folders organized by levels of proficiency or by language, or to individual folders, with one folder for each participant or alternatively, the folders may be divided by tasks. The data organization may also be modified later to conduct different types of analyses.

As mentioned, an extremely useful feature of Claris Homepage® is that the forms (e.g., text field or pop-up menus) can be designed so that participants are required to enter information before submitting the form. When some information was missing, participants received a message asking them to click on "Back" on their browser to return and complete the task by supplying the missing information. When the task was properly completed, participants received a positive confirmation message. This feature allowed us to avoid the common problem of missing data. The resulting analyses are thus more reliable in that we obtained complete data sets.

In the present study, Task 5 was installed on hard drives to be accessed quickly. It is also available through anonymous file transfer protocol (FTP): the task can be downloaded simply by clicking on it.[7] It was created in both MacOS and Windows versions. As already mentioned, Task 5 was also designed so that all the stimuli had to be judged: first, before the participants could move on to the next screen; and second, so that the data they entered could be properly submitted by clicking on the SUBMIT button at the end of the task. As in the Claris Homepage tasks, the learners then received a positive confirmation message. The data were saved in the appropriate folders on the server to be later accessed for statistical analysis.

Participants

The participants (13 females and 4 males with an average age of 20) were enrolled in various classes: most of them (*n*=11) were taking only one French course which ranged from second year to graduate level for one participant. There were 5 French majors and 12 non-language majors.

RESULTS

The raw data were computed in the following manner. In the scaled GJTs and magnitude estimation AJT, participants were awarded one point per correct answer, namely, one point for each correctly accepted sentence (grammatical sentence accepted as "grammatical") and one point for each correctly rejected sentence (ungrammatical sentence rejected as "ungrammatical"). In the scaled GJT a correctly accepted sentence was rated 4 or 5, while a correctly rejected sentence was rated 1 or 2. In the magnitude estimation AJT, a correctly accepted sentence was any positive number from 1 to 100, while a correctly rejected sentence was any negative number from -1 to -100. 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 grammaticality/preference task reflect the participants' ability to make two decisions: first, selecting the correct sentence from a pair of stimuli and second, correctly accepting (when grammatical) or correctly rejecting (when 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.

And finally, in the production task, participants were awarded one point per sentence for correctly producing the relevant structure. Vocabulary errors or gender agreement mistakes were ignored for 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%).

Results by Properties

If parameter setting is taking place in L2 acquisition, there should not be any significant differences between properties as they progressively cluster. Again, we do not expect a specific clustering order since all the properties are governed by the same abstract principle. They should simply be acquired at about the same time. Table 1 presents the accuracy percentage means obtained on the properties for each parameter across all four experimental tasks for the L2 learners as a group. Table 1 presents the means obtained on the properties for both parameters.

Table 1. Results by Properties

Vb-Mvt	NegFin	NegInf	AdvInf	AdvInf	FQ	ProInv	average
	77.6%	62.7%	66.1%	73.5%	67.3%	93.0%	73.5%
Null subj.	ProSub	ProEmb	Expl.	Extr	NomSub	NomInv	average
	68.1%	88.2%	89.2%	52.6%	94.6%	79.1%	78.6%

The learners' performance appears to vary from property to property and the ANOVA presented in Table 2 confirms that there were significant differences among verb movement properties.

Table 2. ANOVA on Verb Movement Properties

Verb movement properties	*SS*	*df*	*MS*	*F*	Sig. of *F*
Between groups	10315.48	5	2063.09	5.60	.001
Within groups	29436.00	80	367.95		

A post hoc Tukey test (confidence level is consistently at .05) revealed that there are significant differences between:

- neginf and proinv
- advfin and proinv
- advinf and proinv
- FQ and proinv

These significant differences are also displayed in Table 3 where they are indicated by (*) in the bottom row:

Table 3. Significant Differences Between Verb Movement Properties

VerbMvt	NegFin	NegInf	Advfin	AdvInf	FQ	ProInv
NegFin						
NegInf						
Advfin						
AdvInf						
FQ						
ProInv	*	*	*	*		

The post hoc Tukey thus indicates that the properties subsumed under the verb movement parameter do partially cluster since they form two homogeneous subsets (i.e., the highest and lowest means are not significantly different) as follows:

Table 4. Clustering of Verb Movement Properties

Subset 1 means	NegInf 62.7%	AdvFin 66.1%	FQ 67.3%	AdvInf 73.5%	NegFin 77.6%
Subset 2 means	NegFin 77.6%	ProInv 93.0%			

These two subsets indicate that most of the verb movement properties cluster in the L2 learners' grammar, which confirms our first hypothesis. Learners performed significantly worse on negative placement, adverb placement and floating quantifiers in comparison to verb-subject pronoun inversion where they did very well, obtaining an overall average of 93%. The other properties are much closer together but with relatively low scores.

Table 5 indicates that there are significant differences among null subject properties as well.

Table 5. ANOVA on Null Subject Properties

Null subject properties	*SS*	*df*	*MS*	*F*	*pr* >F
Between groups	21199.22	5	4242.20	23.98	.001
Within groups	14153.60	80	176.92		

A post hoc Tukey test revealed that there are significant differences between:

- prosub and proemb
- prosub and expl
- prosub and nomsub
- proemb and extract
- expletive and extraction
- extract and nomsub
- extract and nominv
- nomsub and nominv

Or as illustrated by (*) in the white cells of the lower triangle in Table 6:

Table 6. Significant Differences Between Null Subject Properties

NullSubj	Extr	ProSub	NomInv	ProEmb	Expl	NomSub
Extr						
ProSub	*					
NomInv	*					
ProEmb	*	*				
Expl	*	*				
NomSub	*	*	*			

The post hoc Tukey test indicates that the null subject properties form four homogeneous subsets (i.e., the highest and lowest means are not significantly different) as displayed in Table 7.

Table 7. Clustering of Null Subject Properties

Subset 1 means	Extr 52.6%		
Subset 2 means	ProSub 68.1%	NomInv 79.1%	
Subset 4 means	NomInv 79.1%	ProEmb 88.2%	Expl 89.2%
Subset 4 means	ProEmb 88.2%	Expl 89.2%	NomSub 94.6%

The participants performed very poorly on extraction with a means of 52.6%. The other properties appear to be acquired as a cluster or rather, as clusters.

To sum up, we may say that our first hypothesis which stated that the properties subsumed under each parameter would be acquired as a cluster is confirmed for the verb movement parameter, providing evidence of initial successful parameter resetting in the grammar of our L2 learners. It is more difficult to reach the same conclusion for the null subject parameter. The participants performed below chance level on extraction, and properties form four clusters as opposed to only two. The null subject properties may thus be acquired individually as opposed to as a parametric cluster. However, the difference between the two parameters is not statistically significant (difference mean = -5.80, *SD* = 16.387, *t* = -1.46, *df* = 16, *p* = .163).

Results by Tasks

The results by tasks allow us to see whether the second hypothesis, which stated that there would be a relationship between the learners' proficiency levels and their performance on the various tasks, was confirmed (see the Appendix for complete results). The scores were computed as explained above for each participant. The means represent overall accuracy percentage scores across all properties. They were averaged to obtain group results which are presented in Table 8.

Table 8. Results by Task

	Task 1 PreTest	Task 2 PrefGT	Task 3 S-GJT	Task 4 ProdT	Task 5 ME-AJT	average
L2 FRE	67.02%	70.48%	71.14%	80.37%	77.57%	74.89%

Or as shown in Figure 1:

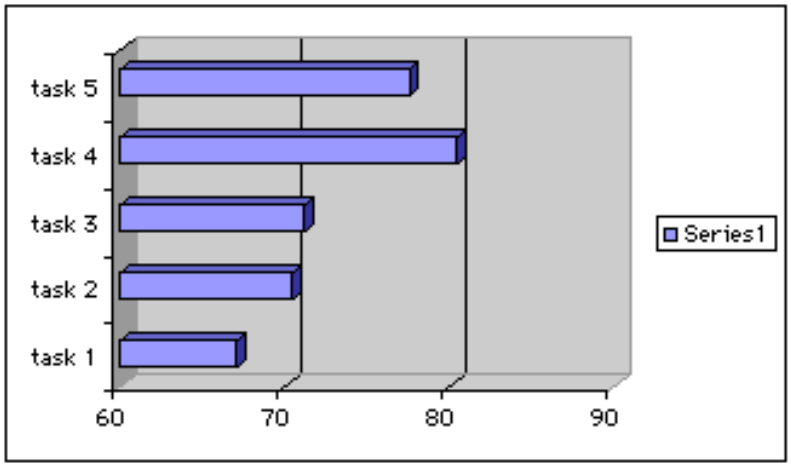

Figure 1.

The pre-test (Task 1) shows an overall low level of proficiency (67%)8 and it does not seem to be a good predictor of the learners' performance on the experimental tasks. The learners' scores are more consistent and higher across all experimental tasks. At least four learners performed in the 80% range, showing that they have acquired the properties tested for both parameters. The group average score was the highest on the production task: 80.4% with a low of 63% and a high of 91.4%. All scores are above the better-than-chance level of 60%. The results of a Pearson correlation test displayed in Table 9 confirmed that learners performed differently from task to task.

Table 9. Correlations Between Tasks

		Task 1 PreTest	Task 2 PrefGT	Task 3 S-GJT	Task 4 ProdT
Task 2 PrefGT	Pearson cor. Sig. (2-tailed)	.476 p =.053			
Task 3 S-GJT	Pearson cor. Sig. (2-tailed)	.600 p =.011	.559 p =.029		
Task 4 ProdT	Pearson cor. Sig. (2-tailed)	.356 p =.160	.022 p =.935	.711 p =.001	
Task 5 ME-AJT	Pearson cor. Sig. (2-tailed)	.354 p =.164	.366 p =.148	.699 p =.002	.660 p =.004

The higher the correlation, the more similarly the participants performed. We thus see that the pre-test was a relatively good predictor of the participants' performance only for task 3 (r = .600, p = .05) but not for the other experimental tasks. This result may be explained by the fact that both the pre-test and task 3 were scaled GJTs. And the results of a oneway ANOVA presented in Table 10 indicate a significant difference between tasks:

Table 10. ANOVA on Tasks

	SS	*df*	*MS*	*F*	*pr* >F
Between tasks	2042.3878	4	510.5969	6.2288	.0003
Within tasks	5246.2720	64	81.9730		

The results of a post-hoc Tukey test displayed in Table 11 show that these significant differences are to found between the following tasks indicated by (*) in the white cells in the lower triangle:

Table 11. Post Hoc Tukey on Tasks

	Task 1 PreTest	Task 2 PrefGT	Task 3 S-GJT	Task 4 ProdT	Task 5 ME-AJT
Task 1 PreTest					
Task 2 PrefGT					
Task 3 S-GJT					
Task 4 ProdT	*				
Task 5 ME-AJT	*	*	*		

There are thus significant differences between

- Task 1 (PreTest) and Task 4 (ProdT), as well as Task 1 and Task 5 (ME-AJT)
- Task 2 (PrefGT) and Task 5 (ME-AJT)
- Task 3 (S-GJT) and Task 5 (ME-AJT)

These results show that participants performed differently from task to task. This may be explained by contending that different elicitation tasks require greater metalinguistic skills than those possessed by our learners at the time of their participation. An alternative explanation is that our learners' L2 competence level is simply not high enough yet to perform consistently on different types of elicitation tasks. Some degree of uncertainty or fuzziness in their interlanguage grammar may be revealed more easily when they are required to perform experimental tasks which solicit a variety of metalinguistic skills. For example, the preference/grammaticality task required the participants not only to correctly select one sentence but to correctly judge the (un)grammaticality of the rejected sentence. Thus in addition to percentage scores for CS/CG (correct sentence/correct grammaticality) already given above, the results included the following:

- The percentage scores for CS/IG (correct sentence/incorrect grammaticality) indicate that the participants selected the right sentence but not the (un)grammaticality status of the rejected sentence. For example when presented with the pair in (12):

  (12) a. Paul souvent va au marché.

  b. Paul va souvent au marché.

  Participants correctly showed a preference for (12b) and then correctly indicated that (12a) was ungrammatical.

- The percentage scores for IS/CG (incorrect sentence/correct grammaticality) show that the participants selected the wrong sentence but the correct (un) grammaticality status of the rejected sentence. Here participants preferred (12a) but judged (12b) to be grammatical.
- The percentage scores for IC/IG (incorrect sentence/incorrect grammaticality) show that participants were unable to select the right sentence or the right (un) grammaticality status. The participants' selection was (12a) for "sentence preferred" and "ungrammatical" for (12b).

All these results are presented in Table 12. The results displayed in the "total" column adds the percentage scores from CS/CG and CS/IG.

Table 12. Preference Task Results

PrefGT	CS/CG	CS/IG	Total	IS/CG	IC/IG
L2 FRE	70.5%	14.7%	85.2%	9.8%	5.0%

The results displayed in "total" now indicate that the learners were actually quite successful at correctly selecting the right sentence (they completely missed only 5% of the stimuli) but had difficulty in consistently judging the (un)grammaticality of the other sentence. This suggests that their interlanguage is still entertaining mixed settings by allowing both the L1 and the L2 options (Ayoun, 1998b; Bolotin 1996a, 1996b, 1996c; White, 1992).

Results by Properties/Tasks

A repeated measures ANOVA was run to determine whether there was any difference in the participants' performance on specific properties in different types of tasks. The results reported in Table 13 reveal (a) a significant task effect, (b) a significant property effect, and (c) a significant property by task interaction which means that the differences among the properties varied by task.

Table 13. Repeated Measures ANOVA on Tasks/Properties

Source	*df*	*MS*	*F*	*Pr* > *F*
Property	10	12901.1052	23.99	.0001
Error (property)	160	863.35		
Task	3	5145.35	5.55	.0024
Error (task)	48	926.66		
Property*Tasks	28	2532.77	4.71	.0001
Error (Property* Tasks)	446	537.74		

Let us now see exactly where the learners' performance on properties differed based on the experimental task. Starting with the properties subsumed under the verb movement parameter, let us present all the task effect significance levels by property in Table 14 (AdvInf is omitted since it was only tested in Task 4).

Table 14. Task Effect by Verb Movement Property

Property	*df*	*F*	*pr* >F
NegFin	(3,45)	.50	.6833
NegInf	(3,48)	7.74	.0003
AdvFin	(3,45)	8.90	.0001
FQ	(3,48)	4.86	.0049
ProInv	(3,48)	3.44	.0239

The following tables will now show the means obtained on each task and the results of post hoc Tukey tests in the form of groupings. Means with the same letter under Tukey groupings are not significantly different. All the means for NegFin--negation placement in finite contexts--in Table 15 are thus in the same group: the participants performed quite consistently from task to task.

Table 15. Means and Tukey Groupings for NegFin

NegFin Tukey grouping	Mean	*SD*	Task
A	76.48	22.86	Task 4 (ProdT)
A	81.25	28.00	Task 3 (S-GJT)
A	82.34	12.44	Task 5 (ME-AJT)
A	86.28	26.50	Task 2 (PrefGT)

On the other hand, Table 16 indicates that the learners' performance was significantly different between the production task (Task 4) and the grammaticality/preference task (Task 2).

Table 16. Means and Tukey Groupings for NegInf

NegInf Tukey grouping	Mean	*SD*	Task
A	82.35	34.76	Task 4 (ProdT)
A	65.68	28.00	Task 5 (ME-AJT)
B	52.91	35.58	Task 3 (S-GJT)
B	48.52	26.65	Task 2 (PrefGT)

The participants did not perform as consistently on the placement of adverbs in finite contexts (AdvFin) but there are no significant differences between tasks due to the high standard deviations, particularly on Task 2.

Table 17. Means and Tukey Groupings for AdvFin

AdvFin Tukey grouping	Mean	*SD*	Task
A	41.17	30.45	Task 2 (PrefGT)
A	68.75	27.29	Task 3 (S-GJT)
A	70.58	23.77	Task 4 (ProdT)
A	73.53	24.33	Task 5 (ME-AJT)

Tukey groupings in Table 18 for floating quantifiers indicate that the L2 learners' performance was significantly different on only two tasks: Task 4 (ProdT) and Task 3 (S-GJT).

Table 18. Means and Tukey Groupings for FQ

FQ Tukey grouping	Mean	*SD*	Task
A	88.24	20.20	Task 4 (ProdT)
B A	69.63	22.24	Task 5 (ME-AJT)
B	66.17	31.79	Task 2 (PrefGT)
B	55.88	30.01	Task 3 (S-GJT)

We obtained similar results for the last verb movement property, inverted questions (proinv), for which a significant difference was found between the production task (Task 4) and the preference/grammaticality judgment task (Task 2) as displayed in Table 19.

Table 19. Means and Tukey Groupings for ProInv

ProInv Tukey grouping	Mean	*SD*	Task
A	100.00	.0	Task 4 (ProdT)
B A	94.11	16.60	Task 5 (ME-AJT)
B	92.64	17.14	Task 3 (S-GJT)
B	75.00	41.45	Task 2 (PrefGT)

These means and Tukey groupings for the verb movement properties thus clearly indicate that learners' performance may vary widely depending on the elicitation task as illustrated in Figure 3.

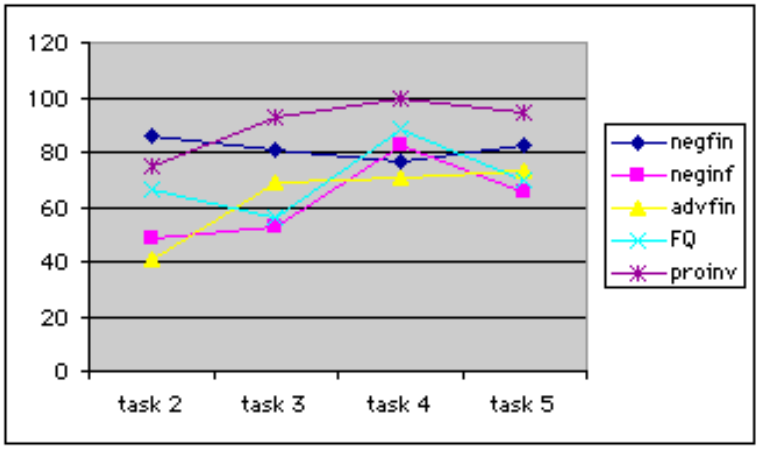

Figure 3. Verb movement properties by tasks

Let us now consider the results for the properties subsumed under the null subject parameter. Again, we first present the results of the ANOVA in Table 20 before showing the means and Tukey groupings for each property.

Table 20. Task Effect by Null Subject Property

Property	*df*	*F*	*pr* >F
ProSub	(3,45)	1.15	.329
ProEmb	(3,48)	5.07	.0003
Expl	(3,48)	1.78	.1639
Extr	(3,48)	7.31	.0004
NomSub	(3,48)	4.63	.0171
NomInv	(3,48)	5.30	.0031

Table 20 indicates that the learners' performance was significantly different from task to task for four properties: embedded null subjects (ProEmb), extraction (extr), nominal subjects (NomSub) and nominal inversion (NomInv). Let us examine the means and Tukey groupings for each property.

The participants did not do very well on null subjects in main clauses particularly on the preference/ grammaticality task (Task 2) as shown in Table 21. However the means for the Tukey groupings have the same letter indicating that there is not a significant difference due to the extremely high standard variation.

Table 21. Means and Tukey Groupings for ProSub

ProSub Tukey grouping	Mean	*SD*	Task
A	75.49	26.43	Task 3 (S-GJT)
A	70.59	39.76	Task 5 (ME-AJT)
A	57.35	49.81	Task 2 (PrefGT)

Our next table, Table 22, shows that the participants did extremely well on all tasks on null subjects in embedded clauses with the exception of the preference/grammaticality task which is significantly different from the production task.

Table 22. Means and Tukey Groupings for ProEmb

ProEmb Tukey grouping	Mean	*SD*	Task
A	100.00	.0	Task 4 (ProdT)
B A	89.70	15.45	Task 5 (ME-AJT)
B	84.32	17.13	Task 3 (S-GJT)
B	76.47	31.21	Task 2 (PrefGT)

The participants did extremely well on expletives as well as shown on Table 23 which reveals high means on all tasks. It is interesting to note that participants obtained the same mean on the preference task and the scaled-GJT, but show a greater standard variation on the former.

Table 23. Means and Tukey Groupings for Expl

Expl Tukey grouping	Mean	*SD*	Task
A	98.82	4.85	Task 4 (ProdT)
A	88.23	17.93	Task 3 (S-GJT)
A	88.23	26.68	Task 2 (PrefGT)
A	86.27	15.84	Task 5 (ME-AJT)

The means and Tukey groupings displayed in Table 24 for extraction show again a significant difference between the preference task and the production task, but this time, the participants performed much worse on the latter than on the former.

Table 24. Means and Tukey Groupings for Extr

Extr Tukey grouping	Mean	*SD*	Task
A	72.55	31.70	Task 2 (PrefGT)
A	56.86	40.42	Task 3 (S-GJT)
A	52.95	44.18	Task 5 (ME-AJT)
B	20.59	22.07	Task 4 (ProdT)

The Tukey groupings displayed in Table 25 indicate that the participants performed significantly differently on the two A/GJTs: they obtained a perfect score of 100% on the magnitude estimation AJT

but did not do as well on the scaled GJT with an average of 85.29%. Thus, although the high means show that the property has been acquired, different experimental tasks yield a significantly different performance.

Table 25. Means and Tukey Groupings for NomSub

NomSub Tukey grouping	Mean	*SD*	Task
A	100.00	.0	Task 5 (ME-AJT)
B A	97.05	8.3	Task 2 (PrefGT)
B	85.29	23.48	Task 3 (S-GJT)

And finally, the means and groupings presented in Table 26 reveal a very good performance on Task 4, the production task, which is significantly different from the Task 2, the preference/grammaticality task.

Table 26. Means and Tukey Groupings for NomInv

NomInv Tukey grouping	Mean	*SD*	Task
A	97.05	12.12	Task 4 (ProdT)
B A	78.42	19.32	Task 5 (ME-AJT)
B	74.51	25.08	Task 3 (S-GJT)
B	61.76	43.40	Task 2 (PrefGT)

Figure 4 sums up the L2 learners' performance on each property from task to task.

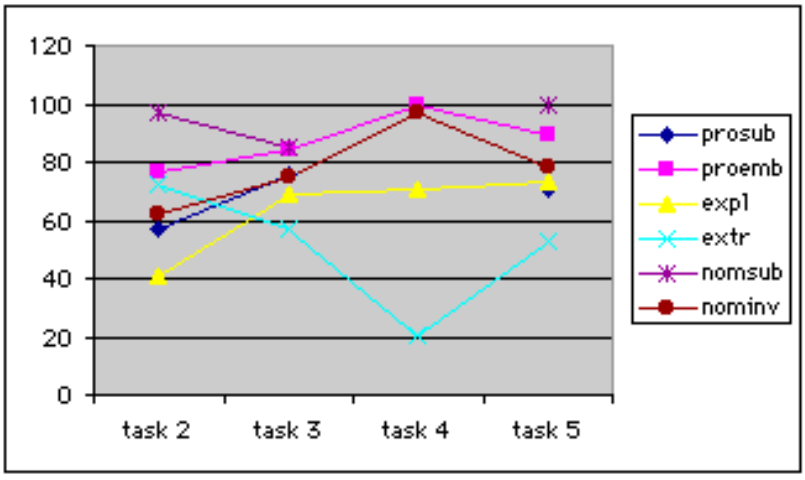

Figure 4. Null subject properties by tasks

The only generalization which holds across all syntactic properties is that the learners' best performance was on the production task and their worst performance was on the preference/ grammaticality task. But

even that generalization is not without a puzzling exception: the fact that learners were able to select and judge stimuli exemplifying extraction with a 72.55% accuracy on the preference/grammaticality task but performed so poorly on the production task (20.59%). Hypothesis 3, which stated that would be a positive relationship between learners' performance and task types, is therefore rejected. The learners did not consistently perform across task types. This finding raises important questions for parameter setting theory which will have to be addressed in future research.

DISCUSSION AND CONCLUSION

It was hypothesized that the syntactic properties subsumed under the null subject and the verb movement parameters would be acquired as a cluster. It is unclear why results revealed that this was the case for the verb movement parameter, which requires a resetting from the L1 value to the L2 value, but not for the null subject which does not require such a resetting, since the L1 and the L2 have the same value. However, this indicates that L2 learners do not simply rely on the L1 and that other psycholinguistic processes are at work.

It was argued that a variety of Web-based elicitation tasks would produce a richer set of data and would be more informative and dependable, and it was indeed the case. First, a study limited to a single traditional scaled GJT would be less informative and would even lead to inaccurate conclusions since it was found that participants performed better on all other tasks revealing a higher level of proficiency, which would otherwise have gone unnoticed. Furthermore, and although no significant difference was found for all properties confounded, most participants (11 out of 17) obtained better results on Task 5 than on Task 3, showing that learners actually perform differently on different types of GJTs/AJTs. And finally, almost as many participants (9 out of 17) obtained their highest score on the production task. We may thus establish a decreasing order of difficulty across all properties: production task (4), magnitude estimation GJT (5)/scaled GJT (3), preference/ grammaticality task (2).

When we considered specific properties for which significant differences were found among tasks we discovered that indeed learners performed better on the production task than on any other task. A particularly important finding was the significant difference on nominal subjects between the two types of judgment tasks, the magnitude estimation AJT and the scaled GJT. A flawless performance of 100% accuracy on the former was significantly different from a very respectable performance of 85.29% on the latter.

Let us be more specific about the various tasks. It seems that participants performed best on the tasks which imposed the least restrictions and were the least artificial. The production task required the participants to translate individual sentences from English into French. This is an exercise the participants may have experienced as students in their respective language classes, although it is certainly not an easy task. The stimuli were controlled for length and simplicity of vocabulary to avoid any unnecessary distraction. The relative open-endedness of the production task may be a good way to actually tap into the learners' competence since it requires them to directly compare the L1 and the L2. The comparison either yields a corresponding structure in their interlanguage grammar or it does not. In the latter case, learners are likely to "fall back" on their L1. On the other hand, learners are very unlikely to be able to correctly produce a structure which is not part of their interlanguage grammar. The drawback of a production task is that it does not rule out the possibility that learners know that the L2 (dis)allows alternatives. This is where acceptability/grammaticality judgment tasks (A/GJT) come into play.

A/GJTs differ from production tasks in several ways. First, they may require less processing time. This may allow learners to access the knowledge stored in long term memory. Second, the artificiality of a A/GJT probably has an impact on learners, which may explain why native speakers rarely perform at a 100% accuracy either. Although there are several other possible explanations such as dialectal and/or individual variations or "fuzzy" grammaticality of the syntactic structures being tested (Sorace, 1996). In

addition, "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" (Schütze, 1996, p. 169).

The preference/grammaticality task showed that even when they were capable of correctly producing a given structure in task 4, L2 learners were not always able to both indicate their preference and correctly determine whether the other sentence was (un)grammatical. As a matter of fact, a significant difference was found between Task 2 and Task 4 on two verb movement properties and three null subject properties.

It may thus be concluded that the various elicitation tasks used in this study provided a better insight into the learners' interlanguage than a single traditional GJT (Chapelle, 1998) in that they yielded a richer and more informative set of data. Web-based elicitation tasks definitely offer a greater versatility. This increases the generalizability of empirical studies and our understanding of second language acquisition processes.

APPENDIX
Individual Results by Task [Return]

L2 FRE	Task 1 PreTest	Task 2 PrefGT	Task 3 S-GJT	Task 4 ProdT	Task 5 ME-AJT	average
kov	44.0%	73.5%	60.0%	74.3%	67.6%	68.8%
hed	51.5%	53.0%	58.7%	77.1%	70.0%	64.7%
mar	53.5%	68.0%	71.3%	86.0%	96.5%	80.5%
mas	56.0%	44.0%	48.0%	77.0%	70.5%	60.0%
mcc	58.5%	82.3%	67.1%	77.0%	84.5%	77.7%
smi	59.0%	55.9%	69.0%	85.7%	67.8%	69.6%
gru	61.0%	73.5%	65.5%	63.0%	71.9%	68.5%
mor	63.2%	74.0%	57.5%	63.0%	61.3%	64.0%
bie	63.5%	61.8%	55.4%	77.1%	69.1%	65.8%
kaz	70.5%	71.0%	86.0%	86.0%	90.6%	83.4%
bro	70.8%	53.0%	80.4%	91.4%	79.1%	76.0%
cha	71.0%	85.3%	85.6%	83.0%	80.4%	83.6%
rac	75.1%	55.9%	76.3%	88.6%	79.8%	75.2%
bur	80.4%	97.0%	85.6%	88.6%	83.4%	88.6%
tru	83.2%	91.2%	92.9%	85.7%	82.6%	88.1%
cox	87.8%	82.3%	91.0%	91.4%	89.6%	88.6%
phi	90.5%	76.5%	59.2%	71.4%	74.0%	70.3%
average	67.9%	70.5%	71.1%	80.4%	77.6%	74.9%

NOTES

1. I am most grateful for the insightful suggestions, constructive comments, as well as the technical and statistical support, the following individuals provided: Roby Ariew, Robert Bley-Vroman, J.D. Brown, Adam Dudsic, George Gutsche, Pat Jones, Kazue Kanno, Gabi Kasper, William O'Grady, and four extremely thorough anonymous reviewers. The usual disclaimers apply.

2. The verb movement parameter is alternatively referred to as the V-Raising parameter (Culicover, 1997), the V-to-I parameter (Deprez, 1994), or the (strength of) AGR parameter (Williams, 1994), since the systematic differences between the two languages most often studied, French and

English, depend on whether overt verb movement is allowed or disallowed due to the [±strong] feature of AGR.

3. This does not necessarily imply that deductive processes are replaced with inductive processes.
4. I would like to very sincerely thank Dr. Chris Johnson for the expert programming of this task, and Adam Dudsic for invaluable technical assistance throughout the data collection.
5. The properties are notoriously difficult for L2 French learners and are often used to determine the learners' level of proficiency (Coppieters, 1987; Birdsong, 1992).
6. Gass (1983) showed that providing a correction to an ungrammatical or unacceptable sentence requires a very different metalinguistic ability than making a holistic judgment. See Birdsong (1989) for the methodological desirata of requiring participants to provide corrections following grammatical judgments.
7. Documents created with Director® can also be made available on a Web page so that they can be easily accessed from any computer equipped with a Web browser.
8. There was also a wide range of variation among participants with a lowest score of 44% and a highest score of 90.5%.

ABOUT THE AUTHOR

Dr. Ayoun is currently Assistant Professor of French Language & Linguistics at the University of Arizona, and Graduate Faculty in the Interdisciplanary PhD program in Second Language Acquisition and Teaching. Her research focuses mainly on the applied pyscholinguistics processes underlying adult second language acquisition.

REFERENCES

Ayoun, D. (1998a). *The verb movement parameter revisited: syntactic theory and language data.* Unpublished manuscript, University of Arizona.

Ayoun, D. (1998b). Visiting traditional assumptions in Principles-and-Parameters theory. *Arizona Working Papers in SLAT, 6*, 61-79.

Ayoun, D. (1999). The verb movement parameter in French L2 acquisition. *Bilingualism: Language & Cognition, 2*, 103-125.

Ayoun, D. (2000). *Parameters of Universal Grammar in first and second language acquisition.* Manuscript in progress.

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

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

Birdsong, D. (1992). Ultimate attainment in second language acquisition. *Language, 68*, 706-55.

Bley-Vroman, R. (1998). *Access to UG: can we have our cake and eat it too?* Colloquium presentation at the Second Language Research Forum, Honolulu, HI.

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

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

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

Chomsky, N. (1981). *Lectures on government and binding.* Dordrecht: Foris.

Chomsky, N. (1995). *The Minimalist Program.* Cambridge, MA: MIT Press.

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

Doughty, C. (1991). Second language instruction does make a difference. Evidence from an empirical study of SL relativization. *Studies in Second Language Acquisition, 13,* 431-469.

Emonds, J. (1978). The verbal complex V'-V in French. *Linguistic Inquiry, 9,* 49-77.

Fukui, N. (1995). The principles-and-parameters approach: a comparative syntax of English and Japanese. In M. Shibatani & T. Bynon (Eds.), *Approaches to language typology* (pp. 327-372). Oxford: Clarendon Press.

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, NJ: Erlbaum.

Hasegawa, N. (1985). On the so-called "zero-pronouns" in Japanese. *The Linguistic Review, 4,* 289-341.

Huang, Y. (1989). *Anaphora in Chinese: toward a pragmatic analysis.* Doctoral dissertation. University of Cambridge.

Hulstijn, J. (1997). Second language acquisition research in the laboratory. Possibilities and limitation. *Studies in Second Language Acquisition, 19,* 131-43.

Jaeggli, O. & Safir, K. (Eds.). (1989). *The null subject parameter.* Dordrecht, The Netherlands: Kluwer.

Lardiere, D. (1998). Dissociating syntax from morphology in a divergent L2-end state grammar. *Second Language Research, 14,* 359-375.

Meisel, J. (1998). *Revisiting Universal Grammar.* Paper presented at the Second Language Research Forum, Honolulu, HI.

Perlmutter, D. (1971). *Deep and surface constraints in syntax.* New York: Holt, Rinehart & Winston.

Platt, E. (1993). Parameter resetting in second language acquisition: a study of adult Spanish and Vietnamese learners of English. In F. Eckman (ed.), *Confluence. Linguistics, L2 acquisition and speech pathology* (pp. 105-134). Amsterdam: John Benjamins.

Pollock, J.-Y. (1989). Verb movement, Universal Grammar, and the structure of IP. *Linguistic Inquiry 20,* 365-424.

Pollock, J.-Y. (1995). Talk given at the Fourth Diachronic Generative Syntax Conference. Université du Québec à Montréal.

Pollock, J.-Y. (1997). *Langage et cognition. Introduction au programme minimaliste de la grammaire générative.* Paris: Presses Universitaires de France.

Roberts, I. (1998). Have/Be raising, Move F, and Procrastinate. *Linguistic Inquiry, 29,* 113-125.

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.

Schwartz, B. & Tomaselli, A. (1990). Some implications for an analysis of German word order. In W. Abraham, W. Kosmeijer, & E. Reuland (Eds.), *Issues in Germanic syntax* (pp. 251-274). New York: Mouton de Gruyter.

Schwartz, B. (1998). *Back to basics.* Paper presented at the Second Language Research Forum, Honolulu, HI.

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, CA: Academic Press.

Uziel, S. (1993). Resetting Universal Grammar parameters: evidence from second language acquisition of Subjacency and the Empty Category Principle. *Second Language Research, 9,* 49-83.

White, L. (1996). Universal Grammar and second language acquisition: current trends and new directions. In W. Ritchie & T. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 85-120). San Diego, CA: Academic Press.

White, L. (1998). *Impaired or unimpaired representation in SLA: evidence from functional categories and features.* Colloquium presentation at the Second Language Research Forum, Honolulu, HI.