Chapter 5

Results and Interpretation

5.0 Introduction

This chapter presents the results obtained from the translation task, the pilot task, and the main grammaticality judgment task. Section 5.1 discusses the coding and data analysis procedures used for the translation task, and it presents the results of this task. Analyzing the translation task posed several challenges, which are discussed in this section. The statistically valid findings of this test are very general, requiring the further refinement that the grammaticality judgment task yielded. Section 5.2 reports on the result of the pilot, focusing mostly on the qualitative issues specific to the purposes of the pilot. Section 5.3 discusses the coding and data analysis procedures used for the grammaticality judgment task and then presents the statistical results of this task. The significant differences between the responses of the various proficiency levels are discussed, organized around the three grammatical properties of null subjects (Section 5.3.1), inversion (Section 5.3.2) and *that-trace* (Section 5.3.3). Section 5.4 interprets these results in light of the implicational hierarchy of Liceras (1989), the issue of ultimate attainment, and the impact of discourse on grammatical choices, leading into the analysis provided in the next chapter. Finally, Section 5.5 briefly discusses limitations of the study revealed by the results.

5.1 Translation task

In the last chapter, we noted that the translation task involved participants from four proficiency levels (beginning, intermediate, advanced, and native speakers) and that each participant was asked to translate two dialogues. Ideally, for the purpose of statistical analyses, it would have been better to have each item translated by each subject, but the time involved in having a single subject translate 36 dialogues would be unreasonable; therefore, each subject (n=124) translated only two dialogues, yielding a total of 248 tokens for analysis. Each of the 36 items was translated at least twice by each proficiency level. These translation tokens were then coded with respect to the preferred responses by native speakers. For example, if native speakers consistently used a null subject in the translation of a particular sentence, then translations by other proficiency groups that lacked a null subject in that context were marked with a '1', indicating a dispreferred translation. If the subject translated the sentence using the preferred translation, it was coded with a '0' to indicate no variance from the native speaker norm; therefore, the coding yielded binary responses (preferred/dispreferred).

To model the binary responses here, the first attempt was to code them as 'logits', or 'logs of odds ratios'. An odds ratio measures the probability of a preferred response versus the probability of a dispreferred response under two different sets of conditions (e.g. *that-trace* and inversion). The analysis was intended to be a logistic regression with two factors. The responses were not independent, since there are two responses from each subject, so it was necessary to include 'subject' in the model as a clustering variable

(or random effect) to induce correlation among the responses for each subject. To do this, the SAS statistical package was used, employing PROC GENMOD.

Analysis of the items for this test proved problematic. This repeated measures model treated the responses of each subject as correlated data, not independent data. In all of the models used for these studies, in which subjects responded to multiple items, it was assumed that the items are correlated within the subject. The initial attempt was to analyze item and group as factors. For the binary response model, however, if any of the items or any of the groups had only one of the two responses present for all subjects who tested on that item or belonged to that group (for example, if for item 16, there were no '1's, hence only '0's), then the algorithm failed to converge. There were many such items in this case, and no further attempt was made to include item as a factor.[1]

Although the translation task did not reveal significant information regarding items, it did reveal that groups differed significantly in their response patterns. Table 5.1 shows the analysis of initial parameter estimates related to each group.

Table 5.1: Analysis of initial parameter estimates for translation task, by group

Parameter	DF	Estimate	Std. Error	Wald 95%	Confidence Limits	Chi-Square	Pr > ChiSq
Beginning	1	-3.5041	0.6376	-4.7535	-2.2544	30.20	<.0001
Intermed.	1	-2.7675	0.6377	-4.0174	-1.5175	18.83	<.0001
Advanced	1	-1.4077	0.6627	-2.7065	-0.1088	4.51	0.0337
Native	0	0.0000	0.0000	0.0000	0.0000	-	-

Table 5.1 sets the native speaker group as the baseline. The 'DF' designator in column 2 refers to the degrees of freedom, or the number of independent pieces of information that go into the estimate of a parameter. The test permitted learner groups to vary from the native group by at most one degree, because for each item the evaluation

was binary, based on whether the translation for an item did or did not use a null subject, inversion, or *that-trace*. The remaining columns show that native speaker response patterns are distinct from all other groups, including the advanced group (although evidence for the difference between advanced and native is marginal). The parameters are listed as log odds ratios. The value of –3.5041 in column 3 for the beginning group indicates that the odds of a preferred response for the native group is exp(3.5041) = 33 times greater than the odds of a preferred response for the beginning group. In other words, natives use the grammatical constructions under investigation much more frequently than beginners. The bounds on these odds are obtained from the Wald 95% Confidence Limits (exp(2.2544), exp(4.7538) = (9.5, 166)). We may be 95% confident that the odds of a native speaker giving a preferred response are 9.5 to 166 times greater than that of a beginning speaker. Intermediate speakers were 15 times less likely to give a preferred response than native speakers. The Wald 95% Confidence Limits for this group ranged from 4.5 to 55.5 times less likely to give a preferred response.

The difference between native and advanced groups did not achieve significance under this test, largely because the Wald limits ranged from 1.1 to 14.9 times less likely for the advanced speakers to give a preferred response. The performance of the advanced group was, however, distinct from the native group, with $p = 0.0337$, and an odds ratio 4 times less likely than natives to provide the preferred response.

The results of the translation task reveal that the proficiency groups tested here do vary in regards to their ability to translate dialogues in a native-like manner with respect to null subjects, inversion, and *that-trace*. This study also revealed progressive movement towards native patterns, with the advanced level as a group differing from

native speakers in a less than significant way. Nevertheless, this study was not designed to indicate the exact loci of the differences in regards to items.

5.2 Pilot study

As discussed in the previous chapter, the purpose of the pilot was to assess the time required to administer the test, check the clarity of the instructions, and uncover difficulties related to vocabulary choice or ambiguities in the discourse context — in other words, to refine the test that would be used for the main study. Since numerous items were changed as a result of the pilot, and a comparison of different versions of the same item would not be valid, a separate statistical package was not run on the pilot. A description of the results, however, is reported here to give a sense of the directionality of responses, along with qualitative results related to the purpose of the task.

Subjects for the pilot study asked very few questions after the form was read and they appeared to clearly understand the directions. Less than 2% of the forms were inaccurately completed (i.e. most subjects correctly checked one and only one box, and did not skip any items). In a few instances, subjects began the test but chose not to finish it. These tests were excluded from the sample. There were also several tests turned in by subjects who reported a language other than English as their L1. These tests were also excluded from the sample. On a number of the tests, subjects marked vocabulary items that they did not understand. These items were replaced with more common vocabulary in the revised test. Following the pilot, two instances of advanced verbal forms were found, as well as a few typographical errors. All of these were corrected in the revised version.

The intermediate learners in the pilot normally selected null subjects when given the choice between a null and overt subject. This was true even for dialogues where an overt subject is preferred by native speakers.[2] For example, for dialogues 17 and 28 below, the preferred native response is the 'A' choice, identified by a check mark, and native speakers made this choice about 90% of the time. In contrast, intermediate level subjects chose the 'A' response only about 50% of the time.[3]

(17) *Después de un paseo por el museo*[4]
    Luis: Fui al museo esta mañana.
    Rosa: ¿Y qué viste?
    Luis: Muchas pinturas de Picasso.
    Rosa: A mí me gusta mucho Picasso.
    Luis: A mí también; y por eso compré un póster en la tienda de regalos.
    Rosa A: Yo también compré uno la semana pasada.   √
    Rosa B: También compré uno la semana pasada.

(28) *Haciendo una acusación*
    Carmen: ¿Entonces, qué está diciendo?
    Rosa: Juana y usted estaban ahí en ese momento.
    Carmen: ¿y?
    Rosa: Que Juana y usted tenían una razón para hacerlo.
    Carmen: Otro lo hizo.
    Rosa A: Ustedes lo hicieron. Nadie más.   √
    Rosa B: Lo hicieron. Nadie más.

Intermediate speakers also registered a strong preference for non-inverted choices, and, again, registered a roughly 50% split even on items for which the preferred native choice was to invert nearly 97% of the time, as in dialogue 4 below:

(4) *Asumiendo responsabilidad*
    Pablo: ¿De qué te ríes?
    Janet: Mamá y Papá van a llegar pronto.
    Pablo: ¡Ay! ¡La casa está muy sucia!
    Janet: Te dije: 'No invites a tus amigos.'
    Pablo: ¿Me ayudas a limpiar la casa?
    Janet A: ¡No! Tienes que limpiarla tú, no yo. √
    Janet B: ¡No! Tienes que limpiarla, no yo.

The pilot also revealed that intermediate learners had great difficulties with the correct distribution of *que* 'that' in these dialogues. For example, in dialogues such as 1 and 19 below, native speakers had a strong preference for the 'B' response (87% and 84%, respectively), but the intermediates in this pilot chose 'B' only 13% of the time for item 19 and not at all for dialogue 1.

(1) *Después de la matrícula en la escuela*
    Alicia: ¿Te has inscrito en tus clases?
    Marta: Sí, pero tengo un problema.
    Alicia: ¿Qué pasó?
    Marta: No pueden encontrar mi cheque.
    Alicia: Se han equivocado con mi cuenta también.
    Marta A: ¿Quién piensas nos puede ayudar?
    Marta B: ¿Quién piensas que nos puede ayudar?   √

(19) *En el aeropuerto*
    Rosalía: ¡Hola! Roberto.
    Roberto: Bienvenida.
    Rosalía: Estoy muy cansada.
    Roberto: No tendrás mucho tiempo para descansar.
    Rosalía: ¿Por qué? Voy a casa.
    Roberto A: ¿Quién esperas va a trabajar esta noche?
    Roberto B: ¿Quién esperas que va a trabajar esta noche?   √

Intermediates were willing to accept *that-trace* violations under certain conditions. For example, for dialogues 14 and 36 (below) intermediates chose the *that-trace* violation at rates of 61% and 69%, respectively.

(14) *Una mujer habla por télefono con su marido*
    Carmen: ¿Hola?
    Felipe: Buenos días, Carmen.
    Carmen: ¿Visitaste al médico? ¿Tienes noticias?
    Felipe: Sí, esta mañana.
    Carmen: ¿Qué te dijo?
    Felipe A: Que no sabe nada todavía.   √
    Felipe B: Que él no sabe nada todavía.

(36) *Haciendo planes para el fin de semana*
      Laura: *Estudias demasiado, Felicia.*
      Felicia: *Tengo un examen en la clase de filosofía.*
      Laura: *Estudiaste toda la semana y el fin de semana pasado.*
      Felicia: *Voy a tomar un descanso pronto.*
      Laura: *¿Vamos a la playa mañana?*
      Felicia A: *No creo yo pueda ir.*
      Felicia B: *Yo no creo que pueda ir.*   √

Interestingly, the pattern of responses here suggests that intermediate learners' difficulty with the distribution of *que* in Spanish may not rest with the violation of *that-trace* but with identifying sentences where *que* is required. This is most clearly seen in dialogues such as 14. This also seems to hold for dialogue 36; however, it should be noted that in that dialogue learners may have been reacting more against the inverted subject than for the *that-trace* sequence.

      To summarize the results of the pilot: subjects in this study appeared to have acquired null subjects, but lacked a sensitivity to discourse conditions that may regulate their use. They displayed dispreference for inversion, preferring non-inverted orders even for sentences where native speakers prefer inversion, although they did select inversion more often when the inverted item was focused. Intermediates had problems with sentences involving t*hat-trace* violations, but it is not clear from this study whether this problem is one of possessing a grammar prohibiting *that-trace* or whether the difficulty is rather one of not recognizing the obligatoriness of *que* in particular contexts.

5.3 Grammaticality judgment task

      This section reports on the grammaticality judgment task which, as noted previously, consisted of learners making binary choices on 36 items differing in the use

of null subjects, inversion, and *that-trace*. There were two forms of this test used for the main study; each presented items in different random orders. Analysis of the results of the two forms of the test found no statistical difference between them, so the order of the presentation of the items was not significant.

The administration of the pilot revealed that the test would take 15–20 minutes for intermediate learners to complete. It was then assumed that, for the main grammaticality judgment task, beginning learners would take slightly longer, and advanced learners slightly less time. This proved to be correct. Advanced speakers and above usually completed the test in 10–15 minutes, while beginners normally took 20–30 minutes. A common consent form was used for all levels, so as a result of the pilot, subjects were all told that the test should be completed in 20–30 minutes. The consent form was refined in two additional ways. First, the original did not inform subjects that as they completed the test they might experience some fatigue, boredom, or frustration. These reactions were observed in the pilot group, but the presence of the advisory in the subsequent tests appeared to mitigate these problems.

A total of 207 subjects responded to each of the 36 dialogues, yielding a total of 7452 tokens for analysis. Each of these tokens was coded with a '0' if the respondent chose choice 'A' and a '1' if the respondent chose choice 'B'. These responses were then analyzed compared to the baseline responses of the native group. An analysis of initial parameter estimates (Table 5.2) yielded results similar to those of the analysis of the translation task:

Table 5.2: Analysis of initial parameter estimates for gramm. judgment task, by group

Parameter	DF	Estimate	Std. Error	Wald 95%	Confidence Limits	Chi-Square	Pr > ChiSq
Beginning	1	-0.5142	0.0784	-0.6678	-.03606	43.06	<.0001
Intermed.	1	-0.3428	0.0814	-0.5024	-0.1832	17.72	<.0001
Advanced	1	-0.3649	0.0801	-0.5219	-0.2079	20.75	<.0001
Near-Nat.	1	-0.5030	0.1551	-0.8070	-0.1990	10.52	0.0012
Native	0	0.0000	0.0000	0.0000	0.0000	-	-

Table 5.2 shows that native speaker response patterns were again significantly distinct ($p <.0001$) from beginning and intermediate groups; this test additionally showed significance between the advanced and native groups ($p <.0001$) and a clear, but not significant, difference between the near-native group and the native speakers ($p=0.0012$). The log odds ratios in the grammaticality judgment task were smaller than in the translation task, but the larger number of respondents and the fact that each subject responded to each item still permitted valid analysis of significances.

The grammaticality judgment task shared one problem with the translation task in that running a global model to analyze item and group parameters simultaneously did not work due to some groups always getting the same number on an item. Fortunately, there are other default methods that can be used to test for item–by–group differences. The standard way of checking whether there is a group effect on a particular item is to use a Fisher Exact Probability Test, a useful alternative to chi-square when that test would not yield valid results. The Fisher test involves a more complicated formula than chi-square, but the Fisher test has an important advantage over the chi-square test in that does not involve approximations; rather, it calculates exact probabilities whatever the sample size. Given the test design here, the Fisher test was able to reveal significant interactions between groups and items. A conservative alpha level was set for this test by inflating

the overall experimental wide error to .20 and dividing this number by the total number of comparisons (36). This set the significance level to $\alpha < .006$.

The basic approach was to question whether choice depended on group in a significant way. Using the Fisher test, significant interactions were found for 18 of 36 items. In an additional 13 of the 36 items, there were observable differences between responses that approached, but did not reach, the conservative significance threshold set for this test. Responses to only 5 of the 36 items displayed little or no difference between all levels. In general, acceptance of null subjects surfaced early, but was overgeneralized until the more advanced levels. Acceptance of inversion and *that-trace* effects clearly surfaced later, but the results of the inversion items are somewhat unclear, as we shall see.

5.3.1 Null subject results

There were two types of items related to null subjects: those requiring that topic-connected null subjects be dropped and those requiring that subjects that are not topic-connected be retained. Since English and Spanish differ only in regards to the former condition, the majority of the items in this task tested for the dropping of null subjects, but the task also included some items that tested for the retention of overt subjects in nontopic contexts. English and Spanish do not differ in their grammatical requirements regarding these contexts, so any differences surfacing for this condition would be interesting.

The results of the items testing for the topic subjects are given in Table 5.3, which

provides the raw percentages of respondents choices (selecting null subjects) on each

item by proficiency group:

Table 5.3 Percentage of choice for null subject, by item and proficiency group

Item	Beginning	Intermediate	Advanced	Near-Native	Native
10	0.5938	0.6667	0.8571	1.0000	0.8000
12	0.7344	0.7059	0.9286	0.8333	1.0000
14	0.5156	0.7255	0.5000	0.9333	0.8000
16	0.6406	0.6078	0.8750	0.8333	0.9333
18	0.6094	0.5490	0.6786	1.0000	0.7000
20	0.4375	0.4706	0.7857	0.8333	0.9333
31	0.4531	0.4706	0.4464	0.6667	0.6667
34	0.6719	0.6471	0.8393	0.8333	0.7333
Avg.	.58	.60	.74	.86	.82

For example, for dialogue 20 (below), native speakers preferred the null subject

choice 93% of the time, while beginners and intermediates registered this preference only

43% and 47%, respectively.

(20) *En la universidad*
    Simón: Fue muy difícil esa clase de biología.
    Adriana: Ah, ¿sí? ¿Qué estudias?
    Simón: Historia. No me gustan las ciencias.
    Adriana: Qué lástima. Las ciencias pueden ser muy interesantes.
    Simón: ¿Cómo te interesaste por las ciencias?
    Adriana A: Tuve una maestra buenísima en la escuela secundaria.   √
    Adriana B: Yo tuve una maestra buenísima en la escuela secundaria.

The level of difference between native speakers and both beginners and intermediates for

dialogue 20 is highly significant. Fishers Exact Test reveals a significance level of p

$=.000001$ between advanced and beginners and $p = .000002$ between advanced and

intermediates. Similarly, item 12 (below) registered differences between the beginning

and intermediate groups compared to the native speakers. Beginners differed from native

speakers by selecting a null subject only 73% of the time, while natives always selected

the null choice. Intermediates also judged this sentences differently from natives,

selecting null subjects only 70% of the time.

(12) Pa*blo y Rosa en casa*
     Rosa: *Pablo, ¿has visto mi dinero?*
     Pablo: *¿Dónde lo dejaste?*
     Rosa: *Estaba aquí en la mesa.*
     Pablo: *Mira debajo de esos documentos.*
     Rosa: *¡Ah, lo encontré!*
     Pablo A: *Debes cuidar bien tu dinero.*   √
     Pablo B: *Tú debes cuidar bien tu dinero.*

The differences between the beginning and intermediate groups and native speakers

registered significance levels of $p = .001$ and $p = .00006$, respectively.

     The results on these items support a general pattern found also in other items (10,

16, 34) where beginners and intermediates fail to delete topic subjects at a rate similar to

the other three proficiency levels. This is generally true for the remaining topic-subject

items as well, since beginners and advanced learners in this study never deleted topic

subjects as often as native speakers, but the level of difference for these remaining items

fails to achieve significance.

     Three items (17, 28, and 35) were predicted to require the retention of overt

subjects. For two of these items (17 and 28), the predicted result was obtained. On the

third item (35), native speakers did not perform as predicted, choosing the overt subject

only 40% of the time. Table 5.4 gives the percentages of choice for overt subject on each

item by proficiency group:

160

Table 5.4 Percentage of choice for overt subject, by item and proficiency group

Item	Beginning	Intermediate	Advanced	Near-Native	Native
17	0.6406	0.5882	0.6786	0.6667	0.8333
28	0.5781	0.4902	0.6607	1.0000	0.9333
35	0.1563	0.3333	0.6250	0.8333	0.4000
Avg.	.45	.47	.65	.83	.72

Items 17 and 28 (below) both supplied contexts in which contrastive or emphatic use of subject pronouns triggers overt realization. Native speakers chose overt subjects for these contexts at rates of 83% and 93%, respectively. In contrast, beginners selected overt subjects less frequently, 64% of the time for item 17 and 58% of the time for item 28. The difference from native speakers does not reach significance for item 17 ($p =$ .0891), but does for 28 ($p = .0003$) (cp. with the results of the pilot on pp.155-156).

(17) *Después de un paseo por el museo*
    Luis: Fui al museo esta mañana.
    Rosa: ¿Y qué viste?
    Luis: Muchas pinturas de Picasso.
    Rosa: A mí me gusta mucho Picasso.
    Luis: A mí también; y por eso compré un póster en la tienda de regalos.
    Rosa A: Yo también compré uno la semana pasada.   √
    Rosa B: También compré uno la semana pasada.

(28) *Haciendo una acusación*
    Carmen: ¿Entonces, qué está diciendo?
    Rosa: Juana y usted estaban ahí en ese momento.
    Carmen: ¿y?
    Rosa: Que Juana y usted tenían una razón para hacerlo.
    Carmen: Otro lo hizo.
    Rosa A: Ustedes lo hicieron. Nadie más.   √
    Rosa B: Lo hicieron. Nadie más.

Of interest here is the observation that intermediates select overt subjects for items 17 and 28 at an even lower rate than beginners, and these rates achieve higher levels of significance for both items ($p =$ .0275 and $p =$ .00003, respectively). It should

also be noted that even the advanced group does not perform significantly better on these items, and near-natives patterned more with beginners, intermediates, and advanced than with natives on item 17.

The variable result among native speakers for item 35 (below) most likely arises from the unique discourse setting. The sentence preceding the binary choice in 35 appears to introduce a new discourse topic (car wrecks), which should require an overt subject to return to an earlier topic; however, for a majority of native speakers the salient topic, despite intervening material, appears to remain the two friends.

(35) *Malas noticias*
    Luis: ¿Qué dijo tu hermano?
    Julio: No hay nada más que los doctores puedan hacer.
    Luis: ¡Qué horrible!
    Julio: Sí, es triste.
    Luis: Los choques de automóvil son muy trágicos.
    Julio A: Eran muy buenos amigos.   √
    Julio B: Esos dos eran muy buenos amigos.

Perhaps the saliency of a tragic death provides a persistent topic that here overrides typical topic behavior.[5] Nevertheless, given this discourse situation, native speakers still retain overt subjects more frequently than beginners (40% to 15%, respectively), with a significance level of $p = .0171$.

The results on all the items related to null subjects may be summarized as follows. Learners begin to accept null subjects early, although they do not choose them as often as native speakers. It is not until the advanced level that choice patterns related to the dropping of topic-connect subjects converge on native patterns. Early learners also differ from natives in their choices regarding nontopic subjects. This is an unusual result because the choices made would not be preferred in English or Spanish, indicating that

learners create an interlanguage stage differing from both the L1 and L2. The difference

between learners and native speakers becomes even more pronounced at the intermediate

stage, where even fewer nontopic subjects are retained. Eventually, learners converge on

the native patterns, but this happens only very late in the learning process.

Section 5.3.2 Inversion results

We turn next to the results related to inversion. The 12 inversion items were also

of two types, six items that involved presentational or contrastive focus (for which it was

predicted that inversion would occur) and six items that lacked this special discoursal

feature (for which it was predicted that the uninverted choice would be preferred.) The

investigation of whether inversion would occur with focused elements stemmed from

Grimshaw and Samek-Lodovici (1995) and Samek-Lodovici (1996) who, as discussed in

Section 4.4, argued that subjects in Italian obligatorily appear at the right edge of a VP

when the discourse context calls for them to be focused. The presence of a pro-drop

'parameter' would imply a similar result for Spanish, but the results here suggest this is

not the case.

Table 5.5 shows the percentages by group of how often subjects selected the

inversion choice when no special focus was associated with that item.

Table 5.5 Percentage of choices for inversion, without focus

Item	Beginning	Intermediate	Advanced	Near-Native	Native
2	0.0938	0.0392	0.0714	0.5000	0.0333
3	0.1250	0.0196	0.0536	0.1667	0.1333
22	0.2344	0.1765	0.1250	0.1667	0.0333
23	0.1094	0.1961	0.2321	0.1667	0.0667
24	0.1406	0.0588	0.1964	0.1667	0.0000
30	0.2500	0.1373	0.1607	0.1667	0.0333
Avg.	.16	.10	.14	.22	.05

All proficiency groups preferred not to invert for these items, as was expected. The

Fisher Exact Test revealed that none of the differences in Table 5.5 were significant.

The results with respect to focused constituents were less clear. Table 5.6

displays the percentages of choices for inversion across proficiency levels for those items

for which contrastive or presentational focus may have been thought to trigger inversion.

Table 5.6 Percentage of choices for inversion, with focus

Item	Beginning	Intermediate	Advanced	Near-Native	Native
4	0.2500	0.2157	0.3036	0.8333	0.9667
5	0.2344	0.1765	0.5357	0.6667	0.4667
8	0.3750	0.1765	0.2857	0.1667	0.2333
11	0.3125	0.2549	0.3393	0.3333	0.2667
15	0.2188	0.0392	0.1607	0.6667	0.2000
33	0.0469	0.0784	0.0714	0.5000	0.0333
Avg.	.24	.15	.28	.53	.36

In contrast with Italian, native speakers of Spanish do not regard inversion of these

focused subjects as obligatory, nor do they even demonstrate a preference for inversion,

with the single exception of item (4).

Nevertheless, several points should be noted. First, among native speakers, the

average percentage of choices selecting inversion for focused constituents (36%) is

higher than for non-focused constituents (5%). This is true also of each of the other

proficiency levels, as Table 5.7 shows:

Table 5.7 Average percentage of selection for inversion, by condition

Condition	Beginning	Intermediate	Advanced	Near-Native	Native
Non-focus	.16	.10	.14	.22	.05
Focus	.24	.15	.28	.53	.36

Therefore, rather than inversion being forced by the discourse condition of focus (as Grimshaw and Samek-Lodovici (1995) claimed for Italian), in Spanish the effect of focus appears to be optionality. In other words, inversion may indeed be a 'free' choice in Spanish, at least as the choice relates to declaratives sentences and non-*wh*-questions, such as those under investigation in this study.[6]

Second, we should notice that pairwise comparison of groups did yield two significant differences. For item (4) (below), native speakers strongly preferred (97%) the inverted choice over the non-inverted choice (3%). Near-natives also showed a strong preference for inversion on this item, while all remaining groups preferred the non-inverted choice.

(4) *Asumiendo responsabilidad*
    Pablo: *¿De qué te ríes?*
    Janet: *Mamá y Papá van a llegar pronto.*
    Pablo: *¡Ay! ¡La casa está muy sucia!*
    Janet: *Te dije: 'No invites a tus amigos.'*
    Pablo: *¿Me ayudas a limpiar la casa?*
    Janet A: *¡No! Tienes que limpiarla tú, no yo.*   √
    Janet B: *¡No! Tienes que limpiarla, no yo.*

The highest level of significance for item (4) was between intermediates and natives ($p = .0000000005$), and beginners and advanced also registered significant differences from the near-native and native groups; however the results on this item must be viewed cautiously due to an additional consideration. The choices for (4) contrast an inverted subject with a null subject. Given this emphatic context, native speakers found the inverted subject preferable to the null subject, but their ideal choice may have been an overt, non-inverted subject.

The other significant interaction revealed was between intermediates and advanced learners on item 5:

(5) *En la oficina*
    Esther: *Henry se va de vacaciones.*
    Juana: *¿De veras? El nunca se va de vacaciones.*
    Esther: *Tienes razón pero esta vez sí se va.*
    Juana: *¿A dónde va?*
    Esther: *A Maui, pero no va solo.*
    Juana A: *¿Marta va con él?*   √
    Juana B: *¿Va Marta con él?*   √

Item (5) did not have the confounding factor of item (4); it offered a true choice of inversion versus non-inversion. Native speakers regarded the choice of inversion here as optional, selecting the inverted choice 43% of the time. Advanced selected the inverted choice 53% of the time, and this was significant (p = .0001) in comparison with intermediates, who selected the inversion item only 17% of the time. In fact, it is only at the advanced level and above that choices for inversion ever reach more than 38% for any item.

Finally, although the near-native group converged on the native speaker group for an item such as dialogue 4 (where inversion was predicted), the near-native group generally inverted more than native speakers, whether the dialogue suggested focus or not. Near-native speakers inverted 22% of the time with non-focus items and 53% of the time with focus items—more than any other group. This suggests either hypercorrection, or that these speakers may enter a stage of overgeneralization, such as we saw with the intermediates regarding null subjects.

To summarize the findings regarding inversion, inversion in Spanish does not appear to pattern the same as has been claimed for Italian. Discourse does appear to have

an effect on the acceptability of inversion, but this effect is not an obligatory one. Recognition of the optionality of inversion for focused constituents appears most noticeably at the advanced level, with the lower levels selecting inverted choices an average of less than 25% of the time. The results here demonstrate convergence between natives and near-natives for some items, but this result is confounded by the possible interference of a choice between a null subject or an overt inverted subject in one instance, what appears to be true optionality in another other instance, and the possibility of overgeneralization on the part of the near-native group.

Section 5.3.3 *That-trace* results

The majority of the *that-trace* items in this study contrasted sentences with a complementizer *que* (followed by a null subject) with sentences lacking the complementizer (also followed by a null subject). A few additional sentences tested other manipulations of *que* and subjects: one sentence in which (+que/+null subject) was contrasted with (-que/-null subject), and two sentences in which (+que/-null subject) was contrasted with (+que/+null subject).

Those sentences contrasting a complementizer and null subject with sentences lacking the complementizer revealed clear and significant results. For all items, native and near-native speakers registered strong preferences for the inclusion of the complementizer on an average of 91% and 86%, respectively. This contrasted significantly with beginners, who registered a preference for complementizer inclusion only 46% of the time, and intermediates who did only slightly better (56%). The

advanced group performed squarely between the lower and higher levels, selecting the complementizer 70% of the time.

Table 5.8 records the results by group and item in terms of the percentage of choices for inclusion of the complementizer:

Table 5.8: Percentage of choices including complementizer (+que/+null vs. −que/+null)

Item	Beginning	Intermediate	Advanced	Near-Native	Native
1	0.6563	0.6863	0.6250	1.0000	0.8667
6	0.4219	0.4118	0.5536	0.6667	0.8667
7	0.4062	0.3922	0.7143	0.8333	0.9333
9	0.4219	0.6275	0.6786	1.0000	1.0000
13	0.5469	0.8039	0.9821	1.0000	0.9667
19	0.4063	0.5294	0.7679	1.0000	0.8333
21	0.5469	0.5686	0.6071	0.5000	0.8667
25	0.4844	0.6078	0.6250	0.8333	0.9667
26	0.3437	0.4706	0.8571	1.0000	0.9667
27	0.4688	0.5294	0.7321	1.0000	0.9333
29	0.4375	0.5686	0.6964	0.6676	0.8333
32	0.5156	0.5882	0.6786	0.8333	0.9667
Avg.	.46	.56	.70	.86	.91

Item 9 (below) serves as an example of the pattern displayed regarding the +*que*/+null vs. −*que*/+null sentences:

(9) *En la tienda*
Sr. Papas: ¿*Puedo atenderlo?*
Sr. Campos: *Estoy buscando unos zapatos de tenis.*
Sr. Papas: *Estos son los mejores.*
Sr. Campos: *Pero son muy caros.*
Sr. Papas: *Sí, pero son de la más alta calidad.*
Sr. Campos A: ¿*Quién crees que los va a comprar a ese precio?* √
Sr. Campos B: ¿*Quién crees los va a comprar a ese precio?*

For this type of item, the 2sg present form of *creer* 'to think/believe' requires the use of *que*, and this is unfailingly recognized by the native and near-native speakers in this sample. The Fisher Exact Test reveals highly significant differences between this native

168

speaker norm and the responses of beginning, intermediate, and advanced learners ($p =$ .000000006, $p = .00005$, $p = .0001$, respectively). Therefore, native-like competence regarding this type of *that-trace* item is achieved here, but only at the near-native level.

A second type of *that-trace* interaction is represented by item 36 (below) in which a choice was given between +*que*/+null subject and –*que*/-null subject:

(36) *Haciendo planes para el fin de semana*
      Laura: Estudias demasiado, Felicia.
      Felicia: Tengo un examen en la clase de filosofía.
      Laura: Estudiaste toda la semana y el fin de semana pasado.
      Felicia: Voy a tomar un descanso pronto.
      Laura: ¿Vamos a la playa mañana?
      Felicia A: No creo yo pueda ir.
      Felicia B: Yo no creo que pueda ir.   √

Once again, native and near-native speakers registered a strong preference for the inclusion of *que* (86-100% of the time), but on this item, beginning through advanced learners also preferred the choice with *que* (70-80% of the time), and the level of difference between groups was not significant for this item.

This is an unexpected result for two reasons: first, the choice of including *que* in item 36 involves a violation of *that-trace*, and it might be expected that such a violation would not be the preferred choice in the grammar of a beginning learner of Spanish. Second, on the surface, this item appears to supply counterevidence to the implicational hierarchy of Liceras (1989), because it means that learners in this study first began choosing *that-trace* at the beginning and intermediate levels, while they did not chose items with inversion until the advanced level.

There is good reason to treat this conclusion with caution, however, because early learners may have made their choice on this item not because of the presence or absence

of *que*, but due to another difference between the items. The 'A' choice in item 36 has a null subject in the main clause, whereas the 'B' choice uses an overt subject. If early learners made their choice based on the overt subject, then we are left with the possible conclusion that both beginning and native groups made the same choice, but for different reasons.

The pattern apparent for item 36 also appears with items 14 and 18, items in which (+*que*/-null subject) was contrasted with (+*que*/+null subject). Item 18 is provided below as an example of this type:

(18) *Carmen y Felipe en la oficina*
     Carmen: *¿Alguien me llamó cuando yo no estaba aquí?*
     Esther: *Sí. Luis Pérez y Lilia Enríquez.*
     Carmen: *Ok. ¿Alguien más?*
     Esther: *No, no llamó nadie más.*
     Carmen: *¿Dijeron Luis y Lilia lo que querían?*
     Esther A: *No, pero dijeron que ellos iban a volver a llamar.*
     Esther B: *No, pero dijeron que iban a volver a llamar.*   √

There were no significant differences between groups on items such as 18. In addition to testing whether a *that-trace* sequence is allowed, these items also provide an interaction between null and overt subjects. The percentages of choices for inclusion of the complementizer are give in Table 5.9.

Table 5.9: Percentage of choices including complementizer (+que/+null vs. +que/-null)

Item	Beginning	Intermediate	Advanced	Near-Native	Native
14	0.5156	0.7255	0.5000	0.8333	0.8000
18	0.6094	0.5490	0.6786	1.0000	0.7000
Ave.	.56	.63	.58	.91	.75

It is tempting to conclude on the basis of items such as these that it is not the *that-trace* sequence that causes difficulties for the early learners since, for these items, learners sometimes choose *that-trace* over the option with an overt subject. Once again, some caution is required here, because the choice could be the result of the overall preference of respondents for null subjects. If this were the case, however, one would need to explain why early learners accepted the *that-trace* sequence in (36), an item with an overt subject in the matrix clause, and also accepted the *that-trace* sequence in items (14) and (28), items with null subjects. If learners are accepting *that-trace* even when there is an option that avoids it, then it is likely, or at least possible, that *that-trace* is not the source of the difference between their response patterns and those of native speakers. Future research on this issue will require more items of this type to provide stricter controls on conditions; results regarding this type of item in the present study remain somewhat tentative.

To summarize the findings on *that-trace*, when early learners were given a choice between sentences containing *that-trace* and sentences with a null subject but no complementizer, the learners included the complementizer a little less than 50% of the time. This contrasts sharply with native speakers, who consistently chose inclusion of the complementizer. When forced to choose between two sentences that include the complementizer but vary in regards to null or overt subjects in the subordinate clause, early learners demonstrated a slightly stronger preference for *that-trace* sequences, though at a rate significantly less than native speakers.

This result introduces the possibility that learners do not reject *that-trace* sentences based on the interaction between *que* and null subjects; rather, there may other

factors in play. For example, learners may not realize that the use of the complementizer in these contexts, which is frequently optional in English, is obligatory in Spanish. If so, then respondents are having less difficulty with the violation of syntactic constraint (e.g. *that-trace*) than they are with identifying the conditions which require the use of *que*.

5.4 Interpretation

The results shown in this chapter now permit certain conclusions to be drawn regarding three of the four hypotheses presented in Section 4.1 of the previous chapter. The first of these hypotheses was that the results would indeed find evidence for the implicational hierarchy of Liceras (1989) (i.e. null subjects would appear before inversion, and inversion would appear before *that-trace* in the grammars of learners). The results of the study did not find strong support for the implicational hierarchy, due to the fact that, in this study, some *that-trace* effects appeared as early as the beginning or intermediate levels, and noticeably by the advanced level.

Nevertheless, the evidence here may have an interpretation within Liceras' hierarchy for two reasons: First, although early learners did not often select the choice with inversion, the discourse condition of focus greatly raised the acceptance rate. Beginners accepted inversion 24% of the time with focus and only 16% without, and advanced learners accepted inversion 28% of the time with focus and only 14% without. In contrast, beginners double their average acceptance rate of *that-trace* only much later, at the advanced and near-native stages. Second, since inversion appears to be optional even for native speakers, little can be made of the fact that early learners use their L1 pattern when it does not conflict with the L2. The important question, then, is not why

learners choose inversion at a rate lower than that of native speakers, but what in their grammars permits them to choose inversion at all. The choice of inversion at a 24% rate, even though such a choice is optional in the target grammar, may reveal more than the acceptance of *that-trace* at a 40% rate when that choice is obligatory in the target grammar. For these reasons, Liceras' hierarchy of null subjects > inversion >*that-trace* cannot be considered disconfirmed by this study's results.

A second hypothesis that was presented stated that 'initial acceptability' would be distinct from 'correct use' for each of these grammatical properties. This hypothesis suggested that Liceras's implicational hierarchy requires refinement, because initial acceptance of inversion does not imply a fully native grammar regarding null subjects, nor does initial acceptance of *that-trace* imply accurate use of inversion. There is enough evidence in this study to claim strong support for this hypothesis.

Learners in this study accepted null subjects very early, but their judgments regarding the acceptable use of null subjects did not demonstrate sensitivity to discourse conditions (topic/nontopic) until the advanced and near-native stages. Some time prior to learners' correct usage of null subjects, *that-trace* begins to surface (though variably) in learners' grammars. In the discourse context of focus constituents, choices for inversion increase among advanced and near-native groups. But near-native choices do not replicate native judgments regarding inversion as closely as for the null subject and *that-trace* conditions. All these findings demonstrate that a distinction must be drawn between initial acceptability (resulting from awareness of a particular grammatical property) and correct use (resulting from the proper ranking of discoursal constraints in regard to this grammatical property).

A third hypothesis presented claimed that at least some L2 learners would converge on native-like usage of null subjects, inversion and *that-trace*, and that this convergence would come as a result of sensitivity to discoursal constraints of the target language. The results of this study showed no statistically significant differences between near-native and native speakers. This was most clearly seen with *that-trace* and null subjects, where near-natives strongly converged on native response patterns. The results regarding inversion are less clear, largely due to the finding that inversion does not appear obligatory under any context even for native speakers. Near-native speakers do converge with native speakers even here, if 'convergence' is defined as 'using inversion more frequently, but not obligatorily, with focused constituents'; but under such a broad definition, all groups (even beginners) can be said to converge on native speaker competence. Nevertheless, the question of ultimate attainment appears to have an affirmative answer here: the grammatical properties of null subjects, inversion, and *that-trace* can be acquired, by at least some learners, to a level such that L2 and native speakers do not differ significantly in their judgments.

These interpretations regarding an implicational hierarchy, initial acceptability vs. correct use, and ultimate attainment are summarized in Table 5.10:

Table 5.10: L2 developmental order: null subjects, inversion, *that-trace*

	Beginners	Intermediates	Advanced	Near-Native	Native
null-subjects	selected at a lower rate than native speakers	selected at a higher rate than native speakers	appear to be selected with sensitivity to the discourse context, similar to native use		select when topics, retain when nontopics
inversion	selected at a much lower rate than native speakers, but rate increases when items are focused		inverted order is selected more frequently, but not obligatorily, in certain discourse contexts. Only weak evidence that learners converge on native judgments, some overgeneralization		optionally select inversion, but only when focused
that-trace	complementizer + null subject selected for certain contexts, but choices differ significantly from native speakers regarding the distribution of *que*		learners start to select *that-trace* with greater frequency	selection of *that-trace* indistinguishable from native judgments	consistently select *that-trace* sequence

5.5 Limitations

The investigation presented here attempted to improve on previous studies in several ways: by asking subjects to make grammaticality judgments in the context of certain discourse conditions, by supplying choices intended to vary minimally in regards to the condition under investigation, and by supporting the findings with a production measure that would provide an additional measure of learner competence. All of these measures were taken, but several limitations remain.

First, although discourse contexts in the form of a dialogue were provided, certain interpretational differences regarding the saliency of topic or focus lingered. For some items it was unclear whether the reader was cueing off of the immediate topic of the preceding sentence or a more distant topic a couple of sentences earlier. Controlling for

this factor is difficult, because if the topic is held consistent throughout the dialogue, an 'out-of-the-blue' introduction of a new topic in the final response would seem very pragmatically odd. Nevertheless, the results of such an approach would be of interest.

Second, although the attempt was made to isolate the grammatical conditions under investigation through the use of sentential minimal pairs, even this approach permitted some seepage of conditions. For example, it is useful to investigate directly learners' differences in judging between [+*que*/-null] and [+*que*/+null] sentences. One difference between such sentences is the presence or absence of *that-trace*, but these sentences also vary in the more simple sense of the null/non-null condition. Learners may make their selection based on this difference alone. One way that future research might begin to address this is first by testing a matrix sentence for the null/non-null distinction, and then by using the same sentence in a subordinate environment. That this approach was not taken here is a limitation of the present study.

Third, in terms of the distribution of subconditions (topic/nontopic, inverted/uninverted, etc.), the present study could have used more items that demonstrate clear native speaker preferences for inversion or retention of topic. The lower number of these items in the current study was the result of using the intuitions of a small group of native speakers for creating the test and conducting the pilot study using only intermediate learners. Intermediate learners were selected for the pilot because it was thought that this group would best reveal the approximate amount of time needed to administer the test and any difficulties with vocabulary; however, the results demonstrate that a more extensive pilot involving native speakers would have been useful.

A final limitation of the present study was that the translation task involved only a small number of total speakers from each level for a particular item. This was due to each subject's translating only two dialogues. The problem with this approach is that, although it permitted conclusions to be made regarding significant differences between groups, it did not make specific item analysis possible. The reason each subject was not asked to translate every item was due solely to time and fatigue considerations. It would be a tiresome task to ask advanced speakers to translate all 36 dialogues, 216 lines of text, and it would far more daunting for beginners. One way around this problem for future studies might be to ask speakers to translate only the last sentence of each dialogue. This revised task would still likely involve a longer testing period than the current study, but it would be completable, particularly if a lower number of items for each condition (e.g. six, rather than 12) were given.

Notes

[1] For this particular test, each item was answered by about only 8 people, so item analysis failed. This was not a problem in the larger grammaticality judgment task, in which every subject answered every item. It would have been tempting to simply create a table and look at the percentage of people that translated each item in a dispreferred way. The problem with this approach is that different groups of people responded to different items. A particular item (translated by only two representatives of a proficiency level) may appear to be 'easy' because it received a greater preponderance of preferred responses than it would typically, whereas if a different item received many dispreferred responses, then it appears that the latter item is 'hard', but this is confounded with the group of subjects that analyzed it. If this were easily modeled, there may have been a chance to disentangle group and item effects and arrive at adjusted item estimates. As it was, only group conditions were successfully modeled.

[2] This was not an unexpected result. LaFond, Hayes, and Bhatt (2001) also showed that intermediate learners overgenerate null subjects in contexts where retention of the subject is required by the native grammar. Possible reasons for this overgeneration are suggested in Sections 6.2 and 6.3 of Chapter 6.

[3] In the grammaticality judgment task, 83% of native speakers chose 'A' for (17) and 93% choose 'A' for (28). These two items did not undergo significant revision from the pilot to the grammaticality judgment test.

[4] Translations of all dialogues in this chapter may be found in Appendix A.

[5] After the death of a loved one, it would not be necessarily unusual, for example, for one sibling to tell another, 'I can't believe she's gone', even when the referent is not a topic in the conversation or immediate context.

[6] This does not imply that inversion is never obligatory. For example, Baauw (1998) cites examples of obligatory inversion in *wh*-questions, both when the moved element is an argument (a) and when it is an adjunct (b), although inversion is sometimes optional in *wh*-adjuncts such as *por qué* (c) and *cómo* (d).

    (a)   *¿Qué dijo Juan ayer?*
                *¿Qué Juan dijo ayer?
                'What (did) John say yesterday?'

    (b)   *¿Dónde escondió Juan el paraguas?*
                *¿Dónde Juan escondió el paraguas?
                'Where (did) Juan hide the umbrella?'

(c)   *¿Por qué compró Juan un coche tan feo?*
        ¿Por qué Juan compró un coche tan feo?
        'Why (did) John buy such an ugly car?'

(d)   *¿Cómo compró Juan este coche tan ridículo?*
        ¿Cómo Juan compró este coche tan ridículo?
        'How come John bought such a ridiculous car?'

Inversion with *wh*-elements does not involve the same type of focus-related constraint that is of interest in the current study. The relationship between *wh*-inversion and inversion based on focus is left for further research.