Article

Clitic-Doubled Left Dislocation in Heritage Spanish: Judgment versus Production Data

Jose Sequeros-Valle 1,*, Bradley Hoot 2 and Jennifer Cabrelli 1✉

1 Department of Hispanic and Italian Studies, University of Illinois at Chicago, 601 S. Morgan St., Chicago, IL 60607, USA; cabrelli@uic.edu
2 Department of Modern Languages, DePaul University, 2320 N. Kenmore Ave., Chicago, IL 60614, USA; bhoof@depaul.edu
* Correspondence: jseque3@uic.edu

Received: 1 September 2020; Accepted: 20 October 2020; Published: 29 October 2020

Abstract: This project examines whether heritage speakers of Spanish distinguish when Spanish clitic-doubled left dislocation (CLLD) is discursively appropriate via an acceptability judgment task (AJT) and a speeded production task (SPT). This two-task experimental design is intended to determine whether heritage speakers diverge from an L1 Spanish/L2 English baseline and, if so, whether such divergence is due to their grammatical knowledge, processing constraints, or other task effects. The baseline group accepted and produced CLLD significantly more than other constructions in anaphoric contexts, with the opposite pattern in non-anaphoric contexts, as expected for Spanish. The heritage speakers showed the same significant differences in production in both conditions and in the AJT’s anaphoric condition; in the non-anaphoric condition, however, they did not show any differences between CLLD and the other relevant constructions. We argue that this group of heritage speakers knows the discursive distribution of CLLD just as the baseline speakers do, as attested by the similar performance pattern in production. Furthermore, we posit that their AJT performance, which shows evidence of overextension of CLLD beyond its anaphoric context and into non-anaphoric contexts, may be due to the metalinguistic nature of AJTs.

Keywords: clitic-doubled left dislocation; focus fronting; information structure; Spanish; clitics; syntax–discourse interface; heritage speakers; acceptability; production

1. Introduction

Heritage speakers are “bilingual speaker[s] of a non-majority language who ha[ve] acquired this language naturalistically within a majority language societal context” (Leal Méndez et al. 2015, p. 85). As adults, these speakers have developed high proficiency in the majority language, typically to a point of dominance over the heritage language. The heritage language, on the other hand, shows a much wider range of outcomes, from extremely low to extremely high proficiency. This social situation commonly leads to linguistic systems in the heritage language that diverge from those of monolinguals as well as speakers that grew up monolingual in the language in question and acquired the majority language in adulthood, perhaps as the result of immigration. The input in the heritage language provided by these monolingual and late bilingual speakers to heritage speakers in childhood is called the baseline input, and many years of research have demonstrated that the grammars heritage speakers acquire can differ from those of the baseline in many ways, yet not all parts of the linguistic system are affected equally.

One major line of research with bilinguals—including heritage speakers—has focused on why certain parts of bilingual grammars seem to show a larger degree of deviation from the baseline than others. One area found to be especially vulnerable to divergence is information structure, that is,
how sentences are adapted to fit into the discourse context. For example, consider the sentence in (1), which includes an example of clitic-doubled left dislocation (CLLD) in Spanish:

1. \[
\text{[CLLD A Pedro] María } \text{lo } v-e \text{ en el parque (CLLD)} \\
\text{DOM}^{1} \text{ Pedro María CL\text{acci\text{o}}} \text{ see-3SING in the park}
\]

‘Pedro, Maria sees him in the park.’

CLLD is a contextually restricted syntactic operation, which involves the interface of syntax—here, movement of a constituent to the left—and discourse—here, anaphoric reference to something present in the previous discourse. Because it refers to something present already in the discourse, CLLD is a classic topic construction.

Topic constructions like CLLD have received ample attention from researchers in bilingualism and language acquisition, and, like other information–structural phenomena, they seem to pose special difficulty for bilinguals of all stripes, whether because they are especially vulnerable to transfer (Hulk and Müller 2000), overwhelm bilinguals’ available processing resources (Sorace 2011), or are potentially ambiguous (Polinsky and Scontras 2020a), among other reasons. Indeed, CLLD appears to pose a problem for second language learners (e.g., (Valenzuela 2006)), and, in light of existing research on interface vulnerability in heritage language acquisition (see (Montrul 2016, pp. 272–7) for a review), there is good reason to expect CLLD to also be vulnerable to divergence in heritage speaker populations. Yet, in the only study conducted thus far on CLLD in heritage Spanish, Leal Méndez et al. (2015) showed that heritage speakers patterned with the Spanish-dominant baseline group on an acceptability judgment task (AJT), which the authors took as counterevidence to approaches claiming special vulnerability for the syntax–discourse interface, such as the Interface Hypothesis (Sorace 2011).

Nevertheless, CLLD in heritage Spanish deserves continued inquiry. Although Leal Méndez et al. (2015) found evidence that heritage speakers did not diverge from the baseline group with regard to CLLD, given the fact that ample previous evidence points to topic constructions such as CLLD as likely sites of variation, their claim is worth verifying via replication. Indeed, throughout the behavioral sciences, including linguistics, many researchers have called for widespread replication to improve the field’s confidence in its empirical foundations (e.g., (Marsden et al. 2018)).

Our study’s first contribution is thus to replicate Leal Méndez et al.’s AJT with a similar population of heritage speakers of Spanish in the Chicago area. Furthermore, some researchers have argued that AJTs alone are not the best method to investigate the syntax–discourse interface (Sorace 2011) or heritage speakers (Polinsky 2018). Regarding the syntax–discourse interface, Sorace contends that its vulnerability stems from processing limitations in real-time integration of syntactic and discursive information; therefore, she argues, offline tasks such as AJTs, which do not tap processing resources to the same degree as other tasks, do not effectively reveal the facts. A similar problem may emerge when using AJTs with heritage speakers. Polinsky notes that heritage speakers’ linguistic insecurity may lead to a “yes-bias” in responses to AJTs, skewing results. She thus recommends complementing AJTs with other tasks. With these issues in mind, our study’s second contribution is to compare the replication of the AJT with a speeded production task (SPT). The SPT induces strain on the processor and provides a less metalinguistic task for heritage speakers, making it an appropriate complement to the AJT replication.

Previewing our findings, the group results show divergence on the AJT and convergence on the SPT. While the SPT data suggest that increased processing load does not impede the accurate production of the structures in question (challenging (Sorace 2011, 2012)), the AJT data suggest that possible linguistic insecurity due to the nature of acceptability tasks (Polinsky 2018) yields judgements that diverge from the baseline judgments. The findings also show distinct differences between heritage speakers and L2 learners for the same linguistic phenomenon (Leal Méndez et al. 2015; Sequeros-Valle et al. 2020;

---

1 DOM = Differential object marker, which here marks a constituent as an [+animate, +specific] direct object.
Slabakova et al. 2012); these populations are often compared for their differences in age and context of acquisition (among others), which we address in our discussion.

The rest of the paper is outlined as follows: The rest of Section 1 provides additional details on the test case, relevant hypotheses for heritage speaker acquisition of CLLD, and the research questions of the study. Section 2 discusses the participants, design, procedures, and findings for the AJT experiment, and Section 3 does the same for the SPT experiment. Lastly, Section 4 summarizes the findings, discusses them in light of the research questions, and provides some avenues for further research.

1.1. The Test Case: Spanish CLLD

In order to test whether heritage speakers are sensitive to the discursive restrictions on CLLD, this construction needs to be examined along with structures that are similar on the surface but appropriate in different discourse contexts. Because CLLD involves dislocation of an argument to the left periphery, comparison structures should also have a non-canonical word order in which the first constituent is unambiguously a dislocated object. Starting the sentence with a human differential object marker a can cue CLLD as well as focus fronting (FF) and clefts. Syntactically, these three constructions differ as follows: compared to a canonical sentence (2), CLLD (3) involves the left dislocation of a constituent (A Pedro), which is then doubled with a clitic pronoun (lo) (see López 2009)). FF (4) and clefts (5) involve a similar movement to the left periphery, but without a clitic pronoun (see Feldhausen and del Mar Vanrell 2015; López 2009). In FF, the fronted constituent appears to the left periphery (A Pedro), where it receives emphatic stress, and triggers obligatory subject–verb inversion. In clefts, however, there is a copular verb and a relative pronoun, in addition to the fronted constituent.

2. María ve a Pedro en el parque (canonical)

   ‘Maria sees Pedro in the park.’

3. [CLLD A Pedro] María lo v-e en el parque. (CLLD)

   ‘Pedro, Maria sees him in the park.’

4. [FF A PEDRO] (*lo) v-e María en el parque. (FF)

   ‘PEDRO Maria sees in the park.’

5. [Cleft A Pedro] (*lo) e-s a quien María v-e en el parque. (Cleft)

   ‘Pedro is the one that Maria sees in the park.’

Despite the surface similarities, these constructions are felicitous in distinct contexts. CLLD is acceptable only if there is an anaphoric relation between the CLLD-ed constituent and an antecedent, meaning that CLLD provides information already present in the discourse context. López (2009) encodes this fact as a pragmatic feature on the dislocated constituent, which he calls [±a(naphor)]. In example (6a), a mis hermanos ‘my brothers’ can be considered a subset of ‘your family’, which is why the relationship is anaphoric, [+a]. The use of CLLD in this discourse context is thus acceptable. However, a mis amigos ‘my friends’ in (6b) is part of a set different than ‘your family’, making the relationship non-anaphoric, [−a]. As a result, the use of CLLD in this discourse context is not acceptable. The opposite is true for FF and clefts; the fronted constituent in these two constructions references new information, which does not have an anaphoric relationship with the discourse context. Thus, a [−a] reading is acceptable for (6d) and (6f), while a [+a] reading is not acceptable for (6c) and (6e).
In summary, speakers of Spanish use three constructions that are similar on the surface in having a fronted constituent in the left periphery but appropriate in different discourse contexts: Speakers use CLLD when the fronted constituent has a [+a] relation to previous discourse, while they use FF and clefts when there is a [−a] relation.

1.2. Hypotheses for Heritage Speakers at the Discourse–Syntax Interface

To be able to use CLLD and FF/clefts in discursively appropriate contexts in a conversation minimally requires (i) acquisition of the syntax of clitic-doubling, (ii) acquisition of the relationship between the presence/absence of the clitic pronoun and the [±a] discourse context, and (iii) the ability to connect these two sources of knowledge (the syntax of clitics and the pragmatic features) in real time. As we have seen above, previous evidence suggests heritage speakers are likely to diverge from the baseline, especially with grammatical constructions that involve integrating syntactic and pragmatic knowledge, but scholars debate the source of observed divergences. Understanding when and why heritage speakers employ their heritage language differently from baseline speakers is essential for a fuller understanding of heritage languages and their acquisition, and it also has the potential to shed light on broader questions about the nature of language knowledge and use (Benmamoun et al. 2013a). We consider three potential sources of divergence based on previous proposals in the literature (along with the possibility that no divergence is observed), summarized in Table 1 and discussed in the following sections.

1.2.1. Divergent Competence

The first relevant line of research predicts that heritage speakers acquire grammatical knowledge that is different from that of the baseline. Because heritage speakers experience reduced quantity and diversity of input and often experience a dramatic shift in usage away from the heritage language and toward the majority language over the lifespan, the mental representation of their heritage language likely has different properties than the mental representations of baseline speakers. For example, Polinsky (2006) has found that the case and gender system of heritage speakers of Russian is significantly different than that of baseline speakers, and heritage Spanish speakers present a reduced range of verbal morphology (Silva-Corvalán 1994).
Table 1. Summary of predictions.

| Prediction                        | Summary                                                                 | Expected Task Performance                      |
|-----------------------------------|-------------------------------------------------------------------------|------------------------------------------------|
| Divergent competence              | Heritage speakers acquire a grammar with different features than the grammar of baseline speakers | Unlike baseline (because grammars are different) |
| Processing and production issues  | Heritage speakers acquire a grammar like the baseline, but processing resource limitations reduce their ability to integrate knowledge in real time | Like baseline (reflecting offline competence)   |
| Convergent competence masked by task effects | Heritage speakers acquire a grammar like the baseline, but certain task types (especially AJTs) are inappropriate for them; some previous results are thus potentially misleading | Unlike baseline (due to task effects) |
| Convergence                        | Heritage speakers acquire a grammar like the baseline and process it without difficulty | Like baseline                                  |

With regard to the syntax/discourse interface, the initial formulation of the Interface Hypothesis contended that bilingual grammars are subject to divergent linguistic knowledge especially where interface properties are involved. Although initially formulated to apply only to near-native adult L2 learners and bilingual adults whose L1 has undergone attrition, later researchers called for the hypothesis to be extended to L2 populations at lower levels of proficiency (Lardiere 2011; White 2011) and heritage speakers (Montrul and Polinsky 2011). Initial evidence for divergent linguistic representations at the syntax/discourse interface came mostly from subject pronouns. For example, Sorace and Filiaci (2006) examined the pragmatic interpretation of overt pronouns in Italian via a Picture Verification Task. Their results showed that L2 learners performed differently than the baseline group in their interpretation of overt subject pronouns despite similar patterns within each group’s data. Furthermore, Tsimpli and Sorace (2006) found L2 overextension of overt subjects and object fronting in data from oral interviews. Lastly, Belletti et al. (2007) reported an overuse of overt pronouns and a use of null pronouns that diverged from the native-like pattern in data from four different tasks, including judgments and production. Based on (i) group results from offline and production tasks and (ii) between-group comparisons, these authors claim that L2 speakers lack knowledge of the syntax–discourse interface in their L2. The same claim is made by Valenzuela (2006) for Spanish CLLD, who found that near-native L2 learners of Spanish differed from the control group in their sensitivity to specificity restrictions on dislocated constituents in two judgment tasks and an untimed written production task.

Although the initial proposal concerned only L2 learners and L1 attritors, a second group of researchers updated this type of claim for heritage speakers. For example, Benmamoun et al. (2013a, 2013b) predicted divergent heritage language representations, making specific reference to the syntax–discourse interface, such as the discursive distinction between null and overt pronouns in Spanish, and to the Interface Hypothesis. Along the same lines, Polinsky and Scontras (2020a, 2020b) reviewed divergence in heritage speaker linguistic knowledge in a number of areas (not only the syntax–discourse interface). For instance, they carefully review empirical evidence for what they call the “Ambiguity Problem”, noting that heritage speakers demonstrate difficulty with constructions that are ambiguous or subject to multiple interpretations and arguing that heritage speakers instead construct linguistic systems with one-to-one mappings between structures and interpretations. Information–structural phenomena,
including CLLD, generally do not display such one-to-one mappings; there are many ways to mark a constituent as an anaphor or as new information. CLLD’s anaphoric construction can map to several possible meanings, including set-subset relations, identity relations, changing topics, and others (e.g., (Bianchi and Frascarelli 2010; López 2009; Sequeros-Valle 2020)).

If heritage speakers acquire grammars that are simply different from those of baseline speakers, as predicted by the initial formulation of the Interface Hypothesis extended to heritage speakers, or as part of constructing more parsimonious grammatical systems in acquisition as a response to reduced input, the prediction that follows is that their performance in our study will diverge from that of the baseline, regardless of the characteristics of the task, given that their grammars are simply different from the baseline in this regard. We would thus expect divergent judgments on the AJT and divergent production in the SPT.

1.2.2. Processing and Production Issues

The Interface Hypothesis inspired intense research on the acquisition of the syntax/discourse interface, and, in response to new data, Sorace (2011, 2012) refined its tenets. Building on the thread of psycholinguistic evidence showing that bilinguals simultaneously activate both their languages and must therefore inhibit one language to use the other, which requires expending cognitive resources, Sorace argued that bilingual divergence is caused by limits in resource allocation and that these limitations are not evident in AJTs. In a nutshell, inhibiting the dominant language consumes attentional resources, resulting in fewer resources to apply to language use. This attentional limit may not become apparent when testing a construction that involves only core syntax or certain language–internal interfaces, but when the interface with the broader discourse context becomes involved, and speakers have to integrate knowledge across mental domains, the resulting processing load proves to be too much for the available resources. Processing breakdown is the result.

As a test of this revised ‘Processing’ Interface Hypothesis, Sequeros-Valle et al. (2020) tested the discursive restrictions of CLLD and FF/clefts with L2 learners of Spanish (L1 English) in an SPT. The authors included a replication of Slabakova et al.’s (2012) AJT and an SPT. This second task was implemented with the goal of increasing the processing load by the addition of both time constraints (Sorace 2011) and unplanned production (Ellis 2009). Sequeros-Valle et al. (2020) reported findings that align with Sorace’s predictions: AJT data indicated that L2 knowledge does not differ from knowledge exhibited by an L1 Spanish/L2 English comparison group, but the SPT data suggest that the L2 performance differed when processing pressure was increased. Specifically, the learners did not show significant differences between the anaphoric and non-anaphoric experimental conditions, even though the L1 Spanish group did show a difference between the two conditions. The authors, however, argue that the same pattern has been found with L2 speakers when studying non-interface phenomena such as L2 gender (e.g., (Grüter et al. 2012)). Therefore, bilinguals may be subject to processing limitations related to real-time production in general that are not specific to the interface nature of CLLD and FF/clefts.

Although Sequeros-Valle et al. (2020) tested Spanish CLLD with L2 learners, the same logic applies to heritage speakers (see (Montrul and Polinsky 2011)). If processing limitations are the fundamental cause of apparent bilingual grammatical divergence, Sorace (2011) argues, then AJTs are not appropriate to test interface phenomena. Instead, Sorace proposes to test the hypothesis with tasks that measure processing directly or put additional pressure on the processing resources of the speaker.

---

2 Sorace (2011) excludes heritage speakers from the scope of the hypothesis. Yet others (Montrul and Polinsky 2011; Hoot 2017) have argued that the hypothesis should be extended to this population, a position Sorace (2012, p. 214) accepts “as long as the differences between individual and generational attrition are clear”.

3 Although the initial formulation of this hypothesis compared “narrow syntax” to interface structures, Sorace (2011, p. 10) recognizes that it is “difficult to identify structures that are sensitive to exclusively syntactic constraints” while also noting that researchers have “repeatedly emphasized syntactic principles and dependencies as having a different status from non-syntactic ones in terms of acquisition and processing”. 
In this case, we would expect that heritage speakers would perform like the baseline group on the AJT, but would diverge on the SPT once processing pressure becomes a factor.

1.2.3. Convergent Competence Masked by Task Effects

Although some approaches predict divergent linguistic systems in heritage speakers, whether due to different language knowledge or processing limits, other lines of research posit that divergent outcomes could be an artefact of task type rather than a reflection of their grammars. AJTs have been of particular concern to researchers in this regard. AJTs have been a foundational method in linguistics and have been profitably employed to elicit reliable data for a range of phenomena from many different types of speakers (Sprouse and Almeida 2017). Data provided by AJTs have been and continue to be crucial to the field, with several advantages over other tasks (Schütze and Sprouse 2013). Nonetheless, it is clear to anyone who has designed an AJT that many nonlinguistic factors intervene between the task and the speakers’ grammars (which is why many prefer calling them acceptability tasks rather than the traditional name, grammaticality judgment tasks, given that they don’t measure grammaticality directly). Among these nonlinguistic factors are individual characteristics of the speakers, and it is for this reason that researchers have been concerned about the use of AJTs with heritage speakers.

Polinsky (2018) provides a useful overview of research on AJTs with heritage speakers, highlighting evidence that shows that heritage speakers demonstrate a “yes-bias”, in which participants “tend to correctly accept grammatical structures but are reluctant to reject ungrammatical ones”, a behavior that “is probably rooted in uncertainty about language” (Polinsky 2018, p. 96). She goes on to note that bilinguals largely have greater metalinguistic awareness than monolinguals (they are more aware of language because they have to navigate two of them) while at the same time heritage speakers likely suffer from significant insecurity about their heritage language (because at some point most have been scolded, shamed, or belittled for not speaking “right”) or uncertainty about the limits of their language knowledge (because they likely have limited exposure to certain aspects of the language). These social facts about their language use may well create heightened awareness of their insecurities or uncertainties, resulting in the yes-bias that previous researchers have observed and making AJTs a task that is difficult to employ and interpret with this population. Furthermore, there seems to exist a broader correlation here: the more metalinguistic or explicit the task, the less well it does with heritage speakers, while the more naturalistic or implicit the task, the better it measures heritage speaker linguistic abilities (see (Polinsky 2018) for a thorough discussion).

For this reason, some previous results showing that heritage speakers diverge from the baseline could in fact be due to task effects. It is also worth keeping this possibility in mind in the present study: if heritage speakers pattern differently from the monolinguals on the AJT, one possible cause to consider is whether said pattern could be due to a yes-bias on the part of the heritage speakers. On the other hand, a task that lies further toward the naturalistic end of the task spectrum (such as an oral production task) may be more likely to serve as a valuable window into heritage speaker competence. Regardless, Polinsky (2018, p. 100) recommends “that the overall successful strategy of experimental testing of heritage speakers should combine production and comprehension”, which is an approach we adopt here.

1.2.4. Convergence

Although substantial previous work makes clear that heritage speakers diverge from the baseline on many grammatical features and suggests that the syntax–discourse interface is especially likely to suffer from such divergence, Polinsky and Scontras (2020a) note many areas of heritage grammars that are resilient to apparent change. A final possibility we should consider, then, is that heritage speakers may indeed fully acquire the syntax of CLLD and its discursive restrictions and be able to process them in real time, just like monolinguals. In fact, the one previous study we are aware of that tested CLLD with heritage speakers, Leal Méndez et al. (2015), found precisely such convergence, albeit from a judgment task with low processing pressure. The authors investigated the performance of heritage speakers of
Spanish on the discursive restriction of Spanish CLLD and FF using the task in Slabakova et al. (2012) (see Section 2 for study details, which we replicate herein). The authors tested the acceptability of these two constructions via an AJT completed by monolingual Spanish speakers, L1 Spanish/L2 adult English learners, intermediate heritage speakers, and advanced heritage speakers. All four groups judged felicitous responses differently than infelicitous responses, and Leal Méndez et al. concluded that the heritage data pattern with the monolingual and the sequential bilingual groups. Contrary to what Leal Méndez et al. expected, the heritage speakers apparently did not struggle with the integration of CLLD and FF in their discourse contexts as reflected in the AJT data, which the researchers interpreted as a challenge to the Interface Hypothesis. Nonetheless, as previously mentioned, Sorace contends that AJTs do not test the hypothesis appropriately, and AJT data should be complemented with data from tasks that tap possible processing limitations. We thus take Leal Méndez et al.’s results as a starting point, replicating their task while also administering a production task designed to tax participants’ processing resources.

1.3. The Present Study

Taking into account the literature reviewed thus far, we propose four research questions. Let us start by looking into the first two:

RQ1: Do heritage speakers diverge from the baseline when interpreting CLLD?
RQ2: Do heritage speakers diverge from the baseline when producing CLLD?

In order to be able to test the distinction between interpretation and production predicted by some of the literature presented thus far, we propose an experimental design divided into two tasks. On the one hand, a replication of Slabakova et al.’s (2012) and Leal Méndez et al.’s (2015) AJT (Section 2) will test the predictions related to language interpretation. On the other hand, an SPT (Section 3) will test the predictions related to language production. The relevant set of predictions for RQ1 and RQ2 combined are presented in Table 2:

| Hypothesis                                         | AJT   | SPT   |
|----------------------------------------------------|-------|-------|
| Convergence                                        | Convergence | Convergence |
| Processing and production issues                   | Convergence | Divergence |
| Convergence competence masked by task effects      | Divergence | Convergence |
| Divergent competence                               | Divergence | Divergence |

The results related to the first two research questions are to be analyzed using within-subject, within-task comparisons, following Leal Méndez et al. (2015).

These main research questions are complemented by two additional questions that have been found to be relevant in previous studies of heritage language grammars. First, in addition to looking at overall patterns, given the heterogeneous nature of heritage speaker populations, it is important to consider the role of proficiency in the heritage language. This consideration motivates the third research question.

RQ3: Does proficiency play a role in the knowledge and/or production of Spanish CLLD by heritage speakers?

To answer this question, we examine data from both tasks and how the data are modulated by proficiency, using within-group, between-subjects, within-task comparisons.

Second, one of the ways data from heritage speakers can be valuable to our understanding of bilingualism at large is to compare it to data from L2 learners. Throughout Section 1.2, the parallels in heritage and L2 outcomes regarding interface vulnerability are evident, and hypotheses about heritage language acquisition have historically been intertwined with hypotheses about L2 acquisition. In an overview of these comparisons, Montrul (2016, pp. 272–77) reports that heritage speakers generally
outperform proficiency-matched L2 learners in direct comparisons, particularly when the property under investigation is not part of explicit classroom instruction, although results vary by task and target structure. The differences between groups are posited to be gradient rather than categorical and might differ depending on the interface, and they have been hypothesized to be a result of input timing (i.e., age of acquisition), input quality (naturalistic vs. classroom exposure), successful coordination of linguistic and cognition domains and other factors. Comparing L2 and heritage speaker outcomes thus contributes to understanding how these diverse factors contribute to language acquisition.

With regard specifically to CLLD and the syntax–discourse interface, the comparison of AJT data from L2 learners in Slabakova et al. (2012) with data from heritage speakers in Leal Méndez et al. (2015) found that heritage speakers and L2 learners both patterned with Spanish monolingual and L1 Spanish-dominant bilingual baseline group, similar to the advanced L2 learners in Sequeros-Valle et al. (2020), which used the same AJT. It remains to be seen, however, whether the heritage data in the present study will align with the L2 data from Sequeros-Valle et al., and whether heritage speakers pattern with L2ers in a task with a higher processing load. With this comparison in mind, we posit RQ4.

RQ4: How do heritage speakers’ knowledge and production of Spanish CLLD compare to that of L2 learners?

To answer this question, we compare our results with data from a previous L2 study that used the same tasks (Sequeros-Valle et al. 2020).

2. Experiment 1—Acceptability Judgment Task

Given the linguistic and experimental nature of this research project, human subjects were required. The initial contact with participants, obtention of consent for participation, data collection procedure, and data de-identification were carried out following the indications from the Institutional Review Board at the University of Illinois at Chicago, under Protocols 2009-0121 and 2015-0040.

2.1. Methodology

The complete experimental design includes a proficiency test, the SPT, the AJT, and a clitic test, administered in that order. The original design did not include the AJT; therefore, participants in a first round of testing (i.e., all baseline participants and three of the heritage participants) completed the proficiency test, the SPT, and the clitic test in one in-person session and the AJT in a second online session once the AJT was included in the design. The remaining 12 heritage participants, who completed the study after the AJT was added, completed all four tasks in one in-person session. The proficiency and clitic tests are presented in Section 2.1.1 along with the description of the participants, the AJT is described in Section 2.1.2, and the SPT is presented in Section 3.

2.1.1. Participants

Twenty-nine participants completed the tasks in both the AJT and the SPT, divided into two groups: fourteen baseline speakers of Spanish (nine female) and 15 adult heritage speakers (13 female) (see Section 2.1.3 for power analysis). Table 3 presents the age distribution of both groups of participants. Heritage participants are U.S.-born bilinguals who grew up with exposure to Mexican Spanish from birth and to North American English starting between birth and 9 years old. Our baseline Spanish participants are originally from different Spanish-speaking countries (eight from the Basque Country, three from southern Spain, two from Colombia, one from Chile, and one from Cuba). All learned English after the age of 12 and are thus categorized as late sequential bilinguals (see (Lightbown and Spada 2006)). All participants lived in the Chicago, Illinois, area at the time of testing.

Let us discuss the selection of this group briefly. First, we acknowledge that there is some debate about who the most appropriate comparison group for heritage speakers is, and many recommend including both homeland monolinguals and baseline bilinguals (as in (Leal Méndez et al. 2015)). We were not able to collect data from Spanish monolinguals for practical reasons, and therefore we
only include a baseline group. The baseline speakers in the study provide an adequate representation of the input to which heritage speakers are exposed in one regard, because they are L1 Spanish learners of L2 English, likely similar to the parents of the heritage speaker group in that respect. However, this baseline group includes speakers of several varieties, including substantial representation from Peninsular dialects, rather than only Mexican Spanish, which may differ from the actual baseline to which the heritage speakers were exposed, given some dialectal variation in the phonology of CLLD (Feldhausen 2016). We acknowledge this limitation, yet note that previous studies (e.g., (Leal et al. 2018)) suggest that the presence and absence of clitic-doubling in Mexican Spanish do follow the same pattern as the one described above (López 2009), which is also the pattern found in our baseline group (see results sections below).

Each heritage speaker participant completed two tasks to determine eligibility for participation. Before the experimental tasks, the participants completed a 50-item written proficiency measurement composed of sections from the Diploma de Español como Lengua Extranjera (DELE) and Modern Language Association proficiency exams, commonly used in L2 Spanish research (e.g., (Montrul 2002, 2004; White et al. 2004)), but also in studies with heritage speakers (e.g., (Leal Méndez et al. 2015)). In our study, scores ranged from 38 to 46 (out of 50); in previous research, scores of 30–39 have been considered to reflect intermediate proficiency and 40–50 to reflect advanced proficiency (e.g., (Cabrelli Amaro 2017; Giancaspro 2015)). Following Leal (2018), we treat proficiency as a continuous variable rather than dividing participants into proficiency groups. However, it is worth noting that our sample was overall highly proficient: ten speakers scored at the top of the scale (40–46) and five would be considered intermediate (38–39). Our sample thus contains speakers at the high end of the proficiency scale that are comparable to the more advanced group in Leal Méndez et al.’s study.

Table 3. Participants’ mean written proficiency score, mean clitics score, and mean age for the baseline and heritage speaker groups.

|                | n  | Proficiency (1–50) M | Proficiency (1–50) SD | Clitics (1–50) M | Clitics (1–50) SD | Age M | Age SD |
|----------------|----|----------------------|-----------------------|------------------|------------------|------|--------|
| Baseline       | 14 | -                     | -                     | -                | -                | 30.00| 6.90   |
| Heritage speakers | 15 | 40.87                 | 2.50                  | 47.47            | 1.96             | 20.80| 2.42   |

After the experimental tasks, participants completed a 10-item multiple-choice test to assess knowledge of the syntax of clitics (specifically, clitic placement and contexts in which clitics are obligatory), which, as Leal Méndez et al. (2015) note, is a prerequisite for the knowledge of clitics’ semantic and discourse restrictions. The 10 items were similar to (7) (example 11 in (Slabakova et al. 2012, p. 330)):

7. Lucia: ¿José le trajo los libros al profesor? ‘Did Jose bring the books to the professor?’ Pedro: Sí, __________ (Yes, he-brought them to-him) se los trajo

* los se trajo
* trajo
* se trajolos
* trájoselos

Participants scored one point for marking a correct answer and one point for each unmarked incorrect answer. Following Leal Méndez et al. (2015), the minimum score for inclusion was 35 out of 50 (10 items, five points per item). No participant was excluded for obtaining a score under 35 in this test. The group results for this test can be found in the column ‘Clitics’ in Table 3.
2.1.2. Experimental Task

The first experimental task is the AJT used by Slabakova et al. (2012) and Leal Méndez et al. (2015). The task includes 40 items, 15 experimental (ten for [+a] and five for [−a]) and 25 fillers. Slabakova et al. included two types of [+a] items: (a) equivalence, in which the CLLD constituent recalls the same referent from previous discourse (e.g., Pedro–Pedro), and (b) set–subset, in which the CLLD constituent recalls a member of an antecedent set (e.g., your friends–Pedro). In the end, both Slabakova et al. and Leal Méndez et al. concluded that this semantic distinction did not make a difference for any of the groups included. Nonetheless, because our purpose was to replicate their experiment, we included both types of items just as they did, which is why there were ten [+a] contexts and five [−a] contexts. Because neither Slabakova et al. nor our analysis below finds any difference between the two types, we hereafter treat all these items as simply [+a] items. We refer the reader to Slabakova et al.’s original article for further details.

Given a context and a dialogue, participants judged two possible answers on a 4-point Likert scale from 1 (very strange) to 4 (perfect), or “I don’t know”. One of the two possible answers always included clitic-doubling, while the other one did not. Participants were presented with the context/dialogue, as well as the two answers, both aurally and visually. The inclusion of an aural presentation of the stimuli aligns with Polinsky (2018), who considered this mode more natural and appropriate for heritage speakers. Example (8) is an item from the AJT, taken from example (12) in Slabakova et al. (2012, p. 331).

8. Juan y Mónica están recién casados y se acaban de mudar a su nueva casa. Mónica trabajó todo el día mientras Juan se quedó en casa y acomodó los muebles. Mónica lo llamó para ver cómo estaban las cosas. ‘Juan and Monica are newlyweds and have just moved into their new home. Monica worked all day long while Juan stayed at home and arranged the furniture. Monica called him to see how things were going.’

Mónica: Entonces, ¿qué hiciste con los muebles?

Monica: ‘So, what did you do with the furniture?’

A) Juan: Bueno, sillas Cl.ACC.pl put.past.1sing in the kitchen and los sillones Cl.ACC put.past.1sing in the living-room

‘Well, I put the chairs in the kitchen, and I put the armchairs in the living room.’

B) Juan: #Bueno, las sillas put.past.1sing in the kitchen and los sillones put.past.1sing in the living-room

‘Well, I put the chairs in the kitchen, and I put the armchairs in the living room.’

Given that las sillas ‘the chairs’ and los sillones ‘the armchairs’ are subsets of the larger set los muebles ‘the furniture,’ the utterance in (8) is an example of the [+a] condition. As a result, answer A (which includes a clitic pronoun) should obtain a higher degree of acceptance (closer to 4) than clitic-less answer B (closer to 1). The utterance in (9) (example (13) in Slabakova et al. (2012, p. 331)) is an example of the [−a] condition; la carne ‘the meat’ is a different set than la sopa ‘the soup.’ In this case,

---

4 As reported in Leal Méndez et al. (2015), these 25 fillers included five corrective focus clitic-doubled right dislocation (CLRD), five ambiguous CLRD, five Rheme constructions, and 10 additional constructions that did not manipulate information structure.

5 We used the original voice recordings from Leal Méndez et al. (2015), and they were at least a male and a female native speaker of Mexican Spanish, but we do not have additional information. Furthermore, each and every word presented visually was also presented aurally, and vice versa.
we predict a higher degree of acceptance for the clitic-less construction A, and a lower degree of acceptance for the construction with a clitic pronoun in B.

9. Juan y Mónica invitaron a María a comer. La cena se sirvió en la terraza y todo estaba muy rico. María felicitó a Juan por la sopa que había hecho. Cuando Mónica escucha esto, responde:

‘Juan and Mónica invited Maria for dinner. Dinner was served on the terrace and everything was very tasty. Maria complimented Juan on the soup he made. When Monica hears this, she responds:’

A) Mónica: LA CARNE preparó Juan, no la sopa.
    the meat prepare.3sing.past Juan, NEG the soup
    ‘Juan prepared the meat, not the soup.’

B) Mónica: # LA CARNE la preparó Juan, no la sopa.
    the meat CL.ACC prepare.3sing.past Juan, NEG the soup
    ‘Juan prepared the meat, not the soup.’

The reader may find the complete data set in the Supplementary Materials of this manuscript.

2.1.3. Variables and Analysis

We fit a linear mixed model in SPSS 26 (IBM Corp.) with Rating (1–4) as the dependent variable and three fixed effects—Group (baseline, heritage), Discourse Context ([+a], [−a]), and Clitic (clitic, no clitic)—plus their interactions. Following Barr et al. (2013), we fit the maximal random effects structure supported by the data. The model included a random intercept by subject and item and a random by-subject slope over Clitic and its interaction with Discourse Context. To examine the possible effect of Proficiency, we fit a second model to the heritage speaker data that included Discourse Context and Clitic as fixed effects and Proficiency (for which the participants’ scores ranged from 38 to 46) as a continuous covariate. We again used the maximal random effects structure, which in this case was one with a random intercept by subject and item, along with a by-subject slope over Discourse Context and its interaction with Clitic.

As noted earlier, Slabakova et al.’s (2012) and Leal Méndez et al.’s (2015) original studies included two types of [+a] items, which is why there are 10 [+a] items and five in the [−a] condition. Considering the results from the two original studies, we treat both [+a] types as a single category in this analysis and throughout the paper. While this results in an unbalanced design, linear mixed models are a statistical technique able to accommodate such data.

Given the group sizes, which are typical of many linguistics studies but are smaller than those typically used in some allied disciplines (e.g., psychology), a concern about statistical power is reasonable. We therefore conducted a power analysis by simulation, adapted from Lane and Hennes (2018), for both linear mixed models. The challenge for any power analysis is defining the minimum practically meaningful effect size, especially for higher-order interactions. For the linear mixed model that included both groups of participants, we began with the conservative assumption that the three-way interaction, which is the main result of interest for this test, would have an effect of $\beta = -0.5$, roughly half as big as the effect we actually observed. Under this assumption, the first linear mixed model had 57.5% power to detect the Discourse*Clitic*Group interaction; that is, if the true effect size in the population for that interaction were $-0.5$, we would detect it 57.5% of the time with our sample of $n = 29$. However, if we assume that the true population effect was something closer to the effect we observed in our sample, then our test had 97.9% power. In other words, if the real population effect is closer to our observed effect of $\beta = -0.91$, the simulation revealed that our sample would be sufficient to find that effect 97.9% of the time. The second test, examining the role of proficiency, however, had less power: under conservative assumptions about effect sizes (based on effects of a similar test reported in (Sequeros-Valle et al. 2020)), our sample size resulted in only 36.3% power for the three-way interaction that was most of interest. If, instead, the true population effect is
closer to our observed effect of $\beta = -0.08$, the power of our test to detect it would be even less, because the effect is so small (a fact which is also consistent with there being no real effect in the population). For this reason, although we report the results of the inferential statistics, we rely more heavily on the visualization of the data to understand the proficiency results.

2.2. Results

2.2.1. Group Results

Figures 1 and 2 present the group results for Experiment 1 in the [+a] and [−a] conditions for the baseline group (Figure 1), and the heritage group (Figure 2).

![Figure 1](image-url)  
**Figure 1.** Baseline group ratings on the acceptability judgment task (AJT) by discourse context and clitic presence. Error bars represent 95% confidence intervals.

Recall that our primary interest lies in whether learners make a distinction between the presence and absence of a clitic in [+a] and [−a] discourse contexts, with a lesser interest in how the degree of this distinction compares with baseline speakers with different patterns of language usage and experience. With this in mind, while we report main effects, we focus on whether there is a significant interaction between the Discourse, Clitic, and Group fixed effects. We highlight the most relevant results produced in SPSS, which consist of (i) the $F$ statistics and $p$ values yielded by the test of fixed effects and (ii) the $p$ values and effect sizes (Hedges’ $g$) of the pairwise comparisons.

A Type III test of fixed effects showed no main effect of Group ($\beta = -0.09, F(1,28) = 0.05, p = 0.819$), but a significant effect of Clitic ($\beta = 0.96, F(1,28) = 6.84, p = 0.014$) and Discourse ($\beta = 0.39, F(1,19) = 4.54, p = 0.046$), as well as a significant Discourse × Clitic interaction ($\beta = -1.27, F(1,55) = 131.85, p < 0.001$). More important, though, is the significant Group × Discourse × Clitic interaction ($\beta = -0.91, F(1,55) = 9.19, p = 0.004$), which is the outcome of interest. Bonferroni-corrected pairwise comparisons for the Group × Discourse × Clitic interaction showed significant differences between sentences with clitics and those without clitics for the baseline group in both the [+a] discourse context ($p < 0.001, g = 0.98$) and the [−a] discourse context ($p < 0.001, g = -0.82$). For the heritage group, results show a difference in the [+a] discourse context ($p < 0.001, g = 0.76$), but no significant difference in the [−a] discourse context ($p = 0.079, g = -0.27$).

---

6 All figures generated in Microsoft Excel.
First, within each condition ( [+a] and [−a]), we looked at whether the 95% Confidence Interval (CI) of each participant’s mean ratings for sentences with clitics overlapped with the 95% CI of their ratings of clitic-less sentences. If the two ranges did not overlap, we took it as evidence that the speaker treated each construction type differently. However, like with the one-point difference, only eight of the 14 baseline group showed the expected differences, making this an invalid procedure for our data. Second, within each condition ( [+a] and [−a]), we calculated the 95% CI of the difference between clitic-ed and clitic-less sentence for the baseline group overall ([+a] = 0.84–1.61; [−a] = −1.36—0.52). This second system, obviously, does not create any issues within the baseline group. However, we later considered that this is the type of between-group comparison that we were to avoid. To address this limitation going forward, as very helpfully noted by an anonymous reviewer, we can incorporate statistical methods such as Best Linear Unbiased Prediction (BLUP), which allows for the estimation of random effects whereby model predictions for each participant can be extracted.
2.2.3. Results by Proficiency Score

To examine the role of proficiency, a second linear mixed model was fit with only the heritage speakers. A Type III test of fixed effects found no main effect of Proficiency ($\beta = -0.001, F(1,13) = 0.06$, $p = 0.805$), meaning that there is no evidence that overall ratings varied as a function of proficiency. More importantly, the Discourse × Clitic × Proficiency interaction was not significant either ($\beta = -0.08, F(1,29) = 1.30, p = 0.264$), which could indicate that the effect on ratings of the presence or absence of a clitic in each discourse condition does not change as proficiency increases. Nonetheless, as we noted in Section 2.1.3, this statistical test had relatively low power, so it is also possible that there exists a difference by proficiency score that we simply failed to observe. Therefore, in order to better understand this result, it is important to visually examine the data.

Figure 3 presents the difference in judgments between sentences containing a clitic pronoun and those which do not in both the [+a] (dark blue) and [−a] (light blue) discourse contexts. A positive value means that participants judge utterances with a clitic pronoun more acceptable than clitic-less ones, while a negative value means that they prefer clitic-less utterances over ones with a clitic. We predicted a positive difference for the [+a] discourse context and a negative difference for the [−a] discourse context. Figure 3 shows that heritage speakers preferred utterances with clitics in the [+a] condition by an average of about 1 point over clitic-less utterances at all levels of proficiency, while they prefer clitic-less utterances in the [−a] condition to a lesser degree (less than 0.5 points on the four-point scale). In both conditions, though, Figure 3 demonstrates the apparent lack of an effect by proficiency: the fact that both lines are relatively flat and have mostly similar slopes likely reflects a lack of a change in preferences tied to increased proficiency.

![Figure 3](image_url)

**Figure 3.** Difference between utterances containing a clitic pronoun minus those which do not in [+a] (dark blue circles) and [−a] (light blue squares) discourse contexts, on the AJT, by proficiency score.

2.3. **Interim Discussion 1**

Considering the descriptive and inferential results presented, our interpretation is as follows. For the group results, the pairwise comparisons show that the baseline group makes a distinction in their judgments between utterances containing a clitic pronoun and those which do not in both conditions. The heritage speaker group makes the same type of distinction as the baseline group in the [+a] condition, but not in the [−a] condition. Although the heritage speakers distinguish between contextually appropriate sentences and contextually inappropriate ones in the [+a] discourse context in a baseline-like manner, they do not show evidence of this distinction in the [−a] discourse context.
For the results by proficiency score, the pattern we observed suggests no effect of proficiency. In other words, the divergent performance in the [−a] discourse context remains despite the advanced level of proficiency of some of our heritage speakers.

3. Experiment 2—Speeded Production Task

3.1. Methodology

3.1.1. Participants

The same 29 participants who completed the AJT also completed the SPT. (See Section 2.1.1).

3.1.2. Experimental Task

This study requires participants to produce sentences with the direct object in the left periphery by relying on an unrelated syntactic phenomenon in Spanish: differential object marking (DOM). Because Spanish direct objects that are [+animate] and [+specific] are obligatorily marked using the marker a, the experiment prompts are unambiguously direct objects, requiring participants to produce CLLD, FF, or clefts. For example, the DOM-marked constituent A Pedro as a prompt can only be interpreted as a direct object, and never as a subject. While it has been argued that heritage speakers may not always notice the presence of DOM (e.g., (Montrul 2014; Montrul et al. 2015)), this possible lack of sensitivity is not problematic for the AJT since complete utterances were always presented with subject and verbal inflections to mark the subject. In the SPT, however, the participant is given a lone direct object as a prompt to produce an utterance with that object (e.g., A Pedro . . . ). Therefore, if the participant does not notice a, she may produce an utterance in which the given constituent is treated as a subject. In other words, a participant may produce a canonical subject-verb-object sentence instead of direct object fronting (CLLD, FF, clefts). Despite this potential issue, there were only 14 (out of 270) invalid sentences for the entire heritage group. Within those 14 utterances, ten were unfinished, and four reflected a misinterpretation of the meaning of the discourse context. None of the invalid utterances included a direct object misinterpreted as a subject.

Participants were presented with twenty-four trials. Each trial consisted of three parts. First, the participant was presented with a short context and, second, a question based on the context. As in the AJT, the context and the question were simultaneously presented visually and aurally in Spanish (Polinsky 2018), and we limited the response time to 10 seconds (following (Leal et al. 2018)). Third, participants were presented with a written prompt to begin constructing an oral response to the question. Example (10) is one of the trials presented to the participants:

10. Tus hijos se llaman David y Antonio. Llevas a David al doctor el jueves, y a Antonio el viernes. Juan te pregunta:

‘Your children are named David and Antonio. You are taking David to the doctor on Thursday, and Antonio on Friday. Juan asks:’

Juan: ¿Cuándo llevas a tus hijos al doctor?
When take.2SING DOM your children to-the doctor
‘When are you taking your children to the doctor?’

Tú: A David ...
DOM David
‘David . . . ’

8 The aural stimuli were recorded by the first author (male L1 Peninsular Spanish/late L2 English speaker) and a female (female L1 Peninsular Spanish/late L2 English speaker). The context and question were presented aurally, while the beginning of the answer was only presented visually.
The expected answer from the baseline group would be a sentence such as (11). The differential object marker \( a \), which indicates that what follows is an \([+\text{animate}, +\text{specific}]\) direct object, forces the speaker to produce a sentence containing movement to the left periphery. The only grammatical options would be CLLD, FF, and cleft constructions. Since this sample discourse is \([+a]\), the only pragmatically adequate option is CLLD.

11. [CLLD A David] lo llevo el jueves. (CLLD)
   DOM David CL.ACC take.1sing the Thursday
   A Antonio lo llevo el viernes.
   DOM Antonio CL.ACC take.1sing the Friday

   ‘I am taking David in on Thursday. I am taking Antonio in on Friday.’

In a \([-\text{anaphor}]\) discourse context (12), we expected sentences such as FF (13a) and clefts (13b).

12. Tú traes a tu viejo amigo Rubén a una fiesta. Juan hace un comentario, pero quieres decirle que tú traes a Rubén y no a Carlos:

   ‘You bring your old friend Rubén to a party. Juan makes a comment, but you want to tell him that you are bringing Rubén and not Carlos:’

   Juan: ¡Gracias por traer a Carlos a la fiesta!
   ‘Thanks for bringing DOM Carlos to the party!’

   Tú: ¡No! ¡A Rubén...
   ‘No! DOM Rubén...’

13 a. ¡No! ¡A Rubén traigo a la fiesta! (FF)
   NEG DOM Rubén bring.1sing to the party

   ‘No! Rubén I bring to the party!’

13 b. ¡No! ¡A Rubén es al que traigo a la fiesta! (Cleft)
   NEG DOM Rubén is DOM.the-one that bring.1sing to the party

   ‘No! Rubén is the one I bring to the party.’

The twenty-four trials were divided into three conditions: (i) 12 trials with \([+a]\) contexts, for which we predicted the production of CLLD constructions, (ii) six trials with \([-a]\) contexts, for which we predicted FF or a cleft construction, and (iii) six fillers. We mirrored the unbalanced design from Slabakova et al.’s (2012) and Leal Méndez et al.’s (2015) AJT in our SPT with the intention of controlling its possible effect across tasks (see Section 2.1.2). To avoid item-ordering effects, we presented trials in six pseudo-randomized orders, and versions were counterbalanced across participants.

In order to control for potential confounding variables that could condition the validity of the data, all questions were in present tense and all words used were among the 5000 most frequent Spanish words (taken from (Davies 2006)). Additionally, we limited the length of the left-dislocated direct objects to a maximum of seven syllables, all of which were \([+\text{specific}], [+\text{definite}], [+\text{animate}], [+\text{human}], \text{masculine}, \) and required differential object marking.

The reader may find the stimuli and the complete data set in the Supplementary Materials of this manuscript.

---

9 López (2009) notes the following about the inclusion of negation preceding this type of utterance: “Since negation does the contrastive work, the focused constituent may just be plain information focus and the function ‘contrast’ and ‘focus’ are distributed among two different constituents” (p. 56). However, negation was included to help the participant understand that the focus would go to Rubén, and not Carlos. Although the presence of negation may explain the overextension of CLLD to \([-a]\) discourse contexts in our group results, the baseline participants were able to consistently distinguish between \([+a]\) and \([-a]\) discourse contexts.

10 These six fillers included dative experiencer predicates (verbs like gustar ‘to like’) in order to force the presence of a preposition \( a \) ‘to’ at the beginning of the sentence, as a parallel to the DOM in the experimental items.
3.1.3. Variables and Analysis

We fit a binomial logistic regression using the GENLINMIXED procedure for generalized linear mixed models (GLMM) in SPSS with the binary dependent variable Response (CLLD, FF/Cleft). The fixed effects were Discourse Context ([+a], [−a]), Group (baseline, heritage), and their interaction. We again chose the model with the maximal random effects specification that converged, which included a random intercept by subject and item, plus a random by-subject slope over Discourse. Additionally, to examine the possible effect of proficiency, a second model was fit to the heritage data. Its fixed effects were Discourse Context, Proficiency (38–46) as a continuous covariate, and their interaction; it also included a random intercept by subject and item, plus a random by-subject slope over the interaction of Discourse and Proficiency.

3.2. Results

3.2.1. Group Results

Figure 4 presents the proportions of CLLD responses by discourse context for each group.

![Figure 4](image_url) Proportion of clitic-doubled left dislocation (CLLD) responses by discourse context in the speeded production task (SPT) for the baseline and heritage groups. Error bars represent 95% confidence intervals.

As before (see Section 2.2.1), we report the tests of fixed effects which are central to answering our research questions. A Type III test of fixed effects yielded an effect of Discourse ($\beta = 3.95$, $F(1,17) = 72.24$, $p < 0.001$) and Group ($\beta = -1.96$, $F(1,28) = 5.43$, $p = 0.027$), but no Discourse × Group interaction ($\beta = 1.00$, $F(1,29) = 1.03$, $p = 0.319$). Pairwise comparisons revealed significant differences in the production of clitics between [+a] and [−a] discourse contexts for both the baseline group ($p < 0.001$) and the heritage group ($p = 0.003$). Furthermore, an odds ratio analysis indicated that the baseline group is 93.33 times more likely to produce CLLD in the [+a] discourse context than in the [−a] discourse context. Similarly, the heritage speakers are 37.46 times more likely to produce CLLD in the [+a] discourse context than in the [−a] discourse context.

3.2.2. Individual Variation

As with the AJT, we examined the results of the SPT individually. However, as before, establishing an a priori cut-off point to distinguish convergence from divergence was a challenge. As a result,
and similar to the AJT data, we refrain from making any claims based on the individual data for this SPT.

3.2.3. Results by Proficiency Score

A Type III test of fixed effects found no effect for Proficiency ($\beta = 0.005, F(1,40) = 0.48, p = 0.492$) and no significant Discourse × Proficiency interaction ($\beta = -0.49, F(1,88) = 0.76, p = 0.386$). Figure 5 presents the percentage of CLLD produced in the [+a] and [−a] discourse contexts by proficiency. As with the AJT, the fact that the lines are relatively flat and nearly the same slope reflects the lack of change in responses as proficiency increases.

![Figure 5. Proportion of CLLD in [+a] (dark blue) and in [−a] (light blue) on the SPT, by proficiency score.](image)

3.3. Interim Discussion 2

The group results show that the heritage group performs differently in the [+a] and [−a] discourse contexts in production, producing significantly more instances of CLLD in the [+a] discourse context. This is the same pattern found in the baseline group. Nonetheless, both the odds ratios and visual inspection of Figure 4 show that the heritage speakers made a weaker distinction between the two conditions in comparison to the baseline group, producing CLLD an average of about half the time in [−a] contexts. Yet what is important for understanding whether or not the heritage speaker group has acquired the relevant discourse restrictions is not whether they match the baseline group to the same degree, but rather whether they distinguish their production according to context. In other words, a significant distinction within subjects is enough to conclude that this group of speakers distinguishes in the use of CLLD between different discourse contexts. Therefore, we contend that these results do in fact show convergence between the heritage and baseline data.

For the analysis by proficiency score, the level of proficiency does not show evidence of affecting our heritage speakers. Visual inspection of Figure 5 suggests that part of this effect may be driven by two high-proficiency speakers who use CLLD 100% of the time in both contexts; without those speakers, the slope of the [−a] line may well have been more negative, differing from the [+a] condition and indicating a decrease in overmarking of CLLD with proficiency. However, in the absence of clear reasons to exclude these participants based on their individual variation, we take them as merely representative of the heterogeneity of heritage speaker grammatical knowledge and use, and it remains true that, overall, we do not observe any patterns by proficiency.
4. General Discussion

4.1. Summary of the Results

The goal of this project is to discover whether heritage speakers’ distinctions pattern with those of baseline speakers at the syntax–discourse interface under low (AJT) and high (SPT) processing pressure, by using Spanish CLLD in [+a] and [−a] discourse contexts as a test case (RQ1 and RQ2). The AJT shows that the baseline group makes a distinction in their judgments between utterances containing a clitic pronoun and those which do not in both discourse contexts. The heritage group makes the same type of distinction as the baseline group in the [+a] condition, but not in the [−a] condition, and shows no effect of proficiency level. On the other hand, the SPT shows a significant difference in the production of clitics in the [+a] versus [−a] conditions by both groups, and no effect of proficiency for the heritage group.

Comparing the results between two tasks that differ in processing pressure allows (a) for the replication of Leal Méndez et al. (2015) and (b) for moving the discussion beyond the limits of the Interface Hypothesis (Sorace 2011, 2012) and into additional considerations of the source(s) of heritage speaker divergence. Table 2, repeated here as Table 4, states the hypotheses we considered and the predictions that each made for the two tasks.

| Hypothesis                                | AJT  | SPT  |
|-------------------------------------------|------|------|
| Convergence                               | Convergence | Convergence |
| Processing and production issues          | Convergence | Divergence |
| Convergence competence masked by task effects | Divergence | Convergence |
| Divergent competence                      | Divergence | Divergence |

Our findings show the third pattern—divergence from the baseline on the AJT and convergence on the SPT—which corresponds with the hypothesis that heritage speaker grammars converge on the baseline grammar, yet issues related to the nature of comprehension tasks with a metalinguistic quality, like an AJT, may limit their performance.

4.2. Is Heritage Speaker Convergence Possible at the Syntax/Discourse Interface?

The SPT results show that the heritage group makes a distinction between the [+a] and [−a] discourse contexts in production, which aligns with the baseline pattern. Therefore, on the more naturalistic of the two tasks, they demonstrate convergence. Our reasoning is that if they can produce the appropriate distinction, even with some variability, it must be present in their grammars. We thus understand our data to confirm Leal Méndez et al.’s (2015) argument that CLLD does not pose special difficulty for heritage speakers, who are indeed able to converge on the baseline (i.e., distinguish between discourse contexts in production) for this construction. What to make, then, of the apparent divergence on the AJT? It is worth exploring this result in more detail.

4.3. Sources of Divergence

We have thus far contended that the pattern we observe in our results indicates relatively unproblematic heritage language acquisition of the discourse restrictions on CLLD (because they can use the construction appropriately), along with task effects that mask that acquisition. Previous work has proposed at least two possible sources of such task effects: linguistic insecurity leading to a yes-bias on judgment tasks and difficulties processing morphology in comprehension or interpretation tasks.

The first possibility is that the heritage speakers are unwilling to reject infelicitous uses of CLLD due to linguistic insecurity. Given that CLLD is relatively rare and that its misuse produces mere infelicity rather than ungrammaticality, CLLD is a plausible candidate for an insecurity-driven yes-bias that has been reported in heritage research (e.g., (Benmamoun et al. 2013a, 2013b; Polinsky 2018)).
However, note that the heritage speakers correctly reject the clitic-less sentences in the [+a] condition; it is only in the [−a] condition that they fail to make a distinction. If a generic yes-bias were at play, we might expect to see inflated ratings for all the infelicitous sentences, whereas what we observe instead is a selective lack of difference.

The second possibility is that the apparent deficit in the AJT is due to its nature as fundamentally a task of interpretation. Some previous evidence has shown that heritage speakers can have difficulty on comprehension or interpretation tasks, especially when there exists some ambiguity in the interpretation, which Polinsky and Scontras (2020a) call the Ambiguity Problem. It could be that our [−a] condition, which is the condition in which the fronted constituent is not clearly linked to a topic reading, resulted in a somewhat ambiguous context for the heritage speakers, which could explain their middling ratings of both sentence types in that condition.

What is curious for both these explanations of our AJT results is that they differ from those of Leal Méndez et al.’s (2015) original study. Their heritage speaker participants preferred clitics in the [+a] discourse context and absence of clitics in the [−a] discourse context, just like their control group (and ours). Our heritage speaker participants, however, did not show the expected difference in the [−a] discourse context. In any replication, despite careful planning, it is possible that any number of factors may not be identical to the first run of the study, including its participants, or when, where, or how it is conducted. Comparing our procedure to Leal Méndez et al.’s does not reveal any salient differences that could explain our different results. Instead, comparing the two studies may reveal the value of multiple replication in behavioral science research that includes multiple tasks in the experiment design: If we had presented only a replication of the AJT, we might have interpreted our findings as fully contradictory to Leal Méndez et al.’s, but taking into account the findings of the SPT leads us to interpret the results differently. When considering both tasks together, we argue that it is more likely that the group of heritage speakers is sensitive to the contextual appropriateness of clitic-doubling but that their awareness is masked in the AJT by task effects.

4.4. An Alternative Explanation

We have thus far argued that our results show that the heritage speakers are able to make the relevant discursive distinction in their production, if not to the same degree as the baseline, and that apparent differences in the AJT result from the nature of the AJT as a comprehension task with a metalinguistic element. Yet we recognize that the heritage speakers also produce substantially more CLLD than the baseline speakers in [−a] contexts, which is the same context in which they fail to make a distinction in the AJT. While we contend that the fact that they differentiate by context in their production shows successful acquisition of the discursive restrictions on CLLD, in line with Leal Méndez et al.’s (2015) results, it is worthwhile to consider alternative explanations.

An alternative interpretation of our findings could be to understand that the heritage speakers diverge from the baseline on both the tasks. If that is the case, it is worth considering the nature of the divergence—namely an extension of clitic-doubling even to inappropriate contexts. In other words, under this interpretation of our results, our speakers oversupply clitic-doubling. Polinsky and Scontras (2020a) provide two sets of evidence suggesting heritage speakers tend to oversupply morphological marking. The first is what they call the Silent Problem: the difficulty that heritage speakers encounter with expressing or interpreting meaning without a directly related overt form. For example, heritage speakers tend to overproduce overt pronouns in pro-drop languages, and they have difficulty interpreting VP ellipsis in heritage Dutch (Koornneef et al. 2011) and Russian (Polinsky 2016, 2018). The second is the Morphology Problem, which refers to the difficulty that heritage speakers have attending to inflectional morphology. Polinsky and Scontras argue that one of the most common compensation strategies on which heritage speakers rely is to explicitly overmark morphology. For example, heritage English speakers have been found to overuse regular past tense -ed (Duffield 2018), and heritage Russian speakers to overmark genitive case (Polinsky 2018). In informal terms, these speakers seem to prefer marking too much instead of not enough. Both these explanations
take the overmarking of overt morphology to be a strategy to improve processing economy in the heritage language, which tends to be these speakers’ less dominant language. If we take the SPT results to indicate such overmarking in production, our results could well be compatible with such an explanation.

If we re-interpret our results as overall divergence from the baseline, could we instead point to dominant-language transfer as its source? For some insight, we can compare our results with those of Laleko and Polinsky (2013, 2016, 2017) for topic marking in heritage Japanese. In baseline Japanese, the suffix -wa marks a constituent as a topic. Heritage speakers in these studies, however, used constituents in certain topic contexts without such a marker. According to Polinsky and Scontras (2020b), this sort of undersuppliance may be due to direct transfer from the dominant language. In the case of Laleko and Polinsky’s participants, the argument is that heritage Japanese lacks the topic marker -wa because there is no topic marker in their dominant language, English. If our heritage participants were subject to the same effect from their English, we might have expected them to undersupply clitic doubling, considering that, although a similar topicalization construction exists in English, it is not morphologically marked. Yet we find the opposite. For that reason, transfer from English seems an unlikely source of divergence from the baseline.

4.5. The Role of Proficiency

The third research question of the study asked whether proficiency plays a role in the knowledge and/or performance of heritage speakers with Spanish CLLD. As we saw in Sections 2.2.3 and 3.2.3, proficiency did not show any significant effect, nor significant interactions, in either experiment.

4.6. The Role of the Population: Heritage vs. L2 Speakers

Lastly, the fourth research question asked whether heritage speakers differ from the L2 learners in Sequeros-Valle et al. (2020) in their knowledge and/or production of Spanish CLLD. As we have seen up to this point, the heritage speakers in this study overextend clitic-doubling from the [+a] condition to the [−a] in the AJT, while making the same type of distinction as the baseline group in the SPT. We have argued that this result is indicative of issues related to the metalinguistic nature of acceptability tasks (Polinsky 2018). In contrast to the heritage speakers, the L2 learners in Sequeros-Valle et al. align with the baseline group in the AJT and overextend clitic-doubling from the [+a] to the [−a] condition in the SPT. The authors suggested that the learners are subject to processing limitations due either to the interface nature of CLLD (following Sorace 2011, 2012) or to real-time production issues beyond the interfaces (following Grüter et al.’s (2012) findings for L2 gender).

Given Montrul and Polinsky’s (2011) proposal that the Interface Hypothesis (Sorace 2011, 2012) should extend to heritage speakers by virtue of their being bilinguals, we should expect L2 learners and heritage speakers to show similar patterns. Instead, their patterns are opposite. This could be due to the (i) age of onset—our heritage speakers were exposed to Spanish from birth, while Sequeros-Valle et al.’s L2 learners were exposed to Spanish after the age of 12—(ii) the context of acquisition—our heritage speakers acquired Spanish naturally, while Sequeros-Valle et al.’s L2 learners acquired it in a classroom setting—or (iii) other factors (literacy, amount of formal education in the language, etc., including a combination of the above). Although we are not able to tease these apart with the current experimental design, the fact that the heritage speakers diverge from the baseline on the AJT while the L2 learners diverge from the baseline on the SPT could indicate a role of acquisition context. On the one hand, those speakers who acquired Spanish naturally (the heritage speakers) present issues on the more metalinguistic task (the AJT). On the other hand, those speakers who acquired Spanish in a classroom context (the L2 learners) present issues on the more naturalistic task (the SPT). It could well be that both populations show evidence of convergence toward the baseline patterns, but with some task effects specific to their acquisition histories.

Montrul (2008) reviews several studies that look into the correlation between population (L2 vs. heritage) and performance on specific task types (viz., written vs. oral). Although both our AJT and
SPT include written and aural language, her comparison provides some relevant parallels to ours since the task effects cited are related to acquisition context. For example, Matsunaga (2003) found that Japanese heritage speakers had more advanced oral proficiency than Japanese L2 learners; however, this difference was neutralized in a reading task. Furthermore, Montrul et al. (2008) found that L2 learners of Spanish were significantly more accurate than a group of heritage speakers on a written comprehension task and a written cloze test; however, on an oral picture naming task, the heritage speakers were significantly more accurate and had faster reaction times. Similarly, Montrul et al. (2006) found no differences between heritage and L2 learners in clitic placement in Spanish on a judgment task; however, on a visual matching task, the heritage speakers presented faster reaction times. In sum, these between-population differences according to task type lend support to the hypothesis that the different outcomes in the AJT versus SPT are due (at least in part) to task effects.

4.7. Future Research

Our conclusions regarding L2 and heritage speaker acquisition of Spanish CLLD takes the discussion of divergence and convergence beyond the interfaces: What has been interpreted previously as an interface issue may be the result of interlinking issues related to the tasks employed and the populations examined. In future work it would be worth focusing the lens beyond the interfaces to compare interface and non-interface phenomena with the same tasks and participants. This way, the predictions from the Interface Hypothesis (Sorace 2011, 2012; Sorace and Filiaci 2006) can be teased apart from general predictions on bilingualism within the same groups of participants.

Furthermore, as the comparison of our results with those of Leal Méndez et al. (2015) and Sequeros-Valle et al. (2020) has shown, there is value in triangulation across multiple tasks types. We therefore adopt Sequeros-Valle et al.’s suggestion that an ideal follow-up study of CLLD would include an offline AJT, an online processing task (like self-paced reading or eye-tracking), and a more naturalistic production task. Such triangulation could even lead to re-evaluating previous conclusions. For example, although it is widely claimed that topic constructions are especially vulnerable for bilinguals, our heritage speakers only appear to diverge on knowledge of topics on the AJT; they produce clitic-doubling to mark [+a] discourse contexts without any major issues. If the same is true of other interface phenomena, some of the previous evidence marking them as problematic based on AJTs may be missing something due to a lack of other sources of evidence. We do not intend to pile on the “AJTs are bad” bandwagon (we do not think they are), but rather to highlight the value of multiple sources of evidence, including production.

Supplementary Materials: The data collection tool for the SPT experiment (https://www.iris-database.org/iris/app/home/detail?id=york:938016) and the data set from both the AJT and the SPT experiments (https://www.iris-database.org/iris/app/home/detail?id=york:938017) are available at the IRIS database.

Author Contributions: Conceptualization, J.S.-V., B.H., and J.C.; methodology, J.S.-V. and B.H.; software, J.S.-V. for data, B.H. and J.C.; for statistical analysis; validation, J.S.-V., B.H., and J.C.; formal analysis, J.S.-V., B.H., and J.C.; investigation, J.S.-V., B.H., and J.C.; resources, B.H. and J.C.; data curation, J.S.-V.; writing—original draft preparation, J.S.-V., B.H., and J.C.; writing—review and editing, J.S.-V., B.H., and J.C.; visualization, J.S.-V.; supervision, J.S.-V., B.H., and J.C.; project administration, J.S.-V., B.H., and J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: We would like to thank many people, but especially Tania Leal and Roumyana Slabakova for sharing their data collection tool with us, Jason Rothman for his feedback on the general shape of the project, Luis López for his support on the theoretical aspects of the project, and the anonymous reviewers for their ideas to improve the manuscript. Lastly, we would not want to forget Aletia Rebollar, the research assistant who took care of a large extent of the data collection for this project.

Conflicts of Interest: The authors declare no conflict of interest.
References

Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68: 255–78. [CrossRef]

Belletti, Adriana, Elisa Bennati, and Antonella Sorace. 2007. Theoretical and developmental issues in the syntax of subjects: Evidence from near-native Italian. *Natural Language Linguistic Theory* 25: 657–89. [CrossRef]

Benmamoun, Elabbas, Silvina Montrul, and Maria Polinsky. 2013a. Heritage languages and their speakers: Opportunities and challenges for linguistics. *Theoretical Linguistics* 39: 129–81. [CrossRef]

Benmamoun, Elabbas, Silvina Montrul, and Maria Polinsky. 2013b. Defining an ‘ideal’ heritage speaker: Theoretical and methodological challenges. *Theoretical Linguistics* 39. [CrossRef]

Bianchi, Valentina, and Mara Frascarelli. 2010. Is topic a root phenomenon? *Iberia: An International Journal of Theoretical Linguistics* 2: 43–88.

Cabrelli Amaro, Jennifer. 2017. Testing the phonological permeability hypothesis: L3 phonological effects on L1 versus L2 systems. *International Journal of Bilingualism* 21: 698–717. [CrossRef]

Davies, Mark. 2006. *A Frequency Dictionary of Spanish*. New York: Routledge.

Duffield, Nigel. 2018. *Reflections on Psycholinguistic Theories. Raiding the Inarticulate*. Cambridge: Cambridge University Press.

Ellis, Rod. 2009. The differential effects of three types of task planning on the fluency, complexity, and accuracy in L2 oral production. *Applied Linguistics* 30: 474–509. [CrossRef]

Feldhausen, Ingo. 2016. Interspeaker variation, Optimality Theory and the prosody of clitic left-dislocations in Spanish. *Probus* 28: 293–333. [CrossRef]

Feldhausen, Ingo, and Maria del Mar Vanrell. 2015. Oraciones hendidas y marcación del foco estrecho en español: Una aproximación desde la Teoría de la Optimidad Estocástica. *Revista Internacional De Lingüística Iberoamericana* 26: 39–59.

Giancaspro, David. 2015. Code-switching at the auxiliary-VP Boundary. *Linguistic Approaches to Bilingualism* 5: 379–407. [CrossRef]

Grüter, Therese, Casey Lew-Williams, and Anne Fernald. 2012. Grammatical gender in L2: A production or a real-time processing problem? *Second Language Research* 28: 191–215. [CrossRef] [PubMed]

Hoot, Bradley. 2017. Narrow presentational focus in heritage Spanish and the syntax-discourse interface. *Linguistic Approaches to Bilingualism* 7: 63–95. [CrossRef]

Hulk, Aafke, and Natascha Müller. 2000. Bilingual first language acquisition at the interface between syntax and pragmatics. *Bilingualism: Language and Cognition* 3: 227–44. [CrossRef]

Koornneef, Arnaout W., Sergey Avrutin, Frank Wijnen, and Eric Reuland. 2011. Tracking the preference for bound-variable dependencies in ambiguous ellipses and only-structures. In *Experiments at the Interfaces*. Edited by Jeffrey Runner. Leiden: Emerald Publishers/Brill, pp. 67–100.

Laleko, Oksana, and Maria Polinsky. 2013. Marking topic or marking case: A comparative investigation of heritage Japanese and heritage Korean. *Heritage Language Journal* 10: 40–64.

Laleko, Oksana, and Maria Polinsky. 2016. Between syntax and discourse: Topic and case marking in heritage speakers and L2 learners of Japanese and Korean. *Linguistic Approaches to Bilingualism* 6: 396–439. [CrossRef]

Laleko, Oksana, and Maria Polinsky. 2017. Silence is difficult: On missing elements in bilingual grammars. *Zeitschrift für Sprachwissenschaft* 36: 135–63. [CrossRef]

Lane, Sean P., and Erin P. Hennes. 2018. Power struggle: Estimating sample size for multilevel relationships research. *Journal of Social and Personal Relationships* 35: 7–31. [CrossRef]

Lardiere, Donna. 2011. Who is the Interface Hypothesis about? *Linguistic Approaches to Bilingualism* 1: 48–53. [CrossRef]

Leal, Tania. 2018. Data analysis and sampling. In *Critical Reflections on Data in Second Language Acquisition*. Edited by Aarnes Gudmestad and Amanda Edmonds. Amsterdam: John Benjamins, pp. 63–88.

Leal, Tania, Emilie Destruel, and Bradley Hoot. 2018. The realization of information focus in monolingual and bilingual native Spanish speakers. *Linguistic Approaches to Bilingualism* 8: 217–51. [CrossRef]

Leal Méndez, Tania, Jason Rothman, and Roumyana Slabakova. 2015. Discourse-sensitive clitic-doubled dislocations in heritage Spanish. *Lingua* 155: 85–97. [CrossRef]

Lighthown, Patsy, and Nina Spada. 2006. *How Languages are Learned*. Oxford: Oxford University Press.

López, Luis. 2009. *A Derivational Syntax for Information Structure*. Oxford: Oxford University Press.
Marsden, Emma, Kara Morgan-Short, Sophie Thompson, and David Abugaber. 2018. Replication in second language research: Narrative and systematic reviews and recommendations for the field. *Language Learning* 68: 321–91. [CrossRef]

Matsunaga, Sashiko. 2003. Instructional needs of college-level learners of Japanese as a heritage language. Performance based analysis. *Heritage Language Journal* 1: 45–62.

Montrul, Silvina. 2002. Incomplete acquisition and attrition of Spanish tense/aspect distinctions in adult bilinguals. *Bilingualism: Language and Cognition* 5: 39–68. [CrossRef]

Montrul, Silvina. 2004. Subject and object expression in Spanish heritage speakers: A case of morphosyntactic convergence. *Bilingualism: Language and Cognition* 7: 125–42. [CrossRef]

Montrul, Silvina. 2008. *Incomplete Acquisition in Bilingualism: Re-examining the Age Factor*. Amsterdam: John Benjamins.

Montrul, Silvina. 2014. Structural changes in Spanish in the United States: Differential object marking in Spanish heritage speakers across generations. *Lingua* 151: 177–96. [CrossRef]

Montrul, Silvina. 2016. *The Acquisition of Heritage Languages*. Cambridge: Cambridge University Press.

Montrul, Silvina, and Maria Polinsky. 2011. Why not heritage speakers? A response to Sorace. *Linguistic Approaches to Bilingualism* 1: 58–62. [CrossRef]

Montrul, Silvina, Rebecca Foote, Silvia Perpiñán, Dan Thornhill, and Susana Vidal. 2006. Full access and age effects in adult bilingualism: An investigation of accusative clitics and word order. In *Selected Papers from the Hispanic Linguistic Symposium*. Edited by Nuria Sagarriga and Jacqueline Toribio. Somerville: Cascadilla Proceedings Project, pp. 217–28.

Montrul, Silvina, Rebecca Foote, and Silvia Perpiñán. 2008. Gender agreement in adult second language learners and Spanish heritage speakers: The effects of age and context of acquisition. *Language Learning* 58: 503–53. [CrossRef]

Montrul, Silvina, Rakesh Bhatt, and Roxana Girju. 2015. Differential object marking in Spanish, Hindi, and Romanian as heritage languages. *Language* 91: 564–610. [CrossRef]

Polinsky, Maria. 2006. Incomplete acquisition: American Russian. *Journal of Slavic Linguistics* 14: 191–262.

Polinsky, Maria. 2016. Structure vs. use in heritage language. *Linguistics Vanguard* 2: 1–14. [CrossRef]

Polinsky, Maria. 2018. *Heritage Languages and Their Speakers*. Cambridge: Cambridge University Press.

Polinsky, Maria, and Gregory Scontras. 2020a. Understanding heritage languages. *Bilingualism: Language and Cognition* 22: 4–20. [CrossRef]

Polinsky, Maria, and Gregory Scontras. 2020b. A roadmap for heritage language research: Response to the commentaries. *Bilingualism: Language and Cognition* 23: 50–55. [CrossRef]

Schütze, Carson T., and Jon Sprouse. 2013. Judgment data. In *Research Methods in Linguistics*. Edited by Robert J. Podesva and Devyani Sharma. Cambridge: Cambridge University Press, pp. 27–50.

Sequeros-Valle, Jose. 2020. A Morphological Approach to Discourse in Spanish: A Theoretical and Experimental Review. Ph.D. dissertation, University of Illinois, Chicago, IL, USA.

Sequeros-Valle, Jose, Bradley Hoot, and Jennifer Cabrelli. 2020. Clitic-doubled left dislocation in L2 Spanish: The effect of processing load at the syntax-discourse interface. *Language Acquisition* 27: 306–30. [CrossRef]

Silva-Corvalán, Carmen. 1994. *Language Contact and Change: Spanish in Los Angeles*. Oxford: Oxford University Press.

Slabakova, Roumyana, Paula Kempchinsky, and Jason Rothman. 2012. Clitic-doubled left dislocation and focus fronting in L2 Spanish: A case of successful acquisition at the syntax-discourse interface. *Second Language Research* 28: 319–43. [CrossRef]

Sorace, Antonella. 2011. Pinning down the concept of ‘interface’ in bilingualism. *Linguistic Approaches to Bilingualism* 1: 1–33. [CrossRef]

Sorace, Antonella. 2012. Pinning down the concept of interface in bilingual development: A reply to peer commentaries. *Linguistic Approaches to Bilingualism* 2: 209–17. [CrossRef]

Sorace, Antonella, and Francesca Filiaci. 2006. Anaphora resolution in near-native speakers of Italian. *Second Language Research* 22: 339–68. [CrossRef]

Sprouse, Jon, and Diogo Almeida. 2017. Setting the empirical record straight: Acceptability judgments appear to be reliable, robust, and replicable. *Behavioral and Brain Sciences* 40: E311. [CrossRef]

Tsimpli, Ianthi, and Antonella Sorace. 2006. Differentiating interfaces: L2 performance in syntax semantics and syntax discourse phenomena. In *Proceedings of the 30th Annual Boston University Conference on Language
Valenzuela, Elena. 2006. L2 end state grammars and incomplete acquisition of Spanish CLLD constructions. In *Inquiries in Linguistic Development: In Honor of Lydia White*. Edited by Roumyana Slabakova, Silvina Montrul and Philippe Prévost. Amsterdam: John Benjamins, pp. 283–304.

White, Lydia. 2011. The Interface Hypothesis? *Linguistic Approaches to Bilingualism* 1: 108–10. [CrossRef]

White, Lydia, n nonnatiElena Valenzuela, Martina Kozlowska-MacGregor, and Yan-Kit I. Leung. 2004. Gender and number agreement ive Spanish. *Applied Linguistics* 25: 105–33.

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.