

Article

Implicit and Explicit Knowledge of a Multiple Interface Phenomenon: Differential Task Effects in Heritage Speakers and L2 Speakers of Spanish in The Netherlands

Brechje van Osch *, Aafke Hulk, Suzanne Aalberse and Petra Sleeman

Faculty of Humanities, Linguistics Department, University of Amsterdam, Spuistraat 134,
1012 VB Amsterdam, The Netherlands; a.c.j.hulk@uva.nl (A.H.); S.P.Aalberse@uva.nl (S.A.);
A.P.Sleeman@uva.nl (P.S.)

* Correspondence: b.a.vanosch@uva.nl; Tel.: +31-61-785-0113

Received: 31 December 2017; Accepted: 4 July 2018; Published: 10 July 2018



Abstract: This paper compares heritage speakers and second language (L2) speakers of Spanish with Dutch as their dominant language, in order to explore the role of age of onset and manner of acquisition in the nature of the knowledge (implicit vs. explicit) of the subjunctive. Differently from previous studies, all items were presented orally and in written form, so that language mode of presentation could be excluded as a confounding factor. Moreover, the groups were matched on their general proficiency in Spanish using both an explicit and an implicit proficiency task. The results showed that the L2 speakers outperformed the heritage speakers in the explicit knowledge task and vice versa in the implicit knowledge task, suggesting that differential task effects, which thus far have only been attested for morpho-syntactic phenomena, can be extended to interface phenomena as well. These findings imply that age of onset and manner of acquisition have an influence in the way knowledge is represented in these two populations, and moreover emphasize the importance of using different task types in bilingual research.

Keywords: heritage speakers; second language acquisition; Spanish; bilingualism; explicit knowledge; implicit knowledge; subjunctive; interface phenomena; task-based differences; age of acquisition

1. Introduction

This study compares adult L2 speakers and heritage speakers of Spanish with comparable proficiency levels regarding their knowledge of the subjunctive, with the aim to explore the roles of age of onset and manner of acquisition. Heritage speakers are bilinguals who speak both the dominant language of the society they live in, and a minority language, which is passed on to them by their parent(s). Generally, these speakers receive a considerable amount of input in the home language in their first years of life, but experience a so-called dominance shift when they start school and the input in the dominant language increases drastically. By the time they reach adulthood, these speakers are typically dominant in the societal language; their heritage language has become their weaker language. Linguistically speaking, heritage speakers resemble L2 speakers¹ in many ways (Montrul 2012), but there are differences as well, which can be traced back to their respective acquisitional pathways. For instance, phonology and core syntax seem to be more robust in heritage

¹ In this paper, we use the term “L2 speakers” to refer to people who have learned a second (foreign) language postpuberty and in an instructional setting.

speakers than in L2 speakers (e.g., [Au et al. 2002](#); [Montrul 2006](#)), probably due to the fact that these modules are acquired early in childhood. Although morpho-syntax seems to be prone to deviation in both populations, recent research in this area has attested to differences between the two groups depending on the type of task: advantages for heritage speakers are attested in oral tasks which target implicit knowledge, while L2 speakers do better when it comes to written and more explicit tasks e.g., ([Montrul et al. 2008a](#); [Bowles 2011](#)). These task effects can be attributed to at least three differences between heritage and L2 speakers: (1) differences in “age of onset” of acquisition; (2) differences in “manner” of acquisition (naturalistic exposure vs. classroom instruction); and (3) differences in “mode” of exposure (oral vs. written). Clearly, heritage speakers acquire their language in childhood, whereas L2 speakers start learning an L2 in adulthood. On some critical period accounts, a fundamental difference between early and late language acquisition is that while early acquirers can rely on implicit acquisition and only need to be exposed to naturalistic input, late learners have lost this ability, and are forced to fall back on explicit learning mechanisms ([Bley-Vroman 1990](#); [DeKeyser 2000](#)).² But besides age of onset, heritage and L2 learners also differ with respect to the *manner* in which the language is acquired: while heritage speakers acquire their heritage language through exposure to naturalistic input, most L2 speakers learn a great deal of the second language through explicit instruction. Finally, the two populations differ regarding the language “mode” they are more familiar with: heritage speakers are generally more exposed to spoken language than written language, while the opposite is true for L2. The last factor, language “mode”, has often been a confounding factor in studies: generally, the tasks used to target explicit knowledge are written tasks, whereas the implicit tasks are oral production tasks. In the present study, all stimuli are presented both aurally and in written form, with the aim to rule out a confounding effect of language mode.

While task-based effects have only been reported for morpho-syntactic structures, there is no reason to assume that similar effects should not apply in other linguistic domains. In the present paper, we focus on a multiple interface phenomenon: the subjunctive in syntactically, semantically and pragmatically constrained contexts. The results for the heritage speakers have been reported elsewhere ([Van Osch and Sleeman 2016](#); [Van Osch et al. 2017](#)). In this paper, these findings are compared to results by L2 speakers on the same tasks. We will demonstrate that differential task effects apply to this interface phenomenon as well. We moreover explore whether these task effects are more likely to be the result of differences in age of onset of acquisition or in manner of acquisition.

The following section introduces the linguistic background on the subjunctive in the three linguistic contexts that are relevant to the present paper: volitional predicates, relative clauses and negated sentences. Section 3 discusses the distinction between explicit and implicit knowledge and the operationalization of these constructs. A summary of previous research in heritage and L2 Spanish is presented in Section 4, followed by the formulation of the hypotheses in Section 5. Section 6 describes the methods of the present study. Section 7 reports the results, which will be discussed in relation to previous findings and theories about age effects in (second) language acquisition in Section 8. Section 9 contains a brief conclusion of the paper.

2. The Subjunctive

Mood in Spanish is a multiple interface phenomenon, because it is governed by syntactic, semantic and pragmatic constraints. Two moods can be distinguished: indicative and subjunctive, of which the latter is the focus of the present paper. Dutch, the dominant language of the bilinguals in this study, does not exhibit productive use of the subjunctive ([Thieroff 2004](#)). Our study includes three different contexts in which subjunctive is required: first, there are those contexts where mood is syntactically

² Other accounts mention different explanations for the critical period effect, such as affective-motivational factors ([Krashen 1982](#)), L1 influence (e.g., [Flege 1999](#)), socio-educational factors (e.g., [Bialystok and Hakuta 1999](#)), and time on the task (e.g., [Flynn and Manuel 1991](#)).

selected by the verb. Verbs that are volitional in nature, like *querer* ('to want') or *esperar* ('to hope') obligatorily select subjunctive, as illustrated in in example (1).

1. Quiero que me *ayudas/ayudes
 Want-1SG that me help-IND.2SG/help-SUBJ.2SG
 'I want you to help me.'

In this type of context, which is generally considered to be morpho-syntactic in nature, the wrong mood leads to ungrammaticality. However, there are other contexts in which mood alternation occurs and the choice of mood depends on semantic and pragmatics factors. An example is sentences containing relative clauses, such as (2):

2. Busco Una blusa que tiene/tenga botones grandes
 look-1SG a shirt that have-IND.3SG/have-SUB.3SG big buttons
 'I'm looking for a shirt that has big buttons.'

Whenever the sentence starts with a verb like *buscar* ('to look for'), a volitional construction such as *querer comprar* ('to want to buy'), or a future reference like *compraré* ('I will buy'), the choice of mood depends on the specificity of the antecedent. In example (2), if the speaker refers to a specific shirt, indicative is more appropriate, but if he is looking for any shirt (as long as it has big buttons), subjunctive is more felicitous. Given that the specificity (a semantic feature) of the antecedent is involved, this use of the subjunctive pertains to the interface between syntax and semantics following (Borgonovo et al. 2015).

Another type of sentence in which choice of mood depends on the context is sentences with negated epistemic, communication or perception verbs, such as (3):

3. Juan no cree que María está/esté embarazada
 John NEG believe-3SG that Mary is-IND/is.SUB pregnant
 'John does not believe that Mary is pregnant.'

In this type of sentence, the subordinate verb can be in either the indicative or the subjunctive mood, depending on whether the speaker wishes to approach the event expressed in the embedded clause from his own perspective or from the perspective of the matrix subject (Quer 2001). If the speaker disagrees with the statement in the embedded clause (that is, if the speaker thinks that Mary is in fact pregnant), he can use the indicative to emphasize that he views the event from his own epistemic model. Subjunctive on the other hand would indicate a shift from the epistemic model of the speaker to that of the matrix subject. Thus, in the case that the speaker believes the proposition in the embedded clause to be true, both moods are possible, depending on the model of evaluation. However, Van Osch et al. (2017) demonstrated that in this situation the specific matrix verb plays a role as well: with perception and communication verbs monolingual speakers of Spanish prefer indicative, whereas with epistemic verbs subjunctive is the preferred option.

If, on the other hand, the speaker agrees with the matrix subject (i.e., neither speaker nor matrix subject believe that Mary is pregnant), subjunctive is unambiguously the most felicitous option. This use of the subjunctive is generally considered to pertain to the interface between syntax and pragmatics.

As for instruction of the subjunctive in L2 acquisition, Mikulski (2006) reports that the subjunctive is generally introduced early in Spanish L2 curricula. Most textbooks start with an explanation of the formal characteristics of the present subjunctive. The use of the subjunctive is typically formulated in terms of doubt, uncertainty, non-assertion, non-specificity and presupposedness. Generally, the obligatory contexts for the subjunctive are explained first, for instance the subjunctive in volitional constructions (one of the contexts discussed in this study), with affective predicates (*me alegro de que* 'I'm happy that' + SUB) or following certain propositions (*para que* 'so that' + SUB). Variable uses of the subjunctive are introduced later, but the exact order varies by textbook: some textbooks introduce relative clause contexts before negation contexts, while others treat them in the opposite

order. However, all three contexts for the subjunctive tested in this study are typically introduced within the first year of the curriculum. Interestingly, the negation construction is generally only explained with epistemic verbs in the first person (e.g., *No creo que* 'I don't think that'), not with communication and perception verbs and not in the third person, which was the context tested in the present study. This means that any knowledge L2 speakers acquire of the subjunctive in contexts broader than *No creo que* has to come from either exposure in the input or from their ability to recognize the pattern (namely, that in these contexts subjunctive refers to non-assertion from the perspective of the speaker) and extend this pattern to constructions with other verbs.

Despite the extensive instruction L2 speakers receive on the subjunctive, it remains a vulnerable area for this population, especially in those contexts where mood is variable and determined by semantic and pragmatic features (Iverson et al. 2008; Borgonovo et al. 2015). Similar problems are reported for heritage speakers of Spanish e.g., (Silva-Corvalán 1994; Montrul 2009). In this paper, we ask the question whether one of these groups has an advantage over the other, and whether this advantage might differ depending on the type of knowledge we are looking at: implicit or explicit. The following section discusses these notions in more detail.

3. Explicit vs. Implicit Knowledge

What is meant exactly by the terms explicit and implicit? This is not an easy question, since different studies adopt different definitions. However, the key component seems to be awareness/consciousness. Explicit knowledge is knowledge someone is aware/conscious of, whereas implicit knowledge lies outside awareness (e.g., DeKeyser 2003; Ellis 2009a, 2009b). Ellis (2009a) adds that access to implicit knowledge is automatic, and therefore fast and effortless, whereas explicit knowledge requires attentional control and is thus more time-consuming. He furthermore mentions that implicit knowledge can only become evident in linguistic behavior, whereas explicit knowledge can be expressed in words. Not all scholars agree that a distinction between explicit and implicit knowledge is psychologically and neurally real (e.g., Shanks 2003). However, as Ellis (2009b) argues, the mere fact that speakers are capable of correctly applying linguistic rules without being able to verbally explain those rules, as well as the opposite situation, namely that speakers know the rule but are unable to apply it correctly, implies that two different types of knowledge are involved. Moreover, there exists neurological evidence that both types of knowledge are stored in different parts of the brain (Paradis 1994).

An important question is how to operationalize the concepts explicit and implicit knowledge. In early research, many different instruments were used to measure the same construct. Ellis (2005, 2009b) attempted to provide a valid and reliable test battery that could be consistently used by all researchers. The test battery contained: (1) an elicited oral imitation test; (2) an oral narrative test; (3) a timed grammaticality judgment test; (4) an untimed grammaticality judgment test; and (5) a metalinguistic awareness test. These tests were designed so that they would differ maximally on the following 4 criteria:

1. Degree of awareness. Explicit knowledge is considered to be conscious; implicit knowledge is not.
2. Time availability. Implicit knowledge is assumed to be accessed automatically and fast, whereas explicit knowledge requires controlled processing and thus is more time-consuming.
3. Focus of attention. Tasks that focus on fluency (focus on meaning) are considered to test implicit knowledge, whereas tests that prioritize accuracy (focus on form), tap into explicit knowledge.
4. Metalanguage, used to verbalize linguistic rules, is related to explicit, but not implicit knowledge.

Results from both native and L2 speakers of English showed that the 5 tasks could be grouped into two clusters: the elicited oral imitation test, the oral narrative test and the timed grammaticality judgment task on the one hand and the untimed grammaticality judgment task and the metalinguistic knowledge test on the other hand, providing evidence that they load on different factors. Ellis (2005, 2009b) concluded that the first three tasks are more likely to test implicit knowledge and

the last two tap into explicit knowledge. The results furthermore showed no evidence for a distinction between production vs. judgment, given that the timed grammaticality judgment task clustered together with the two oral production tasks. The same pattern of clustering of these tasks was replicated for Spanish by Bowles (2011).

The design of the present study is based on three of the four criteria described above, namely: degree of awareness, time availability and focus of attention, as will be discussed in the method section. In the next section, we look at previous studies comparing heritage speakers and L2 speakers, highlighting those that have used different task types, and focusing on the subjunctive in particular.

4. Previous Research

4.1. Heritage Speakers and L2 Speakers Compared

Montrul (2012) offers an overview of scientific research exploring the similarities and differences between heritage speakers and L2 speakers. She concludes that, although both populations often diverge from the native speaker norm and make similar types of errors, heritage speakers seem to have an advantage over proficiency-matched L2 learners, but only in certain linguistic modules. For instance, when it comes to the perception and production of phonological features, quite a lot of evidence shows that heritage speakers are closer to monolinguals than L2 speakers. This has been found for Spanish as a heritage language (Au et al. 2002; Knightly et al. 2003) but also for other languages like Korean (Oh et al. 2003), Mandarin (Chang et al. 2008) and Russian (Lukyanchenko and Gor 2011).

Another domain where early acquisition seems to present an advantage is core syntax. An interesting pilot study by (Håkansson 1995) showed that heritage speakers of Swedish were more target-like than L2 speakers with V2 (syntax), while they were less accurate with gender agreement (morpho-syntax). Montrul has carried out several studies showing that heritage speakers of Spanish outperform L2 speakers when it comes to purely syntactic phenomena, such as the syntax of subjects and objects (Montrul 2006), wh-movement in questions with (embedded) subject and object extraction in Spanish (Montrul et al. 2008a) and the syntax of clitics (Montrul 2010; Montrul et al. 2006).³

To our knowledge, morpho-syntax is the only domain in which differential task effects have been reported for heritage and L2 speakers. (Bowles 2011) for instance, used the same test battery as (Ellis 2005, 2009b) for a whole range of (morpho-)syntactic structures in Spanish, among which is the subjunctive (but results are not reported for the separate structures), and found that heritage speakers performed better on the implicit tasks and L2 speakers on explicit tasks. Montrul et al. (2008b, 2014) report a similar pattern for gender assignment and agreement in Spanish.

Interface phenomena are generally found to be notoriously vulnerable in both heritage and L2 speakers (e.g., Montrul 2008, Iverson et al. 2008). In her overview, Montrul (2012) does not mention advantages for one group over the other. Nevertheless, some of her studies find that Spanish heritage speakers are more similar to monolinguals than L2 speakers with certain interface phenomena, such as felicitous use of overt subjects syntax-discourse interface (Montrul 2006) and unaccusativity syntax-semantics interface (Montrul 2005). Not all research confirms this pattern, though. Montrul (2004) reports no difference between the two groups with aspectual distinctions (syntax-semantics). Similarly, Montrul and Ionin (2012) fail to attest differences between the two groups with the generic interpretation of definite articles (syntax-semantics interface). Keating et al. (2011) attest mixed results for anaphoric resolution (syntax-discourse interface): the heritage speakers were less target-like with overt pronouns, which they interpreted as referring to the subject more than L2 speakers did, whereas L2 speakers deviated more with null pronouns, in that they interpreted them less often as referring to the matrix subject.

³ But see Bruhn de Garavito (2002) for a study showing no difference between heritage and L2 speakers with verb movement in Spanish.

To sum up, when it comes to phonology and syntax, the evidence for an advantage for heritage speakers over L2 speakers is quite convincing. For morpho-syntax, there seem to be differential task effects. Phenomena lying at the interface between syntax and other domains seem to be vulnerable in both bilingual populations.

4.2. The Subjunctive

The specific phenomenon of interest for the present study is the subjunctive in Spanish. As described in Section 2, this phenomenon can be considered to pertain to morpho-syntax, the syntax-semantic interface or the syntax-pragmatics interface, depending on the context in which it is used. The subjunctive has been investigated extensively, predominantly in Spanish-English bilingual populations in the U.S. The literature generally shows the subjunctive to be a vulnerable area, particularly in those contexts where mood is variable and depends on semantic and pragmatic factors, and this applies to both heritage speakers (e.g., [Silva-Corvalán 1994](#); [Montrul 2009](#); [Van Osch et al. 2017](#)) and L2 speakers of Spanish ([Iverson et al. 2008](#); [Borgonovo et al. 2015](#)).

Some studies have compared heritage to L2 speakers on their knowledge of the subjunctive. [Montrul and Perpiñán \(2011\)](#) found that L2 speakers outperformed heritage speakers on a morphology recognition task targeting obligatory subjunctive and on a sentence conjunction task (which was considered more implicit by the authors) containing sentences with *cuando* ('when'), *de manera que* ('so that'), and relative clauses, in which mood selection depends on the context. [Mikulski \(2010\)](#), on the other hand, demonstrated that heritage speakers outperformed L2 speakers in a grammaticality judgment task and an editing task targeting the volitional subjunctive. In an elicited production task, [Mikulski and Elola \(2013\)](#) also found an advantage for heritage speakers over L2 speakers with the subjunctive in advice constructions (e.g., *aconsejar/recomendar* + SUB 'to advise/recommend that + SUB'), but no such difference was attested by [Lynch \(2008\)](#) in a spontaneous production task. These different findings may be due to differences in experimental design or possibly to differences w.r.t. the specific subjunctive-targeting contexts, which were not specified in [Lynch \(2008\)](#).

Two studies have used both explicit and implicit knowledge measures of heritage and L2 speakers' knowledge of the subjunctive. [Montrul \(2011\)](#) tested heritage and proficiency-matched L2 speakers using an explicit forced choice task and an implicit oral narrative task. The study included several phenomena, among which (obligatory) choice of mood. The L2 speakers outperformed the heritage speakers on the explicit task, and the reversed pattern was attested in the implicit task. However, this pattern was not confirmed by [Potowski et al. \(2009\)](#), who tested heritage and L2 speakers on the past subjunctive in sentences with indefinite or non-existent antecedents using a written interpretation task, a written grammaticality judgment task and a written production task. On all three tasks, the heritage speakers outperformed the L2 speakers.

No clear conclusion can thus be drawn from these studies, in part because different studies include different contexts for mood and we do not always know which aspects of mood (morpho-syntax, syntax-semantics, syntax-pragmatics) are targeted.

4.3. Problems with Previous Studies

At this point, we need to point out certain problematic issues in previous studies. First of all, in most studies reporting differential task effects (i.e., heritage speakers performing better on implicit tasks, and vice versa on explicit tasks), the explicit tasks contain only written language, whereas the implicit tasks are generally oral production tasks (e.g., [Montrul et al. 2008b](#); [Montrul 2011](#); [Bowles 2011](#)). This means that there is a second, possibly confounding factor, namely language mode (oral vs. written). Given that L2 speakers are relatively more familiar with written language, and heritage speakers with spoken language, we cannot unequivocally conclude that the respective advantages for each group are due to the explicitness of the task, and not to the language mode (oral vs. written) in which the items were presented. [Montrul et al. \(2014\)](#) tried to get around this problem by presenting all items only aurally. However, using only aural stimuli may still give the

heritage speakers an advantage over L2 speakers across tasks, given that the former have more experience with spoken language. For this reason, the present study includes a simultaneous bimodal presentation of the items.

Another debatable matter in previous research relates to the way in which heritage speakers and L2 speakers' general proficiency is matched. In order to draw strong conclusions about specific task-based differences between heritage speakers and L2 speakers, it is imperative that the groups do not significantly differ from each other on a general measure of proficiency. Unfortunately, some studies do not match the groups at all (e.g., Lynch 2008; Mikulski 2010; Mikulski and Elola 2013), and those that do often determine proficiency based on unreliable measures such as self-evaluations (e.g., Keating et al. 2011) or course level (e.g., Potowski et al. 2009; Bowles 2011). Now, the question is: what is an appropriate measure of proficiency? Montrul (2005), Montrul and Ionin (2012), and Montrul and Perpiñán (2011) typically apply parts of the DELE (Diplomas de Español como Lengua Extranjera), the proficiency task used by the Spanish Ministry of Education, Culture and Sport to grant official diplomas of competence in the Spanish language (<http://www.dele.org>). A problem with this task is that it is (1) written and (2) rather explicit in nature (untimed, and focused on form). As pointed out by Valdés (1995), while the DELE may give an accurate indication of L2 proficiency, it may not be the most reliable measure to compare L2 speakers to heritage speakers. After all, if our assumptions about the differences between heritage and L2 speakers are accurate, using exclusively the DELE increases the risk of an underestimation of the relative proficiency of heritage speakers. To avoid this issue, in the present study we include both the DELE and an aural lexical decision task as matching criteria, the latter of which contains spoken language and taps into more implicit knowledge due to the time pressure. If this task constitutes an advantage for one of the groups, we assume the advantage will be for the heritage speakers, who are more familiar with spoken language, and presumably do better on implicit tasks.

5. Research Questions and Hypotheses

This study compares L2 and heritage speakers of Spanish with Dutch as their dominant language, regarding their knowledge of the subjunctive. Our research question is:

Will L2 speakers and heritage speakers of Spanish with comparable general proficiency levels have differential advantages depending on whether explicit or implicit knowledge of the subjunctive is tested?

Based on previous findings in the area of morpho-syntax, we hypothesize that the L2 speakers will outperform the heritage speakers on an explicit task and vice versa on a task measuring implicit knowledge.

6. Method

This study was part of a bigger project in which both the subjunctive and the indicative were tested. We focus only on the subjunctive-targeting items in this paper, both because of space limitations, and because of the ambiguous predictions in sentences with negation, as discussed in Section 2. For the heritage speaker's results on the indicative see (Van Osch et al. 2017).

6.1. Participants

In total, 27 heritage speakers, 28 L2 speakers and 18 monolingual speakers participated in the study. However, several participants were excluded in order to maximize homogeneity. First of all, a selection was made based on the participants' scores on a Morphology Recognition Task (MRT), which served to check whether participants knew the correct form of the subjunctive following (e.g., Montrul 2009, 2011; Iverson et al. 2008). Only those participants who scored higher than 80% on this task were included following (Iverson et al. 2008). Furthermore, both the DELE and an aural lexical decision task served as selection criteria. Participants were included if they scored higher than 36 on

the DELE, corresponding to a proficiency level of high-intermediate to advanced, and if they had more than 100 (out of 149) items correct on the lexical decision task.⁴ One L2 speaker was excluded because she was considerably older than the other participants. After these exclusions, 17 heritage speakers, 21 L2 speakers and 18 monolinguals remained. The L2 and heritage speakers did not differ significantly regarding their scores on the DELE ($t = -0.66, p = 0.51$), or the lexical decision task ($t = 0.24, p = 0.81$). However, there was a difference in terms of their self-reported proficiency: heritage speakers rated themselves significantly higher than L2 speakers ($t = -4.07, p < 0.001$). This difference should be taken with a grain of salt, since self-assessments, especially by heritage speakers, are not very reliable (Benmamoun et al. 2010). The L2 speakers also performed significantly better on the MRT than the heritage speakers ($t = 2.12, p = 0.04$), which is not surprising, considering this task is written and explicit in nature. Both bilingual groups differed significantly from the monolingual speakers on all proficiency measures (DELE: heritage speakers (HS) vs. monolinguals: $t = 5.59, p < 0.001$, L2 vs. monolinguals: $t = 6.64, p \leq 0.001$; lexical decision task: HS vs. monolinguals: $t = 7.87, p < 0.001$, L2 vs. monolinguals: $t = 7.54, p < 0.001$; self-reports: HS vs. monolinguals: $t = 4.32, p < 0.001$; L2 vs. monolinguals: $t = 15.79, p < 0.001$) as well as on the MRT (HS vs. monolinguals: $t = 3.59, p = 0.002$; L2 vs. monolinguals: $t = 2.11, p = 0.04$). The three groups did not differ significantly in age. Table 1 gives an overview of the participants' most important characteristics.

Table 1. Proficiency scores, ages and MRT per group.

Group	N	Age	DELE	Lexical Decision Score	Self-Reported Proficiency	MRT Score
Monolinguals	18	26.4	45.22	130.72	5.94	98.97
Heritage speakers	17	25.9	40.94	110.41	5.13	92.16
L2 speakers	21	28.1	40.43	110.95	4.32	96.30

DELE: Diploma de Español como Lengua Extranjera; MRT: Morphology Recognition Task.

An extensive questionnaire gave insight into the participants' linguistic backgrounds.

The heritage speakers (3 male, 14 female) were all university graduates or students at the time of testing. Different varieties of Spanish were included in the sample: mostly peninsular and Mexican Spanish, but also Colombian, Uruguayan and Argentinian Spanish. To our knowledge, no dialectal variation has been reported for the uses of the subjunctive in the contexts discussed in this paper. The majority (16) were second-generation heritage speakers, who had been exposed to both languages from birth as they had one Dutch-speaking parent and one Spanish-speaking parent. While the Spanish-speaking parents spoke mostly Spanish, many of them had also learned Dutch at some point, and spoke some Dutch at home as well. As tends to be the case with heritage speakers, many of the participants experienced a relative drop in input and output in Spanish at home throughout their lives, but for others it remained more or less stable, and in a few cases it even increased. Most heritage speakers reported frequent visits to the home country throughout their lives and a strong emotional bond with both the country and the heritage language. The majority had received some formal instruction in Spanish, either as a subject in primary school, high school or college, or at a Saturday school. Three heritage speakers were enrolled in a bachelor's program of Spanish at the time of testing. There was thus some variability regarding the amount of exposure to explicit instruction in Spanish within the heritage group.

As for their language use at the time of testing, Dutch was clearly the dominant language, especially at work and in contact with friends. At home, about half of the participants reported using both Spanish and Dutch. English was also a commonly used language at university or in the workplace, as well as with friends and in some cases at home as well. When asked about their knowledge of

⁴ For reference, consider that the lowest native speaker score was 110.

languages, while most participants considered themselves to be dominant in Dutch, some of them identified as balanced bilinguals in Dutch and Spanish.

The L2 speakers (5 male, 16 female), like the heritage speakers, were all university students or graduates who were exposed to Spanish after the age of 15. Unlike in most L2 studies, they were not all students in the same course. In fact, there was quite some variation regarding the amount and the type of instruction they had received: from only two-and-a-half months to 10 years in total (on average, the total months of instruction was 2 years and 8 months). Unfortunately, we do not have information about the exact contexts of the subjunctive in which the L2 speakers had received explicit instruction. However, since most of them (15 out of 21 participants) reported having received instruction for at least one year, at least these speakers are expected to have been introduced to all three contexts for the subjunctive tested in this study.

The L2 speakers furthermore varied considerably regarding their amount of exposure to naturalistic input in Spanish. Some L2 speakers had spent only 5 months abroad, whereas others had lived in a Spanish-speaking country for almost 4 years. Some participants had been in a relationship with a native speaker or were so at the time of testing, and used considerably more Spanish than Dutch at home than others. Some people also reported Spanish as (one of the) dominant languages at work/school and/or a common language used with friends.

All L2 participants considered themselves native speakers of Dutch. Regarding their self-reported proficiency in Spanish, the majority rated themselves as either advanced or near-native. Given The Netherlands' internationally oriented society, it is no surprise that all participants (heritage and L2 speakers alike) reported advanced to near-native knowledge of English, as well as of other languages (in varying proficiency levels), such as German, French (obligatory subjects in high school), one of the country's minority languages, and in some cases Italian or Portuguese.

The control group, consisting of 18 monolingually raised native speakers of Spanish (5 male, 13 female), were tested in the Netherlands. All of them were recent immigrants, who had arrived less than six months before the time of testing. Their countries of origin were Spain (9), Mexico (4), Colombia (2), Argentina (1), Nicaragua (1) and Venezuela (1). All speakers were native speakers of Spanish, but had learned English as a second language later in life. Their self-reported proficiency in English ranged from intermediate to highly advanced. They had no knowledge of Dutch whatsoever.

6.2. Tasks and Procedure

First of all, an extensive background questionnaire was administered to obtain detailed information about the participants' family situation, their language dominance, their exposure and use of different languages in various environments and their attitudes towards Spanish. Hereafter followed three tasks that were carried out on a laptop: a lexical decision task, an elicited production task and an acceptability judgment task. The lexical decision task served to measure the participants' overall proficiency in Spanish. This task was included in addition to the DELE, a standardized test generally used in Spanish L2 and heritage studies which is explicit in nature and contains only written language. As discussed in Section 4.3, using only the DELE to match the participants' general proficiency could lead to a relative overestimation of the L2 speakers' proficiency, due to their increased familiarity with written language and their putative advantage with explicit tasks. In the lexical decision task, participants were aurally presented with Spanish words and non-words, and had to decide as quickly as possible whether the word they heard was a real Spanish word or not, by pressing either a green or a red button. While we do not know of research attesting to an advantage for early bilinguals over late bilinguals on word recognition tasks, considering that this task is aural and less explicit (given the time pressure) this test was assumed to provide an advantage for the heritage speakers rather than the L2 speakers and thus minimize the risk of a relative underestimation of the heritage speakers' general proficiency. There was a strong and significant correlation between the two proficiency tasks ($r = 0.72$; $p < 0.001$), and a moderately significant correlation between both proficiency tasks and the participants'

self-reported proficiency (self-reports & DELE: $r = 0.57$, $p < 0.001$; self-reports & lexical decision task: $r = 0.62$, $p < 0.001$).

After the lexical decision task followed the elicited production task and the acceptability judgment task, which are described in detail below. At the end of the session, the participants were administered the MRT and the DELE, which were both paper-and-pencil tasks. The MRT contained sentences targeting the subjunctive in obligatory contexts, such as following volitional predicates or the construction *es + adjective + que ...* ('It is [adjective] that ...') followed by both the indicative and the subjunctive form of the verb, between which the participant had to choose. Following previous research (i.e., Montrul 2009, 2011; Iverson et al. 2008), this task was included to check whether the participants had accurate knowledge of the "form" of the subjunctive. After all, in order to deduce anything about participants' accurate "use" of the subjunctive, it is crucial that we know whether they indeed recognize the subjunctive as such. Given that heritage speakers typically lack meta-linguistic knowledge of their heritage language, it is impossible to explicitly ask them in which mood a certain verb is conjugated. Therefore, the MRT introduces the subjunctive forms in obligatory contexts, assuming that heritage speakers are aware that these contexts require the subjunctive. Even though this design may have resulted in the exclusion of some participants based on their inaccurate knowledge of subjunctive "use" in these contexts, at least we can assume that all participants scoring higher than 80% in any case have highly accurate knowledge of "form" as well.

The DELE consisted of a vocabulary part containing 30 fill-in-the-gap sentences, and a cloze test, targeting both vocabulary and grammatical knowledge of Spanish. In total, the experiment took about 2 to 2.5 h to complete. All subjects were paid 10 euros for their participation and signed an informed consent form (file number ethical committee: 2014–2013).

As mentioned earlier, in the two subjunctive tasks all items were presented both written and aurally, to avoid any influence of the language mode of presentation. The recordings were made by a native speaker of Colombian Spanish, who was instructed to speak slowly and clearly. The two tasks were designed in such a way that they would differ maximally on three of the characteristics mentioned by Ellis (2009b): time availability, focus of attention and awareness.

In the elicited production task, participants were presented with short stories in Spanish. After each story, the participants would read aloud the beginning of a sentence, which they were asked to finish in a way that made sense in the provided context. This task was assumed to mostly tap into implicit knowledge. Firstly because there was no time for controlled processing; after 10 s the test automatically proceeded to the next item. Furthermore, the focus of attention was on meaning, rather than form, since the instruction was simply to end the sentence in a way that made sense in relation to the preceding story. And finally, given the inclusion of fillers and the focus on meaning, it was unlikely that people would be conscious of the topic of investigation. A debriefing session confirmed that none of the participants were aware of what they were tested on.

The task contained three conditions for the subjunctive. The first condition contained purely morpho-syntactic uses of mood, in which the subjunctive was syntactically selected by the main verb, which was always a volitional predicate such as *querer* ('to want'), *esperar* ('to hope'), or *aconsejar* ('to advise'), for instance in example (4).

4. *Estoy molesto porque mi esposa nunca limpia la casa. Esta noche de nuevo no me ayuda a lavar los platos.*
Me enojo y le digo:
 'I'm annoyed because my wife never cleans the house. This evening once again she does not help me with the dishes. I get angry and I say:'
Quiero que...
 'I want that ...'

In this case, any answer containing a verb in the subjunctive would be correct, because the main verb *querer* ('want') obligatorily requires a subjunctive verb in the subordinate clause.

In the relative clause condition, the target sentence would always start with either a verb like *buscar* ('to look for') or a reference to the future, like *compraré* ('I will buy'), followed by an indefinite

inanimate object with a relative clause. The context made clear that the antecedent was non-specific (thus targeting subjunctive). An example of an item of this type is (5):

5. *Camilo está de vacaciones en Málaga. Le gustaría mucho comer tapas en un restaurante y ver un show de flamenco. Va al centro antiguo y pregunta a alguien en la calle:*
 ‘Camilo is on a holiday in Malaga. He would very much like to eat tapas in a restaurant and see a flamenco show. He goes to the old centre and asks someone in the street:’
Busco un restaurante de tapas donde...
 ‘I’m looking for a tapas restaurant where ... ’

In this case, the participant could say something like *toquen flamenco* (‘they play.SUB flamenco’) or any other answer including a subjunctive verb form.

In the third condition, the main clause contained a negated epistemic (e.g., *creer* ‘to think’), perception (e.g., *ver* ‘to see’) or communication predicate (e.g., *decir* ‘to say’). The context served to make clear that the speaker was not committed to the truth of the proposition in the embedded clause, targeting subjunctive mood. An example is shown in (6):

6. *Selma camina por la calle y ve a su tía caminando a 20 metros de ella. La llama, pero hay mucho ruido de los coches así que es imposible oír algo.*
 ‘Selma is walking on the street and sees her aunt walking at a distance of 20 meters. She calls her, but there is a lot of noise from the cars, so that it is impossible to hear anything.’
Selma no cree que su tía la...
 ‘Selma doesn’t think that her aunt her ... ’

Here, the target answer would be *oiga/escuche* (‘hear.3SG.SUB’), or any comparable answer containing a subjunctive verb form.

The task contained 81 items in total: 3 practice items, 9 items in each of the three conditions targeting the subjunctive, 3×9 items in which the indicative was the target answer (which are not discussed in this paper), and 24 filler items targeting a different construction, namely word order. The full task is presented in the Supplementary Materials. The items appeared in randomized order.

After a short break, participants continued with the scalar acceptability judgment task, in which they were presented (again, both aurally and written) with similar stories followed by two sentences that only differed from each other regarding the mood of the verb. Each of the sentences had to be rated on a scale from -2 to 2 . The instruction stated that -2 indicated: “this sentence sounds very unnatural to me and I would never say this sentence”, and 2 indicated: “this sentence sounds completely natural to me and I could say this”. This task was assumed to tap into more explicit knowledge of the subjunctive. First of all, there was no time limit whatsoever; whenever the participants were sure about their answer they could press the space bar to proceed to the next item. Second, since both options—indicative and subjunctive—were always presented simultaneously, the focus was on the form of the verb, and it was assumed that this would moreover automatically make the participants aware of the topic they were being tested on. Most participants confirmed this to be the case during debriefing.

The acceptability judgment task contained the same 3 conditions as the elicited production task, and the same types of stories to target the more felicitous mood. Some of the items for this task were based on items used by [Borgonovo et al. \(2015\)](#) and [Borgonovo and Prévost \(2003\)](#).⁵ An example for this task, from the relative clause condition, is illustrated in (7):

⁵ The items were obtained through personal communication with the authors.

7. *María tiene que dar una presentación sobre Miró para su clase de historia del arte. Quiere dar muchos ejemplos de pinturas de Miró durante la presentación. Va a la biblioteca y le dice a la señora:*

‘Maria has to give a presentation about Miró for her art history class. She wants to give many examples of paintings by Miró during the presentation. She goes to the library and tells the woman:’

*Busco un libro que **tiene** pinturas de Miró.*

‘I’m looking for a book that has. IND paintings by Miró.’

○ ○ ○ ○ ○
-2 -1 0 1 2

*Busco un libro que **tenga** pinturas de Miró.*

○ ○ ○ ○ ○
-2 -1 0 1 2

Half of the target sentences contained regular verbs and the other half irregular verbs, evenly divided over the three conditions. The same goes for present and past tense, except for the relative clause condition, which were all in present tense. In total, the task contained 81 items, just as the elicited production task: 3 practice items, 27 subjunctive-targeting items, 27 indicative-targeting items (which are not discussed in the present paper), and 24 filler items targeting word order. The full task is presented in the Supplementary Materials. The order of the items, as well as the order of the two sentences, was randomized.

7. Results

7.1. Production Task

Of all 1512 responses, 57 were excluded, because (1) there was no (intelligible) response; (2) no verb was included in the response; or (3) the response did not relate to the story in any logical way. The remaining responses were coded as either “subjunctive”, “indicative” or “other”. In total, there were 950 subjunctive (i.e., correct/felicitous) responses, 448 indicative (i.e., incorrect/infelicitous) responses, and 57 ‘other’ responses, which contained future tenses, conditionals or infinitives. These were excluded from the statistical analysis. The results for the production task are depicted in Figure 1.

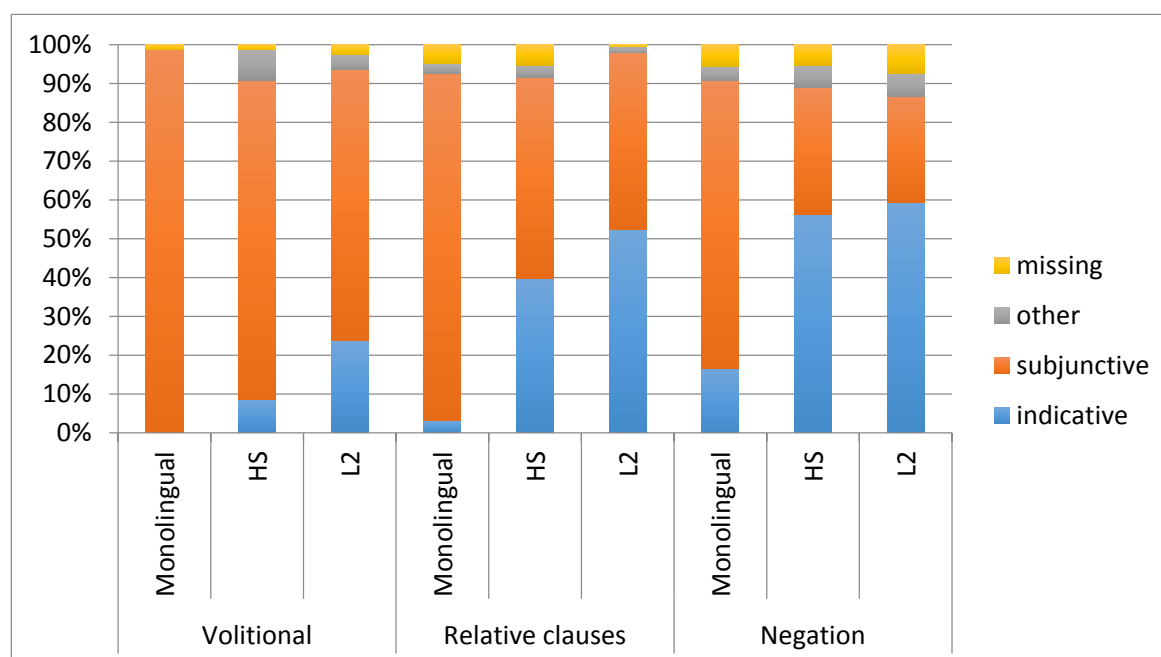


Figure 1. Percentages of indicative (infelicitous), subjunctive (felicitous) and other responses per condition per group. HS: heritage speakers.

To test for quantitative differences between the groups, generalized linear mixed effects models were run in each condition, using the lme4 package from statistical tool R (R Development Core Team 2017). A full overview of all statistical models is presented in the Supplementary Materials. In each analysis, ‘group’ was the independent variable, with one contrast set between the monolingual group vs. the two bilingual groups and another contrast between the heritage speakers and the L2 speakers. In each model, the random structure included intercepts and slopes for ‘item’ and ‘subject’ if this improved the model significantly. For the contrast between the monolingual group and the two bilingual groups, the effect of group was significant in all three conditions (volitional: $\beta = 0.18$, $SE = 0.06$, $z = 3.27$, $p = 0.002$; relative clauses: $\beta = 4.23$, $SE = 0.75$, $z = 5.65$, $p < 0.001$; and negated sentences: $\beta = 3.83$, $SE = 0.83$, $z = 4.62$, $p < 0.001$). For the contrast between the heritage and the L2 speakers, the effect was significant in the volitional condition ($\beta = 0.16$, $SE = 0.06$, $z = 2.53$, $p = 0.016$), but not in the relative clause condition ($\beta = 0.62$, $SE = 0.58$, $z = 1.07$, $p = 0.29$) or the negation condition ($\beta = 0.26$, $SE = 0.81$, $z = 0.32$, $p = 0.75$). This means that in all three conditions, the monolinguals produced the subjunctive relatively more frequently than the two bilingual groups combined and in the volitional condition, the heritage speakers produced the subjunctive significantly more often than the L2 speakers.

7.2. Acceptability Judgment Task

The mean ratings in the acceptability judgment task are depicted in Figure 2.

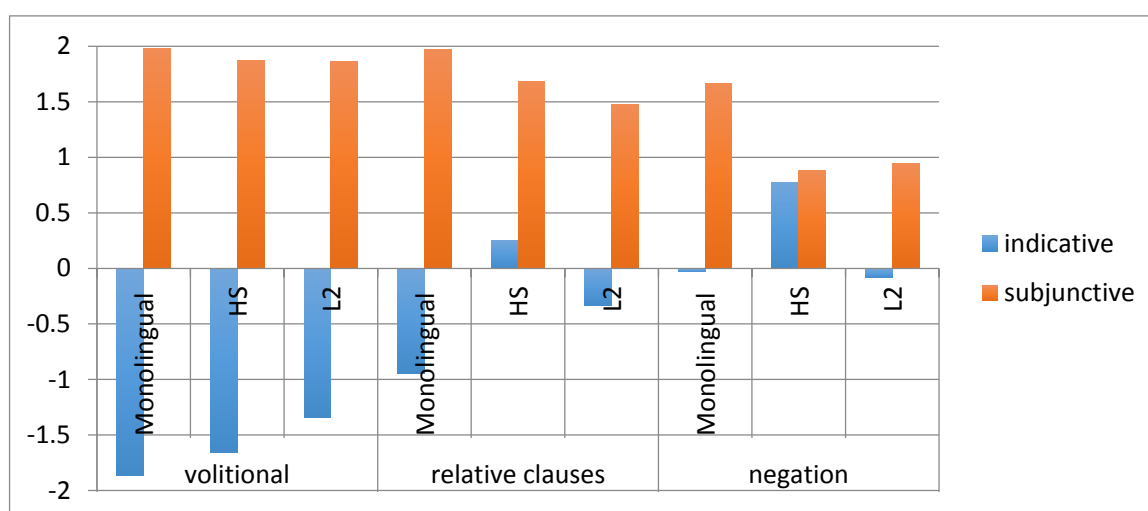


Figure 2. Mean ratings for the subjunctive (felicitous) and indicative (infelicitous) sentences per condition per group. HS: heritage speakers; L2: L2 speakers.

Linear mixed effects analyses were run for the rating for the indicative sentence and the rating for the subjunctive sentence in each condition. Again, ‘group’ was the independent variable with one contrast set between the monolingual group vs. the two bilingual groups and another contrast between the heritage speakers and the L2 speakers. Random intercepts and slopes were included for ‘item’ and ‘subject’, if this improved the model significantly. p -Values were calculated using the Kenward–Roger approximation, from the pbkrtest package (Halekoh and Højsgaard 2014).

In the analyses of the ratings for the subjunctive (felicitous) sentence, an effect of group was found in all three conditions, which was marginal in the volitional condition ($\beta = -0.11$, $SE = 0.06$, $t = -1.93$, $p = 0.058$), and significant in the relative clause ($\beta = -0.39$, $SE = 0.14$, $t = -2.80$, $p = 0.007$) and the negation condition ($\beta = -0.75$, $SE = 0.23$, $t = -3.26$, $p = 0.003$), but only for the contrast between monolingual and bilingual speakers. This means that monolingual speakers rated the

subjunctive higher than both bilingual groups, but the two bilingual groups did not rate the subjunctive significantly differently from each other.

In the analyses of the ratings for the indicative sentences, there were also significant effects of group in all three conditions. For the contrast between monolingual and bilingual speakers the effect was significant in the volitional condition ($\beta = 0.36$, $SE = 0.17$, $t = 2.10$, $p = 0.04$) and in the relative clause condition ($\beta = 0.91$, $SE = 0.26$, $t = 3.52$, $p < 0.001$). This means that the monolingual controls rejected the indicative more than the two bilingual groups in the volitional and the relative clause condition. For the contrast between the two bilingual groups, the effect was significant in the negation condition ($\beta = -0.86$, $SE = 0.34$, $t = -2.57$, $p = 0.014$), indicating that in this condition, the L2 speakers rejected the indicative more than the heritage speakers. In the relative clause condition, there was a similar tendency, which was not significant ($\beta = -0.59$, $SE = 0.33$, $t = -1.76$, $p = 0.08$).

Summing up these results for the judgment task, monolingual speakers show a higher preference for subjunctive than the two bilingual groups, across conditions. L2 speakers are more inclined to reject infelicitous indicative than heritage speakers in the negation condition (and approaching significance in the relative clause condition). Combining the results for the two tasks, we can conclude that there is indeed a task effect: in production, heritage speakers outperform L2 speakers, and the opposite is true in the judgment task.

7.3. Individual Results

Individual results were also explored. For the judgment task, this meant that for each participant in each condition, it was checked whether their average rating for the subjunctive sentences was at least 0.5 points higher than their average rating for the indicative sentences. This cut-off point of 0.5 was chosen because it was the smallest difference between ratings occurring in the monolingual group. For the production task, it was checked for each individual participant in each condition whether they used at least one more subjunctive than indicative. This cut-off point was not based on a monolingual standard; in fact, even in the monolingual group there were two participants who actually used two more indicatives than subjunctives (only in the negation condition). The argument for choosing a difference of 1 (and not, say, two, or three) was that a difference of 1 out of 9 items corresponds roughly to a difference of 0.5 on a scale from -2 to 2 , which was the cut-off point in the judgment task.

For the production task, the individual results showed that in the volitional condition, all monolinguals and heritage speakers used the subjunctive more than the indicative, but only 15 out of 21 L2 speakers did so. In the relative clause condition, all monolinguals, 8 out of 17 heritage speakers and 9 out of 21 L2 speakers used subjunctive more frequently than the indicative. In the negation condition, 16 out of 18 monolinguals, 5 out of 17 heritage speakers and 4 out of 21 L2 speakers showed a similar pattern. As for judgment, all participants, except for one L2 speaker, rated subjunctive higher than the indicative in the volitional condition. In the relative clause condition, all monolinguals, 14 out of 17 heritage speakers and 17 out of 21 L2 speakers showed a similar preference. In the negation condition, all monolinguals preferred the subjunctive, but 12 out of 21 L2 speakers and only 4 out of 17 heritage speakers did the same. These results are summarized in Tables 2 and 3.

Table 2. Number of participants per group who produced subjunctive more than indicative.

	Volitional	Relative Clauses	Negation
Monolinguals	18/18 (100%)	18/18 (100%)	16/18 (88.89%)
Heritage speakers	17/17 (100%)	8/17 (47.1%)	5/17 (29.41%)
L2 speakers	15/21 (71.4%)	9/21 (42.9%)	4/21 (19.05%)

Table 3. Number of participants per group who rated subjunctive higher than indicative.

	Volitional	Relative Clauses	Negation
Monolinguals	18/18 (100%)	18/18 (100%)	18/18 (100%)
Heritage speakers	17/17 (100%)	14/17 (82.35%)	4/17 (23.53%)
L2 speakers	20/21 (95.24%)	17/21 (80.95%)	12/21 (57.14%)

These patterns thus mirror the group results: an advantage for the heritage speakers in the production task, which is most pronounced in the volitional condition, and a clear advantage for the L2 speakers in judgment in the negation condition.

7.4. The Role of Exposure and Instruction

As mentioned in Section 6.1, the amount of instruction and exposure to naturalistic input varied greatly within the L2 group. To explore the role of these two factors, additional analyses were run. For each participant, the number of months of instruction was calculated based on the information provided by the background questionnaire. Similarly, cumulative exposure was calculated by adding up the total amount of time spent in Spanish-speaking countries (including holidays, longer trips, and/or living abroad). With these two variables, another set of mixed effects model was run on the L2 group in all contexts in both judgment and production. In neither task, and in neither context did any of these effects turn out significant.

Another source of considerable variation within the L2 group was the amount of current exposure and use of Spanish, for instance because some of them had a Spanish-speaking partner, or had to use Spanish for their work. The questions in the questionnaire regarding current use and exposure of Spanish and Dutch were formulated in terms of relative frequency. Therefore, this information could not be expressed in a numerical variable and thus could not be included in a statistical analysis. Instead, the individual results of 7 participants who reported being exposed to Spanish equally or more often than Dutch at home or in the workplace were investigated with more scrutiny. However, no particular different behaviour could be deduced for these seven participants: in fact, most of them performed quite poorly on the production task, which would not be expected based on their increased exposure and use of Spanish.

These additional analyses thus provided no evidence for any effect of amount of explicit instruction and/or exposure to naturalistic input on L2 speakers' relative preference for the subjunctive in either task.

8. Discussion

This study tested heritage, L2 and monolingual speakers' implicit and explicit knowledge of the subjunctive in various contexts. Based on previous research comparing heritage and L2 speakers, it was hypothesized that heritage speakers would outperform L2 speakers in the implicit task and vice versa in the explicit task.

The design of our study was innovative in two ways. First of all, we closely matched the two groups' general proficiency levels in Spanish based on both a written explicit proficiency task (the DELE) and an aural implicit proficiency task (a lexical decision task). Including both these tasks as matching criteria reduced the risk of an overestimation of the proficiency of one of the groups due to the language mode they are more familiar with (written for L2 speakers and spoken for heritage speakers) or the type of knowledge they are assumed to possess (explicit for L2 speakers and implicit for heritage speakers). Moreover, unlike other studies, all stimuli were presented simultaneously

aurally and written. This design allowed us to reduce the possibility of a confounding effect of the language mode of presentation (spoken vs. written).⁶

The results showed first of all that the two bilingual groups diverged from the monolingual group in all three conditions: they showed a weaker preference for the subjunctive in both judgment and production. As for differences between the two bilingual groups, task-based differences were indeed attested in two out of three conditions. In the volitional condition, the heritage speakers outperformed the L2 speakers in production, but not in judgment.⁷ In the negation condition, L2 speakers outperformed heritage speakers in the judgment task, but not in the production task. These results were confirmed by individual analyses. One might wonder about the interaction between task advantage and subjunctive context.⁸ Could it be that heritage speakers are better in production in volitional contexts, because these are more frequent in everyday use, and L2 speakers are better with negation contexts in judgment because these have been part of their instruction and are less frequent in everyday use? As for the second point, as discussed in Section 2, the negation context is rarely extended to other verbs than epistemic verbs in textbooks. Any knowledge that L2 speakers have acquired about the subjunctive in the negation context thus cannot reflect mere repetition of what was offered during instruction. We think that L2 speakers' more explicit and metalinguistic knowledge enables them to recognize the pattern that subjunctive is used to refer to non-assertion (from the perspective of the speaker) and extend it to other verbs. Moreover, even if there is an effect of frequency of occurrence of the specific subjunctive contexts in specific situations, we do not think that this invalidates the argument for differences between the two groups in terms of the type of knowledge they possess. Even if the heritage speakers' advantage with volitional subjunctive in the production task would be due to the fact that this construction is more frequent in spoken language (with which they are more familiar) than in explicit instruction, the fact that they are able to apply this knowledge better than L2 speakers in one task, but not in the other, still implies that the nature of this knowledge is different between the groups. Similarly, even if the L2 speakers were helped because there is more evidence for the negation construction in instructed input (with which L2 speakers are more familiar) than in everyday language use, the fact that they are able to use this advantage only in the judgment task, but not in production, still implies that they have relatively more explicit knowledge of the construction.

These findings thus confirm previously attested task-based differences between heritage and L2 speakers, which until now have extensively been attested for morpho-syntactic phenomena (Montrul 2012), but not for other linguistic domains. Our study suggests that the observation can be extended to the interface between syntax and pragmatics, although more studies targeting different kinds of interface phenomena are needed to be able to generalize our findings to the domain as a whole.

The fact that these task-based differences between the groups were attested even though language mode of presentation (oral vs. written) was controlled for provides even more solid evidence suggesting that these two populations indeed possess different types of knowledge of the subjunctive. Nevertheless, it is important to point out that despite the equal *presentation* (written and aurally) of the items in the present study, our two tasks differed from one another in that only in the implicit task oral production was required. Even though Ellis (2005, 2009b) did not find evidence for a distinction between production vs. judgment in his task battery, it would be interesting to see whether the task effect remains if oral production is taken out of the equation. We would like to suggest future

⁶ An anonymous reviewer pointed out that, to completely control for language mode, one would ideally present two separate versions of the tasks: one aural and one written version. However, this would have required doubling the number of participants. A bimodal presentation therefore seemed the best solution to the effect of language mode.

⁷ Interestingly, the morphological recognition task, which was also explicit in nature, and which targeted partially similar contexts as the volitional condition, rendered a significantly higher accuracy on part of the L2 speakers compared to the heritage speakers. However, one of the differences between this task and the judgment task was that the MRT was presented only in written form, not aurally, which may have contributed to the L2 advantage.

⁸ Thanks to an anonymous reviewer for pointing this out to us.

researchers to include more online measures to tap into implicit knowledge, such as a self-paced reading/listening task, which does not include any oral production, but still meets the requirements of an implicit task as proposed by Ellis (2005, 2009b), namely: lack of awareness, time pressure and focus on meaning. We moreover agree with Montrul et al. (2008b) that including neuroimaging techniques could be a promising line of research to provide more insight in this matter.

So what does it mean if heritage speakers' knowledge is more implicit in nature and L2 speakers' knowledge is more explicit? As suggested in the introduction, this can mean two things. It could be related to the age at which each group was first exposed to the language. Certain theories assuming a critical period for language acquisition claim that the fundamental difference between early and late acquisition of a language is that while children are able to acquire language implicitly, based on exposure to naturalistic input only, this ability is lost in adults, who therefore can only resort to explicit learning mechanisms (Bley-Vroman 1990; DeKeyser 2000).

But heritage and L2 learners also differ with respect to "manner" of acquisition: while heritage speakers have been predominantly exposed to naturalistic input, and do not receive much explicit instruction in the language, for most L2 speakers it is exactly the other way around. It seems obvious that explicit instruction will lead to more explicit knowledge and naturalistic exposure leads to implicit knowledge. This may be true regardless of age of acquisition. The problem is that with instructed L2 learners, these two possible explanations are generally impossible to tease apart, since they differ from heritage speakers regarding both age and manner of acquisition.

Nevertheless, the data from the present study may shed some new light on this question. As mentioned in Section 6.2, the amount of instruction and exposure to naturalistic input varied greatly within the L2 group. If experience with the language is part of the explanation for the attested task-based effects, we would expect those L2 learners who received the most explicit instruction to do better on the explicit task, and those who were most exposed to naturalistic input to behave more monolingual-like on the implicit task. However, separate statistical and individual analyses did not reveal any effect of the amount of exposure to naturalistic input on L2 speaker's performance in production, or for an effect of the amount of explicit instruction on their performance on an explicit judgment task. We would therefore like to argue that the observed task-based differences are more likely to be related to age of onset of acquisition than to manner of acquisition. Nevertheless, to unambiguously disentangle these two effects, studies should compare heritage speakers to naturalistic L2 learners who have not received any instruction, and who, moreover, have received a similar amount of input as heritage speakers (and are still comparable in terms of general proficiency). This way, the factor "manner" of acquisition can be completely taken out of the equation. We leave these questions for future research.

9. Conclusions

This study compared proficiency-matched heritage and L2 speakers of Spanish on their implicit and explicit knowledge of the subjunctive in Spanish: a multiple interface phenomenon. An acceptability judgment task measured the participant's explicit knowledge on the topic, and an elicited oral production task tapped into implicit knowledge. The results confirmed task-based differences between heritage and L2 speakers of Spanish for the subjunctive in the volitional subjunctive (morpho-syntactic) condition and the negation (syntax-pragmatics interface) condition: while heritage speakers did better on the implicit task, L2 speakers had the advantage in the explicit task. This suggests that differential task effects for these two bilingual groups can be found not only in morpho-syntax as reported in e.g., Montrul (2012), but also in the syntax-pragmatics interface. We also demonstrated that these effects could not be attributed to a confounding effect of the mode in which the tests were administered (aural vs. written), or to the way in which general proficiency in Spanish was measured.

The findings from this study underline the importance of using different task types in bilingual research: relying exclusively on one task may obscure underlying differences between different bilingual populations. Scholars should be aware that different types of tasks tap into fundamentally

different types of knowledge, and depending on one's research question and on the type of population, an explicit task or an implicit task (or both!) will be more suitable. These findings moreover have potential implications for theories about the relationship between age of onset of acquisition and nature of knowledge (implicit vs. explicit).

Supplementary Materials: The following are available online at <http://www.mdpi.com/2226-471X/3/3/25/s1>, Supplementary 1: EPT Items; Supplementary 2: AJT Items; Supplementary 3: Statistical Models.

Author Contributions: B.v.O. conceived and designed the experiment, performed the experiments, analyzed the data and wrote the paper under the supervision of A.H., P.S. and S.A.

Funding: The research presented in this study was funded by a grant from the graduate program of the Netherlands Graduate School of Linguistics (LOT), which received the funds from The Netherlands Organisation for Scientific Research (NOW) in the context of the project “Language—from cognition to communication” (NOW project number 022.004.015).

Acknowledgments: We would like to express our appreciation to Dirk Jan Vet and Nelson Ogliastri for their assistance in the technical design of the experiment, as well as Paul Boersma for his assistance with the statistical analysis. Much gratitude also goes to all the people who participated in this study. Finally, we want to thank the two anonymous reviewers for their useful comments on an earlier version of this paper. Any weaknesses and errors are entirely ours.

Conflicts of Interest: The authors declare no conflict of interest. The funding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

- Au, Terry Kit-fong, Leah M. Knightly, Sun-Ah Jun, and Janet S. Oh. 2002. Overhearing a language during childhood. *Psychology Science* 13: 238–43.
- Benmamoun, Elabbas, Silvina Montrul, and Maria Polinsky. 2010. *Prolegomena to heritage linguistics*. White paper. Champaign, IL, USA: University of Illinois at Urbana–Champaign; Cambridge, MA, USA: Harvard University.
- Bialystok, Ellen, and Kenji Hakuta. 1999. Confounded age: linguistic and cognitive factors in age differences for second language acquisition. In *Second language Acquisition and the Critical Period Hypothesis*. Edited by David P. Birdsong. Mahwah: Lawrence Erlbaum, pp. 161–81.
- Bley-Vroman, Robert. 1990. The logical problem of foreign language learning. *Linguistic Analysis* 20: 3–49.
- Borgonovo, Claudia, and Philippe Prévost. 2003. Knowledge of polarity subjunctive in L2 Spanish. In *BUCLD 27: Proceedings of the 27th annual Boston University Conference on Language Development*. Edited by Barbara Beachley, Amanda Brown and Frances Conlin. Somerville: Cascadilla Press, vol. 1, pp. 150–61.
- Borgonovo, Claudia, Joyce Bruhn de Garavito, and Philippe Prévost. 2015. Mood selection in relative clauses. *Studies in Second Language Acquisition* 37: 33–69. [[CrossRef](#)]
- Bowles, Melissa A. 2011. Measuring implicit and explicit linguistic knowledge. *Studies in Second Language Acquisition* 33: 247–71. [[CrossRef](#)]
- Bruhn de Garavito, Joyce. 2002. Verb raising in Spanish: A comparison of early and late bilinguals. In *BUCLD 26: Proceedings of the 26th annual Boston University Conference on Language Development*. Edited by Barbora Skarabela, Sarah Fish and Anna H.-J. Do. Somerville: Cascadilla Press, vol. 1, pp. 84–94.
- Chang, Charles B., Erin F. Haynes, Yao Yao, and Russell Rhodes. 2008. A tale of two fricatives: Consonantal contrast in heritage speakers of Mandarin. *University of Pennsylvania Working Papers in Linguistics* 15: 37–43.
- DeKeyser, Robert M. 2000. The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition* 22: 499–533.
- DeKeyser, Robert M. 2003. Implicit and explicit learning. In *The Handbook of Second Language Acquisition*. Edited by Catherine J. Doughty and Michael H. Long. Oxford: Blackwell, pp. 312–48.
- Ellis, Rod. 2005. Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition* 27: 141–72. [[CrossRef](#)]
- Ellis, Rod. 2009a. *Implicit and Explicit Learning, Knowledge and Instruction. Implicit and Explicit Knowledge in Second Language Learning, Testing and Teaching*. Bristol: Multilingual Matters, vol. 42, pp. 3–25.
- Ellis, Rod. 2009b. Investigating learning difficulty in terms of implicit and explicit knowledge. In *Implicit and Explicit Knowledge in Second Language Learning, Testing and Teaching*. Edited by Rod Ellis, Shawn Loewen, Catherine Elder, Hayo Reinders, Rosemary Erlam and Jenefer Philp. Bristol: Multilingual Matters, pp. 31–64.

- Flege, James E. 1999. Age of learning and Second-language speech. In *Second Language Acquisition and the Critical Period Hypothesis*. Edited by David P. Birdsong. Hillsdale: Lawrence Erlbaum, pp. 101–32.
- Flynn, Suzanne, and Sharon Manuel. 1991. Age-dependent effects in language acquisition: An evaluation of ‘critical period’ hypotheses. In *Point/Counterpoint: Universal Grammar in the Second Language*. Edited by Lynn Eubank. Amsterdam: John Benjamins, pp. 351–74.
- Håkansson, Gisela. 1995. Syntax and morphology in language attrition: A study of five bilingual expatriate Swedes. *International Journal of Applied Linguistics* 5: 153–71. [\[CrossRef\]](#)
- Halekoh, Ulrich, and Søren Højsgaard. 2014. A kenward-roger approximation and parametric bootstrap methods for tests in linear mixed models—the R package pbrtest. *Journal of Statistical Software* 59: 1–30. [\[CrossRef\]](#)
- Iverson, Michael, Paula Kempchinsky, and Jason Rothman. 2008. Interface vulnerability and knowledge of the subjunctive/indicative distinction with negated epistemic predicates in L2 Spanish. In *Eurosla Yearbook 8*. Edited by Leah Roberts, Florence Myles and Annabelle David. Amsterdam: John Benjamins, pp. 135–63.
- Keating, Gregory D., Bill Van Patten, and Jill Jegerski. 2011. Who was walking on the beach? Anaphora Resolution in Spanish Heritage Speakers and Adult Second Language Learners. *Studies in Second Language Acquisition* 33: 193–221. [\[CrossRef\]](#)
- Knightly, Leah M., Sun-Ah Jun, Janet S. Oh, and Terry Kit-fong Au. 2003. Production benefits of childhood overhearing. *Journal of the Acoustic Society of America* 114: 465–74. [\[CrossRef\]](#)
- Krashen, Stephen D. 1982. Accounting for child-adult differences in second language rate and attainment. In *Child-Adult Differences in Second Language Acquisition*. Edited by Stephen D. Krashen, Robin C. Scarcella and Michael H. Long. Rowley: Newbury House, pp. 202–26.
- Lukyanchenko, A., and K. Gor. 2011. Perceptual correlates of phonological representations in heritage speakers and L2 learners. In *BUCLD 35: Proceedings of the 35th annual Boston University Conference on Language Development*. Edited by Nick Danis, Kate Mesh and Hyunsuk Sung. Somerville: Cascadia Press, vol. 2, pp. 414–26.
- Lynch, Andrew. 2008. The linguistic similarities of Spanish heritage and second language learners. *Foreign Language Annals* 41: 252–381. [\[CrossRef\]](#)
- Mikulski, Ariana Maria. 2006. Native Intuitions, Foreign Struggles? Knowledge of the Subjunctive in Volitional Con-structions among Heritage and Traditional FL Learners of Spanish. Ph.D. dissertation, University of Iowa, Iowa City, IW, USA, May.
- Mikulski, Ariana Maria. 2010. Receptive volitional subjunctive abilities in heritage and traditional foreign language learners of Spanish. *The Modern Language Journal* 94: 217–33. [\[CrossRef\]](#)
- Mikulski, Ariana Maria, and Idoia Elola. 2013. Heritage and Foreign Language Learner Use of the Spanish Subjunctive in Advice. *Heritage Language Journal* 10: 51–82.
- Montrul, Silvina. 2004. Convergent outcomes in second language acquisition and first language loss. In *First Language Attrition: Interdisciplinary Perspectives on Methodological Issues*. Edited by Monika S. Schmid, Barbara Köpke, Merel Keijzer and Lina Weilemar. Amsterdam: John Benjamins, pp. 259–80.
- Montrul, Silvina. 2005. Second language acquisition and first language loss in adult early bilinguals: Exploring some differences and similarities. *Second Language Research* 21: 199–249. [\[CrossRef\]](#)
- Montrul, Silvina. 2006. Incomplete acquisition as a feature of bilingual and L2 grammars. In *Inquiries in Linguistic Development: In honor of Lydia White*. Edited by Roumyana Slabakova, Silvina Montrul and Philippe Prévost. Amsterdam: John Benjamins Publishing Company, pp. 335–59.
- Montrul, Silvina. 2008. Incomplete acquisition in Spanish heritage speakers: Chronological age or interfaces vulnerability. In *BUCLD 32: Proceedings of the 32nd annual Boston University Conference on Language Development*. Edited by Harvey Chan, Heather Jacob and Enkeleida Kapia. Somerville: Cascadia Press, vol. 1, pp. 414–26.
- Montrul, Silvina. 2009. Knowledge of tense-aspect and mood in Spanish heritage speakers. *International Journal of Bilingualism* 13: 239–269. [\[CrossRef\]](#)
- Montrul, Silvina. 2010. How similar are L2 learners and heritage speakers? Spanish clitics and word order. *Applied Psycholinguistics* 31: 167–207. [\[CrossRef\]](#)
- Montrul, Silvina. 2011. Morphological errors in Spanish second language learners and heritage speakers. *Studies in Second Language Acquisition* 33: 163–92. [\[CrossRef\]](#)

- Montrul, Silvina. 2012. Is the heritage language like a second language? In *Eurosla Yearbook 12*. Edited by Leah Roberts, Christina Lindqvist, Camilla Bardel and Niclas Abrahamsson. Amsterdam: John Benjamins, pp. 1–29.
- Montrul, Silvina, and Tania Ionin. 2012. Dominant language transfer in Spanish heritage speakers and second language learners in the interpretation of definite articles. *The Modern Language Journal* 96: 70–94. [\[CrossRef\]](#)
- Montrul, Silvina, and Silvia Perpiñán. 2011. Assessing differences and similarities between instructed heritage language learners and L2 learners in their knowledge of Spanish tense-aspect and mood (TAM) morphology. *Heritage Language Journal* 8: 90–133.
- Montrul, Silvina, Rebecca Foote, Silvia Perpiñán, Dan Thornhill, and Susana Vidal. 2006. Full access and age effects in adult bilingualism: An investigation of Spanish accusative clitics and word order. In *Selected Proceedings of the 9th Hispanic Linguistics Symposium*. Edited by Nuria Sagarra and Almeida Jacqueline Toribio. Somerville: Cascadia Press, pp. 217–28.
- Montrul, Silvina, Rebecca Foote, and Silvia Perpiñán. 2008a. Knowledge of wh-movement in Spanish L2 learners and heritage speakers. In *Selected Papers of the 8th Hispanic Linguistics Symposium*. Edited by Joyce Bruhn de Garavito and Elena Valenzuela. Somerville: Cascadia Press, pp. 93–106.
- Montrul, Silvina, Rebecca Foote, and Silvia Perpiñán. 2008b. 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, Justin Davidson, Israel De La Fuente, and Rebecca Foote. 2014. Early language experience facilitates the processing of gender agreement in Spanish heritage speakers. *Bilingualism: Language and Cognition* 17: 118–38. [\[CrossRef\]](#)
- Oh, Janet S., Sun-Ah Jun, Leah M. Knightly, and Terry Kit-fong Au. 2003. Holding on to childhood language memory. *Cognition* 86: B53–B64. [\[CrossRef\]](#)
- Paradis, Michel. 1994. Neurolinguistic aspects of implicit and explicit memory: implication for bilingualism and second language acquisition. In *Implicit and Explicit Learning of Languages*. Edited by Nick C. Ellis. Cambridge: Academic Press, pp. 393–421.
- Potowski, Kim, Jill Jegerski, and Kara Morgan-Short. 2009. The effects of instruction on linguistic development in Spanish heritage language speakers. *Language Learning* 59: 537–79. [\[CrossRef\]](#)
- Quer, Josep. 2001. Interpreting mood. *Probus* 13: 81–111. [\[CrossRef\]](#)
- R Development Core Team. 2017. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Shanks, David R. 2003. Attention and awareness in “implicit” sequence learning. In *Attention and Implicit Learning*. Edited by Luis Jiménez. Amsterdam: John Benjamins, vol. 48, pp. 11–42.
- Silva-Corvalán, Carmen. 1994. The gradual loss of mood distinctions in Los Angeles Spanish. *Language Variation and Change* 6: 255–72. [\[CrossRef\]](#)
- Thieroff, Rolf. 2004. The subjunctive mood in German and the Germanic languages. In *Focus on Germanic Typology*. Edited by Werner Abraham. Berlin: Akademie Verlag, pp. 315–58.
- Valdés, Guadalupe. 1995. The teaching of minority languages as academic subjects: Pedagogical and theoretical challenges. *The Modern Language Journal* 79: 299–328. [\[CrossRef\]](#)
- Van Osch, Brechje, and Petra Sleeman. 2016. Spanish heritage speakers in the Netherlands: Linguistic patterns in the judgment and production of mood. *International Journal of Bilingualism*. [\[CrossRef\]](#)
- Van Osch, Brechje, Aafke Hulk, Petra Sleeman, and Suzanne Aalberse. 2017. Knowledge of mood in internal and external interface contexts in Spanish heritage speakers in The Netherlands. In *Multidisciplinary Approaches to Bilingualism in the Hispanic and Lusophone World*. Edited by Kate Bellamy, Michael W. Child, Paz González, Antje Muntendam and M. Carmen Parafita Couto. Amsterdam: John Benjamins, pp. 67–94.

