Article

The *That*-Trace Effect: Evidence from Spanish–English Code-Switching

Bradley Hoot [1,*] **and Shane Ebert** [2]

1   Department of Modern Languages, DePaul University, Chicago, IL 60614, USA
2   Department of Hispanic and Italian Studies, University of Illinois Chicago, Chicago, IL 60607, USA;
    sebert2@depaul.edu
*   Correspondence: bhoot@depaul.edu

Abstract: The *that*-trace effect is the fact that many languages (like English) ban the extraction of embedded-clause *subjects* but not *objects* over an overt complementizer like *that*, while many other languages (like Spanish) allow such extractions. The effect and its cross-linguistic variation have been the subject of intense research but remain largely a mystery, with no clear consensus on their underpinnings. We contribute novel evidence to these debates by using Spanish–English code-switching (the use of two languages in one sentence) to test five contemporary theoretical accounts of the *that*-trace effect. We conducted a formal acceptability judgment experiment, manipulating the extracted argument and code-switch site to test different combinations of linguistic features. We found that subject extraction is only permitted in Spanish–English code-switching when both the C head (*que* 'that') and the T head (i.e., the verb) are in Spanish, but not when either functional head is in English. Our results demonstrate indirect support for two of the five theories we test, failing to support the other three. Our findings also provide new evidence in favor of the view that the *that*-trace effect is tightly linked to the availability of post-verbal subjects. Finally, we outline how our results can narrow the range of possible theoretical accounts, demonstrating how code-switching data can contribute to core questions in linguistic theory.

Keywords: *that*-trace effect; code-switching; anti-locality; Criterial Freezing; syntax/prosody interface; labeling; COMP-trace; experimental syntax; acceptability judgment task

1. Background

This paper uses novel evidence from code-switching to explore a classic phenomenon in linguistic theory: the *that*-trace effect, exemplified in (1).

1.   English: the *that*-trace effect

    a.   Who do you think <u>that</u> Susana saw t_{who}?
    b.   * Who do you think <u>that</u> t_{who} saw Susana?
    c.   Who do you think ___ Susana saw t_{who}?
    d.   Who do you think ___ t_{who} saw Susana?

Extracting a direct object from an embedded clause headed by the complementizer *that* is licit (1a), but extracting a subject is barred (1b). Both extractions are allowed when the complementizer is omitted (1c,d). These facts remain essentially a mystery: despite extensive research over the decades, their cause is still debated (Pesetsky 2017). The effect continues to garner interest because it touches on core issues of the architecture of language, including limits on a core syntactic operation—movement—and the interaction between prosody and syntax. The wide range of approaches to accounting for the *that*-trace effect complicates the process of converging on an explanation, so novel evidence which narrows the range of possible accounts can help push the field toward the most fruitful avenues of research.

We draw such evidence from code-switching, the use of more than one language in a single utterance. Code-switching is not a random admixture of material from each language; it is rule governed, like all natural languages. Because code-switching sentences are constrained by unconscious mental rules just like any other sentences, they can be used in the same way as monolingual evidence to evaluate theoretical claims and contribute to understanding the language faculty. Yet code-switching also offers a special advantage: code-switching combines not only the two languages' words but also their linguistic features, potentially producing combinations of features not available in monolingual sentences in either language. Code-switching thus functions as a microscope of sorts, allowing researchers to make new discoveries by combining dissimilar languages.

To that end, we examine English combined with Spanish, a language that does not display the *that*-trace effect. Spanish differs from English in several relevant ways, and we can use these cross-linguistic differences to profitably combine features in Spanish–English code-switching to put theoretical claims about the *that*-trace effect to the test.

To situate our work, we first describe some key facts about the *that*-trace effect, English and Spanish (Section 1.1) and how code-switching can shed light on these topics (Section 1.2). We then review five theories of the *that*-trace effect we can test (Section 1.3) before presenting the research questions (Section 1.4), followed by the methods (Section 2), results (Section 3), and discussion (Section 4). Previewing our results, we find that subjects can only be extracted over overt complementizers in Spanish–English code-switching when both C and T are in Spanish. We argue that our results show further support for the view that the *that*-trace effect is tightly tied to extraction from a particular syntactic position. We find indirect support for two of the tested theoretical proposals, and we suggest how future hypotheses might be constrained by our data.

1.1. The That-Trace Effect in English and Not in Spanish

The *that*-trace effect in English, as shown above in (1), has two components: (i) a subject/object asymmetry (only subjects are barred from moving over the complementizer) and (ii) a complementizer asymmetry (deleting the complementizer obviates the effect). Spanish differs on both counts. First, Spanish subjects and objects can both be extracted over the complementizer *que* (2) (Perlmutter 1971; Pesetsky and Torrego 2001).

2. Spanish: no subject/object asymmetry

   a.

¿A	quién	crees	que	Susana	vio	$t_{\text{a quién}}$?
ACC	who	believe.2SG	that	Susana	saw.3SG	

       'Who do you think Susana saw?'

   b.

¿Quién	crees	que	$t_{\text{quién}}$ vio	a	Susana?
who	believe.2SG	that	saw.3SG	ACC	Susana

       'Who do you think saw Susana?'

Not only can Spanish subjects be extracted over *que*, but many of their properties differ from English. Most salient for present purposes is the fact that Spanish allows subjects to appear in post-verbal position, whereas English generally requires pre-verbal subjects; see (3). For historical reasons, the property of requiring pre-verbal subjects is called the EPP, and we assume that it entails movement to a particular syntactic position, which we will refer to as Spec-TP (specifier of the Tense Phrase, which hosts pre-verbal subjects).[1]

3. Cross-linguistic differences in subject position

a.	The	boy	cried.	/	*Cried	the	boy.
b.	El	niño	lloró.	/	Lloró	el	niño.
	The	boy	cried.3SG		cried.3SG	the	boy

       'The boy cried.'

Across many languages, post-verbal subjects correlate with lacking the *that*-trace effect (Gilligan 1987; Pesetsky 2017).[2] Why should post-verbal subjects avoid the *that*-trace effect? Rizzi (1982) provides an explanation, demonstrating convincingly using sentences with

partitive clitics that when subjects in Italian are extracted from an embedded clause, they are extracted from their post-verbal position, not their pre-verbal position. Since then, more evidence has emerged from diverse languages—such as Trentino (Brandi and Cordin 1989), Bani-Hassan Arabic (Kenstowicz 1989), and Icelandic and Kinande (Bošković 2016)—that extraction of the subject over an overt complementizer occurs from a position other than Spec-TP. This evidence has led many researchers to concur that subject position and the *that*-trace effect are intertwined: if it is possible for a subject to appear post-verbally, it's possible to extract it. Thus, the data in (2b) might be more correctly represented as (4), showing that, unlike the situation in English, the extracted subject has moved not from Spec-TP but from a post-verbal position.

4.

¿Quién	crees	que	vio	$t_{quién}$	a	Susana?
who	believe.2SG	that	saw.3SG		ACC	Susana

'Who do you think saw Susana?'

Under this view, the *that*-trace effect is not a prohibition on extracting subjects *per se*, but rather a prohibition on extraction from a particular syntactic position, namely Spec-TP. This raises two questions that are central to any theory of the effect. First, what is special about Spec-TP? In other words, why does this specific position block extraction? Second, is the ban on extraction from Spec-TP universal? That is, is extraction from that position always barred, such that what differs across languages is whether the subject is required to raise there? In other words, does cross-linguistic variation in the *that*-trace effect boil down to variation of the EPP?

This latter question is relevant because subject position is not the only way Spanish and English differ. The English subject/object asymmetry disappears when the complementizer does—when *that* is omitted, both subjects and objects can be extracted; see (1c) and (1d)—whereas Spanish can never omit the complementizer—*que* is obligatory; compare (2) to (5).

5. Spanish: obligatory *que*

a.

*	¿A	quién	crees	___	Susana	vio	$t_{a\ quién}$?
	ACC	who	believe.2SG		Susana	saw.3SG	

'Who do you think Susana saw?'

b.

*	¿Quién	crees	___	$t_{quién}$	vio	a	Susana?
	who	believe.2SG			saw.3SG	ACC	Susana?

'Who do you think saw Susana?'

This difference has received less attention in the literature, but the properties of the complementizer in each language must also be considered in any satisfactory explanation of the *that*-trace effect and its cross-linguistic variation (and, indeed, we will see that the complementizer appears to play a role in our data).

Both main contrasts between English and Spanish hint at something deeper underlying the cross-linguistic differences. Spanish has post-verbal subjects, as well as null subjects, V-to-T movement, 'rich' subject/verb agreement, and other properties that have been traditionally connected to what was known at the time as the Null Subject Parameter (Barbosa 2011; Camacho 2013). Although a strict parameterization of these characteristics has mostly been abandoned, it remains the case that a cluster of features centers on the EPP property that Spanish lacks, and most analyses of these properties connect them in some way to the verb and its extended projection, especially the functional head T. At the same time, Spanish displays a number of properties that suggest that the features of the complementizer are key in the difference between the languages: obligatory *que* (whereas English C can be either phonologically null or just omitted, depending on the analysis), no T-to-C movement (whereas English has T-to-C movement in matrix questions), and extraction from *whether*-islands (which English forbids in some cases), among others.

As we previously mentioned, the nature of the *that*-trace effect continues to be debated, but its description is largely agreed upon. Despite some initial results questioning the *that*-trace effect in some varieties of English (Sobin 1987), a recent survey of the empirical research (Cowart and McDaniel 2021) provides convincing evidence of its robustness.

Similarly, the description of Spanish we have presented enjoys nearly universal consensus, although one study (Chacón et al. 2015) presents a wrinkle that complicates the empirical picture slightly, which we discuss in Section 3.3. In this landscape, scholars mainly debate what the facts mean, but new evidence from English or Spanish alone is unlikely to settle those debates. Evidence from code-switching is especially valuable because it can narrow the hypothesis space.

1.2. Code-Switching

Code-switching is the use of two or more languages in a conversation (although the term refers to other practices, such as style or dialect shifting, in other fields). We are concerned with intra-sentential code-switching, the use of two languages in one sentence. Like all forms of natural language, code-switching is constrained by complex, unconscious mental rules that form speakers' grammars. Our view of the nature of these rules largely follows the tradition of Chomsky (1995) *et seq.*; we assume an invariant computational system, with cross-linguistic variation resulting from differences in the features of a language's lexical items, especially the functional heads. We follow MacSwan (1999) in assuming code-switched sentences are only different in that the lexical items, with their attendant syntactic features, come from multiple languages. Rather than special rules specific to code-switching, such sentences are simply the result of the interaction of those features when applying the same universal syntactic principles of monolingual sentences. (For a similar approach, see also, *inter alia*, González-Vilbazo and López 2011, 2012; López 2020; MacSwan 1999, 2004, 2013; Woolford 1983.)

By incorporating syntactic features from multiple languages in a single derivation, we can observe the effect of combinations of features that would not otherwise mix, providing insight that is not available with only monolingual data (González-Vilbazo et al. 2013; González-Vilbazo and López 2011, 2012). In this way, code-switching acts as a microscope of sorts, allowing us to isolate and observe syntactic properties of interest. Such an approach, we contend, is especially valuable for a case like the *that*-trace effect, where the facts themselves are not in serious dispute, but the questions they raise have been the subject of enduring debate.

Given our position that code-switching is a valuable source of evidence for general linguistic theory, we must also consider how previous code-switching studies can inform the predictions made for the *that*-trace effect discussed in the next section. Four such studies are relevant.

First, González-Vilbazo and López (2012, 2013) use data from Spanish–German, Spanish–English, and other code-switching to propose the Phase Head Hypothesis: the phase head (C, in our case here, but also *v*) determines the properties of its complement. Their proposal aligns with Minimalist approaches that derive many properties of the clause from the properties of the phase heads, including approaches like Feature Inheritance, in which the features of T come directly from Vanden Wyngaerd (2020) argues for a similar conclusion using data from Dutch–English code-switching. In her experiment, she found that the language of the verb determined the availability of V2 word order, which suggested that traditional accounts of V2 as a property of the C head were not on the right track. However, acknowledging extant evidence showing that V2 was unlikely to be a property of the verb alone, Vanden Wyngaerd instead concludes that her results support a Feature Inheritance model, in which the features of C are transferred to T, which could explain why the language of the verb was crucial.

Two other experiments, both with Spanish–English code-switching, likewise argue for a closer connection between C and T. Sande (2018) tested null subjects in Spanish–English code-switching and concluded that null subjects are not possible when there is a switch between C and T. She hypothesizes that both C and T together are needed to license null subjects, proposing that T provides some relevant syntactic features (some variant of the EPP that allows null subjects) while C provides relevant discursive features (e.g., the topic reading that generally accompanies null subjects), so both are necessary. She calls this

the C+T Hypothesis, but she does not test it directly (her experiment only included items with switches between C and T). In previous work (Ebert and Hoot 2018; Hoot and Ebert 2021), we tested her hypothesis in the domain of subject position (rather than null subjects), conducting an experiment with Spanish–English code-switching in which we found that, indeed, the only sentences that permitted post-verbal subjects were those in which both C and T were in Spanish; whenever either head was in English, post-verbal subjects were not possible.

All these findings bear directly on our current predictions. Given that we know that null subjects and post-verbal subjects generally entail the availability of subject extraction, and given the hypothesis that both null subjects and post-verbal subjects are only licensed when both C and T are in Spanish, previous findings from code-switching lead us to expect that subject extraction will likewise only be available when both C and T are in Spanish. This prediction has implications for the predictions made by existing theoretical accounts of the *that*-trace effect, to which we turn in the next section.

1.3. Theoretical Accounts of the That-Trace Effect

Pesetsky (2017) reviews the history of theories of the *that*-trace effect, from Perlmutter's (1968, 1971) initial observation through early accounts familiar to many syntacticians, like the Nominative Island Condition, the Doubly Filled Comp Filter, and the Empty Category Principle; we refer readers there for a full review. Here, we focus on five contemporary accounts, summarized in Table 1. Some theorists claim that extraction from Spec-TP is universally barred, so that the difference between languages derives only from the EPP, while others ascribe extraction bans to language-specific causes. Similarly, some posit a common underlying cause for the EPP and the *that*-trace effect, while others derive them separately. We review each proposal's details in turn to motivate its predictions for code-switching.

Table 1. Theoretical accounts of the *that*-trace effect.

Account	Basic Claim	Spanish Is Different Because . . .	Ban on Extraction from Spec-TP Is . . .	Prediction for CS
Anti-Locality (Douglas 2017; Erlewine 2016, 2020)	Movement from Spec-TP to Spec-CP is too short; extraction from Spec-TP thus universally barred.	It does not have the EPP; Spanish allows extraction from post-verbal position, so movement is not too short.	Universal.	Extraction only from post-verbal position in CS, so whatever determines subject position (C+T?) determines extraction in CS.
Criterial Freezing (Rizzi 2006, 2015; Rizzi and Shlonsky 2007)	Positions with interpretive properties (like subjects) are frozen; extraction from Spec-TP thus universally barred.	Null expletive fills the subject position in Spanish, so subject can be extracted from lower position.	Universal.	Null expletives permit extraction, so whatever determines null subject availability (C+T?) determines extraction in CS.
Prosodic Alignment (Kandybowicz 2006, 2009; McFadden and Sundaresan 2018; Sato and Dobashi 2016)	Empty Spec-TP cannot align with left edge of intonational phrase (or cannot form phrase with C) so syntax/prosody matching fails; extraction from Spec-TP thus universally barred.	V-to-T movement means V is highest head in intonational phrase and therefore at left edge, which is thus not empty.	Universal.	Assuming V-to-T is a property of T, language of T determines CS behavior.
Labeling (Chomsky 2013, 2015)	T is deficient and cannot be labeled alone; can only be labeled with subject, so subject is frozen.	T is not deficient (because of 'rich agreement').	Language-specific.	Language of T determines CS behavior.
T-to-C (Pesetsky and Torrego 2001)	T raised to C surfaces as *that*; extracting a subject is more economical and blocks T raising, so **that*-t.	Spanish C is a true complementizer, not an instance of T in C.	Language-specific.	Language of C determines CS behavior.

1.3.1. Anti-Locality

Some scholars have proposed a universal restriction on extraction from Spec-TP due to anti-locality, the notion that there is a prohibition on syntactic movements that are too short (Erlewine 2016). The central claim is that the embedded-clause subject moves from Spec-TP to Spec-CP to escape the C phase, moving only one maximal projection, violating

anti-locality and creating the *that*-trace effect. Languages can avoid the too-local movement by avoiding movement either to Spec-CP or to Spec-TP. English null complementizers are an example of the first strategy: either bundling C and T together (Erlewine 2020) or reducing the CP structure (Douglas 2017) allows the subject to escape the phase without moving to Spec-CP. Spanish takes the second tack: the subject is extracted directly from a lower position to Spec-CP, avoiding the restriction on extraction from Spec-TP. In other words, languages like Spanish that do not require the subject to raise to Spec-TP (i.e., those that do not have a traditional EPP requirement) can leave the subject in situ and directly extract from this post-verbal position.

Traditional accounts of the EPP have traced cross-linguistic differences in subjects to properties located in T (e.g., the "rich agreement morphology" of the verb raising to T to satisfy the subject requirement, as in Alexiadou and Anagnostopoulou 1998), so predicting that the language of T will determine code-switching behavior under the Anti-Locality account seems sensible. However, in this case, we must consider our previous empirical evidence of subject position in code-switching: we found that post-verbal subjects were only acceptable when both C and T were in Spanish (Ebert and Hoot 2018; Hoot and Ebert 2021). That is, Spanish T alone was not enough to permit post-verbal subjects, contrary to the usual view of the EPP as a feature of T. If we use this evidence to inform our prediction, then the Anti-Locality account predicts that subject extraction will only be possible when both C and T are in Spanish, because subject extraction follows directly from the availability of post-verbal subject position, whatever its cause.

1.3.2. Criterial Freezing

Rizzi (2006, 2015; Rizzi and Shlonsky 2007) identifies certain positions in the syntactic derivation that possess semantic, discourse, or scope properties as "criterial positions", and argues "classical EPP, the requirement that clauses have subjects, can be restated as a criterial requirement, the Subject Criterion" (Rizzi and Shlonsky 2007, p. 116). A constituent moved to subject position[3] has a criterial feature connected to interpretive properties, which furthermore freezes it in place due to what Rizzi calls Criterial Freezing. Thus, Rizzi connects the interpretation of subjects, the EPP, and the inability to extract subjects via the notions of criterial positions and Criterial Freezing.

The feature involved in Criterial Freezing is valued via the c-command relationship between C and the subject, and variations in the heads projected in the C domain are what give rise to the differences between English *that* and null C. When there is an overt complementizer, though, Criterial Freezing is held to be universal: even in a language like Italian or Spanish, a subject raised to Spec-TP is frozen in place. What makes Italian or Spanish different is that they have a phonetically null pronoun *pro* in their lexicon, which can occupy Spec-TP and satisfy the Subject Criterion. In subject extraction cases, the null expletive *pro* occupies Spec-TP, allowing the overt subject to remain in a lower position. Since the lower subject never raises to Spec-TP, it can be extracted without trouble.

Criterial Freezing is unlike the other accounts reviewed here in its locus of cross-linguistic difference. For Rizzi, the key difference between Spanish and English resides not on C or T but instead turns on the presence of null expletive *pro* in the lexicon. It is not immediately clear what predictions such a proposal makes for code-switching. If the null expletive can be drawn into any derivation (we see no *a priori* reason it could not be), we would expect *all* code-switching sentences to permit *that*-trace violations and null subjects, on the assumption that they could always use *pro* to satisfy the Subject Criterion, contrary to fact. An alternative is to assume a connection between the availability of null subjects and null expletives more specifically, as Rizzi implicitly does. As mentioned in Section 1.2, Sande (2018) investigated null subjects in Spanish/English code-switching with switches between C and T and found that they were rated lower regardless of which head was in Spanish. She hypothesized that features of both C and T are needed to license null subjects. We can extend this idea to the predictions for Criterial Freezing. The Criterial Freezing approach predicts that subject extraction will be allowed if the Subject Criterion

can be satisfied with a null expletive. Assuming that, for Spanish, null expletives are only available if null subjects more generally are available,[4] and assuming Sande's proposal is on the right track, then the Criterial Freezing approach would also likely predict that subject extraction would be available only when C and T are both in Spanish.

1.3.3. Prosodic Alignment

Although most research on the *that*-trace effect has postulated syntactic explanations, several researchers have instead looked to the syntax/prosody interface. These approaches largely share the perspective that extraction from Spec-TP is disallowed universally; where they differ is the cause. For Prosodic Alignment approaches, no syntactic feature directly prohibits further movement of subjects; rather, requirements on the alignment of syntactic phrases to prosodic or intonational phrases are violated when the Spec-TP position is empty.

The key claim is that, when Spec-TP is phonetically empty, it is not possible to create a licit prosodic structure. The details differ slightly. Sato and Dobashi (2016), drawing on Truckenbrodt's (1999) Lexical Category Condition, argue that a functional head like C and a prosodically empty trace or copy are both invisible to the prosody, resulting in an illicit empty phonological phrase. McFadden and Sundaresan (2018) argue that TP, as a spell-out domain, forms an intonational phrase, whose left edge (i.e., Spec-TP) cannot be empty. Kandybowicz (2006, 2009) combines elements of both. Despite differing details of their syntactic and prosodic analysis, what matters in each case is that the Spec-TP position cannot be phonetically null.

Appealing to prosody has the advantage of accounting for the fact, noticed since at least Bresnan (1977), that inserting a parenthetical aside between *that* and the TP can ameliorate the *that*-trace effect, as shown in (6), which is unexpected in syntactic accounts that take the parenthetical to be outside the structure of the main clause.

6.　　Who do you think that, for all intents and purposes, t_{who} will actually call the shots?

On the other hand, Ritchart et al. (2016) and Toquero-Pérez (2021) have tested several consequences of prosodic approaches to the *that*-trace effect and found counter-evidence, so the question is not at all settled, which is why we endeavor to test them here.

How do these approaches deal with Spanish? For Kandybowicz, who contrasts effects in English and Nupe, the analysis can depend on the independently attested intonational facts in each language. While English and Spanish prosody differ in many ways, we would need to make many assumptions about prosody in English, Spanish, and code-switching to make a prediction based on the languages' different prosodic phrasing. Without direct prosodic or intonational data (and we know little, as a field, about the prosody of code-switching), any such prediction would be ill-advised. Sato and Dobashi's account concerns English alone, so we will not speculate what they would predict. The only one of the prosodic accounts that makes a testable prediction is McFadden and Sundaresan's. They suggest that the difference between Spanish and English may come down to yet another fact that has been bundled together in the past with the null subject parameter: V-to-T movement. It is generally accepted that Spanish verbs raise to T, unlike English verbs (Villa-García 2018). McFadden and Sundaresan speculate that the verb's raising means that the highest head in the TP is filled with phonetic material, such that the left edge of the intonational phrase aligned to TP is no longer empty; it is filled by the verb.[5]

Applying this prediction to code-switching, we expect that whatever determines V-to-T movement in code-switching will likewise determine the availability of subject extraction. If V moves to T, then the left edge of the intonational phrase will be filled whether the subject is extracted or not, so V-to-T movement should entail obviation of the *that*-trace effect. McFadden and Sundaresan do not specify, but it seems sensible to assume that the feature responsible for V-to-T movement is a feature of either V or T; either way, it comes down to the language of the T head. Therefore, we take the prediction of this account to be that extraction will be possible when T is in Spanish.

1.3.4. Labeling

Another proposal that unifies the EPP and the *that*-trace effect is Chomsky's labeling approach. Chomsky (2013, 2015) proposes a Labeling Algorithm for phrases. Crucially for the *that*-trace effect, some heads, including T in English, cannot label a phrase on their own. However, if the subject raises to Spec-TP, the phi-agreement between them allows the phi-features, as the 'most prominent' features of the syntactic object, to label the phrase, thus deriving the EPP from a labeling requirement. Labeling does not occur until Transfer, so when the wh-subject is raised from Spec-TP to Spec-CP in a long-distance wh-question, TP no longer has the subject to provide the label, generating the *that*-trace effect. For a null complementizer, Chomsky assumes that C is deleted. T inherits the properties of C, including phasehood, allowing the subject to remain in Spec-TP at Transfer and therefore permitting TP to be labeled, obviating the *that*-trace effect.

Under this approach, what makes languages like Spanish different is the rich agreement on T, allowing it to label the phrase even if the subject is extracted or if it does not raise. This predicts that subject extraction will be acceptable as long as T is in Spanish.

1.3.5. T-to-C

Pesetsky and Torrego (2001) propose an account that unifies several COMP-trace effects with apparently unrelated phenomena like subject-verb inversion in questions via particular features on C, though we focus here on the *that*-trace effect. In addition to an unvalued wh-feature uWh, the embedded C also has an unvalued T feature uT which can be valued in two ways: by a subject in Spec-CP, since the subject would have first agreed with T, or by T-to-C movement. With a wh-object, both ways are possible. When uT is valued by the subject, C is phonetically null, since Pesetsky and Torrego assume C is always null in English. When uT is valued by T-to-C movement, T in C spells out as *that*. With a wh-subject, which already must move to Spec-CP to satisfy uWh, the subject also values uT, making T-to-C movement redundant and therefore illicit. Without T in C, C is null. For Pesetsky and Torrego, then, it is not that the complementizer blocks subject extraction, but rather the reverse: more-economical subject extraction blocks the T-to-C movement necessary to realize C as *that*.

What about Spanish? Pesetsky and Torrego do not offer a full analysis of Spanish, but they do suggest that the difference is that C in Spanish is unlike English in that it is not phonetically null; it is always realized as *que*. This assumption accounts for the obligatory nature of *que* and correctly suggests no subject/object asymmetry, because two movements will be needed no matter which argument you extract. This makes a clear prediction for code-switching: the features of C will determine the behavior of the derivation.

1.4. Research Questions

Bringing together the previous sections, we use code-switching to test the relevant theoretical proposals by taking advantage of the cross-linguistic variation in features between English and Spanish. More concretely, our goal is to examine how the combination of the relevant functional heads—C and T—from English and Spanish in a code-switching sentence affects the acceptability of subject extraction, and then consider the implications of those findings, leading to the research questions in (7).

7. Research questions

a. What combinations of C and T permit subject extraction over an overt complementizer in Spanish–English code-switching?

b. Which theoretical accounts of the *that*-trace effect are supported by the code-switching evidence?

2. Materials and Methods

We carried out a formal acceptability judgment experiment designed to test the *that-trace* effect in code-switching with the four possible combinations of C and T, as well as monolingual sentences in English and Spanish.

2.1. Participants

Participants were simultaneous or early sequential Spanish–English bilinguals, *n* = 36, who started learning both Spanish and English before age 7, were raised and had completed high school in the United States, and reported hearing and using Spanish in childhood and with one or more caregivers. All were self-reported code-switchers, i.e., they reported using both Spanish and English in the same conversation. Participants completed a background questionnaire, including self-ratings of proficiency and estimated language use; these characteristics are reported in Table 2.

Table 2. Participant characteristics.

Characteristic	Mean (SD, Range)
Age	21.2 (3.2, 18–31)
Overall self-rating, English, 1–5, 5 max.	4.9 (0.3, 4–5)
Overall self-rating, Spanish, 1–5, 5 max.	4.1 (0.9, 2–5)
Self-reported English usage, all domains	69.6% (26.4, 0–100)
Self-reported Spanish usage, all domains	35.9% (29.0, 0–100)
LexTALE_Esp score	28.5 (9.4, 12–49)

Participants also completed a Spanish lexical decision task, the LexTALE_Esp (Izura et al. 2014), designed to measure vocabulary size while controlling for guessing, which we used as a proxy for global Spanish proficiency. Because Spanish is a minority language in the United States, US-raised Spanish–English bilinguals (often called heritage speakers of Spanish) display a wide range of Spanish proficiency, from so-called 'overhearers' who retain only limited receptive knowledge to highly proficient bilinguals (Polinsky 2018). Heritage speakers are native speakers of Spanish (Rothman and Treffers-Daller 2014), but, of course, bilingual language knowledge is necessarily different from that of monolinguals (Grosjean 1985), so we chose to limit our group to those participants most likely to possess language knowledge that resembled that of other native speakers. We therefore excluded those participants whose score on the LexTALE_Esp was at or below the level Izura et al. (2014) found for beginning L2 learners,[6] on the assumption that these low-proficiency speakers likely diverge the most from other Spanish speakers. (We also anticipated that very low proficiency speakers would be unlikely to be able to complete our task, which included long and complex sentences.) We made 18 exclusions based on said threshold (not included in the participant numbers above). We did not test their English proficiency directly because these speakers, like most heritage speakers of Spanish in the United States, reported English dominance and were much less likely to differ in a relevant way from other native speakers of English.

Following Juzek and Häussler (2015), we also excluded "non-cooperative" participants, who did not complete the task in good faith, instead merely clicking through the items without reading them. We calculated the threshold below which meaningful participation was infeasible to be 1000 milliseconds (based on calculations by Bader and Häussler 2010; Juzek 2016); any participant who entered 20% or more of their judgments in less than 1000 milliseconds was excluded. We made 11 such exclusions (also not included in the numbers above).

2.2. Procedure

Participants were recruited in person at two large universities in the Midwest United States, but they completed the experiment over the internet via Ibex (Drummond 2017).

(On the validity of judgment experiments via the internet, see Sprouse 2011). Initial task instructions and training were presented in Spanish–English code-switching to help establish code-switching mode (González-Vilbazo et al. 2013). After the training, participants judged 112 code-switched sentences, presented individually. They registered their judgment by clicking on one number in a seven-point scale, with 7 labeled *bien* '[sounds] good' and 1 labeled *mal* '[sounds] bad', with intermediate values unlabeled. Upon choosing a number, participants were advanced automatically to the next sentence, and each judgment was untimed (although reaction times were recorded). The order of the sentences was pseudo-randomized such that two target sentences did not occur sequentially.

After judging the code-switched sentences, participants judged two blocks of single-language sentences—48 Spanish and 48 English—with the order of the blocks rotated by participant. The presentation of the single-language sentences was the same, using the same judgment scale and pseudorandomization. After each block of sentences, participants were instructed to take a break if needed. Finally, participants completed the background questionnaire, followed by the LexTALE_Esp.

2.3. Materials

As pointed out by González-Vilbazo et al. (2013), in addition to carefully controlling the speaker population to ensure comparability between bilinguals and existing data (usually drawn from monolinguals), it is important to verify that bilinguals' language knowledge in fact aligns with what we expect for a given language, given that bilinguals necessarily have different linguistic experiences than monolinguals do. To that end, we tested the *that*-trace effect in monolingual English and Spanish sentences, as a conceptual preliminary to our code-switching sentences.

The monolingual English stimuli had a 2 × 2 factorial design, as shown in Table 3: Wh-extraction (Subject vs. Object) and realization of C (*that* vs. Null).

Table 3. Monolingual English experiment design.

Wh	C	Example
Object	*That*	What did the editors confirm <u>that</u> the author had written in only one week?
Object	Null	What did the editors confirm ___ the author had written in only one week?
Subject	*That*	Who did the editors confirm <u>that</u> had written the book in only one week?
Subject	Null	Who did the editors confirm ___ had written the book in only one week?

The monolingual Spanish stimuli had the same 2 × 2 factorial design, as shown in Table 4.

Table 4. Monolingual Spanish experiment design.

Wh	C	Example[7]
Object	*Que*	¿Qué confirmaron las editoras que el autor había escrito en tan solo una semana?
Object	Null	¿Qué confirmaron las editoras ___ el autor había escrito en tan solo una semana?
Subject	*Que*	¿Quién confirmaron las editoras que había escrito el libro en tan solo una semana?
Subject	Null	¿Quién confirmaron las editoras ___ había escrito el libro en tan solo una semana?

The code-switching stimuli were designed to test the availability of subject extraction in sentences with different combinations of C and T, so it had a 2 × 2 × 2 design: Wh-extraction (Subject vs. Object), language of C (English vs. Spanish), and language of T (English vs. Spanish), as shown in Table 5.

Table 5. Code-switching experiment design.

Wh	C	T	Example[8]
Object	*That*	En	*Qué asumieron los maestros* <u>that</u> the child had read before the test?
Object	*That*	Sp	What did the teachers assume <u>that</u> *el niño había leído antes del examen*?
Object	*Que*	En	*Qué asumieron los maestros que* the child had read before the test?
Object	*Que*	Sp	What did the teachers assume *que el niño había leído antes del examen*?
Subject	*That*	En	*Quién asumieron los maestros* <u>that</u> had read the text before the test?
Subject	*That*	Sp	Who did the teachers assume <u>that</u> *había leído el texto antes del examen*?
Subject	*Que*	En	*Quién asumieron los maestros que* had read the text before the test?
Subject	*Que*	Sp	Who did the teachers assume *que había leído el texto antes del examen*?

We imposed several constraints on the design of the stimuli to avoid possible confounds, each presented with its rationale in Table 6. The monolingual stimuli were subject to the same controls on frequency, verb forms, and arguments, but obviously not those specific to code-switching.

Table 6. Experiment token constraints.

Constraint	Rationale
Only one code-switch per sentence.	Avoid single-word switches and awkwardly repeated switches.
Arguments and verbs among the 5000 most common Spanish (Davies 2006) and English (Davies 2008) words.	Avoid frequency effects; limit chance of rejection due to unfamiliar words.
English C always *that* and Spanish C always *que*.	Unambiguously identify the language of C.
Matrix clause verbs plural and lower clause verbs singular.	Avoid misinterpretation of extracted wh-word as pertaining to matrix clause.
Matrix clause verbs all verbs of assertion or belief that take clausal complements and are unlikely to take human objects (following Ritchart et al. 2016).	Avoid misinterpretation of extracted wh-word as object of matrix verb.
Embedded-clause verbs all simple transitives chosen to semantically favor animate subjects and inanimate objects.	Avoid confusion of embedded subjects and objects.
Lower clause verb always pluperfect, with auxiliary verb assumed to instantiate T.	Make language of T head unambiguous; three-syllable pluperfect *había* 'had' used rather than one-syllable present perfect *ha* 'has' to make it more salient.
Matrix verbs simple preterit.	Consistent tense across items; past-tense verbs sound natural in information-seeking questions.
Matrix subjects always definite, animate, human, and plural.	Avoid possible extraneous grammatical effects; make subject clear via subject-verb agreement.
Embedded subjects always definite, animate, human, and singular.	Avoid possible extraneous grammatical effects; make subject clear via subject-verb agreement.
Embedded objects always definite and inanimate.	Avoid possible extraneous grammatical effects; make object clear via verb semantics.
No initial inverted question mark for code-switching.	Make sentences consistent whether the first part is in Spanish or English.

Of these design features, we want to highlight our choice to limit each sentence to a single code-switch, such that it always started in one language and switched to the other just once. We chose this design because keeping the language of the matrix clause constant would have resulted in some sentences with awkward, single-word code-switches or with several code-switches in rapid succession, which are also quite unnatural (and would have multiplied the number of sentences unmanageably). However, our choice introduces another possible confound, which we acknowledge: the language of the matrix clause (and thus the wh-word) is always the opposite of the language of T. We agree with an anonymous reviewer who points out that this confound could potentially affect the

assumptions underlying our design: perhaps, despite Spanish C and T licensing post-verbal subjects, the features of English *who* do not allow it to appear post-verbally. If so, perhaps extracted *who* cannot be launched from a post-verbal position, complicating the interpretation of our data for theories that tie subject extraction to that position (e.g., Rizzi 1982). While we recognize this possibility, some existing data suggests that English subjects are not generally barred from post-verbal positions in Spanish–English code-switching. In two acceptability judgment experiments, Ebert (2014) and Ebert and Koronkiewicz (2018) found that some code-switching sentences with post-verbal English subjects, as in (8), received high ratings, suggesting that such a position is in principle available to English subjects, given the appropriate environment.[9]

8.	*No*	*recuerdo*	how much money	*han*	*robado*	those criminals
	no	remember.1SG		have.3PL	stolen	
	en	*la*	*última*	*década*.		
	in	the	last	decade		

'I do not remember how much money those criminals have stolen in the last decade.'

We created 48 lexicalizations (also called 'token sets') with the characteristics in Table 6, resulting in 384 total code-switching tokens. These we distributed across 24 lists using a Latin square, so that each participant saw 4 tokens per cell of the experiment design without repeating any lexicalization. Participants also judged 80 filler items with a range of acceptability, for 112 total judgments.

For the monolingual sentences, we similarly created 32 lexicalizations for each language, a total of 128 tokens per language. We distributed them across 8 lists using a Latin square, so that each participant saw 4 tokens per cell of the experiment design without repeating any lexicalization within a given language. The Spanish and English items were translation equivalents, but a given participant never saw the exact same token in translation. Each monolingual block of 48 items also contained 32 fillers.

As explained in Section 2.2, participants completed some other materials in addition to the three judgment tasks. Prior to the tasks, they completed a brief training with announced practice items and anchor items (following recommendations of Schütze and Sprouse 2013) and a section specifically focused on code-switching (following González-Vilbazo et al. 2013). After the judgment tasks, they also completed the background questionnaire and LexTALE_Esp, as discussed above. All our items are available online; see the Supplementary Materials endnote.

3. Results

3.1. Data Preparation and Analysis

Linguistic theory is generally concerned with *grammaticality*—whether a given sentence can be generated by someone's mental grammar—but grammars cannot be directly observed. Instead, we infer the properties of a speaker's mental grammar by observing their linguistic behavior, which can range from recording their naturalistic speech to measuring their reaction times to asking their judgments of sentences directly. Sentence judgments can be informally collected by analysts or part of a formal experiment, which is the tack we take here. Regardless, sentence judgments provide us evidence not of grammaticality but of *acceptability*, a complex mental construct of a person's subjective reaction to a sentence's apparent well-formedness. Without a doubt, a sentence's grammaticality affects its acceptability, but acceptability also varies according to other factors, including processing difficulty, semantic plausibility, word and structure frequency, and more. One great advantage of collecting acceptability judgments through a formal experiment is the opportunity to account for some of these confounds via the controls that we impose on our stimuli. We explicitly control for frequency, context, semantic plausibility, and many other factors as we craft our tokens, which allows us to hold constant, as much as possible, sources of variation in sentence acceptability that are not of interest to our research questions.

Within formal experiments, the *factorial* design we employ here—adopting the influential paradigm of Sprouse and colleagues (see Schütze and Sprouse 2013; Sprouse et al. 2016)—is especially apt for isolating grammatical effects of interest from other sentence features that can affect ratings. In this design, each factor may affect the acceptability of the sentence for independent reasons. For instance, perhaps the addition of the word *that* always makes a sentence sound slightly stilted, or subject questions just generally sound less plausible than object questions. We might then expect sentences with *that* or sentences with subject questions to be rated slightly lower, yet such a rating does not necessarily reflect ungrammaticality but rather some confounding feature. If we put those features together—a sentence with *that* and subject extraction—and we observe that the rating decreases roughly along the lines of each independent effect added together, we might conclude that there is no special grammatical restriction; instead, we observe a merely additive effect, in which the extraneous non-grammatical (or grammatical but not of interest for our purposes) features of the sentence lower its acceptability. If, on the other hand, we observe that sentences with both *that* and subject questions decrease in acceptability more than expected based on either factor alone—what Sprouse and colleagues call a *super-additive effect*—we can conclude that something else is going on: there is an additional factor lowering the acceptability of such sentences more than either individual piece would predict, which we can then take to be our predicted grammatical restriction. This does not necessarily mean that we can say that the super-additive effect indicates ungrammaticality *per se*, since judgments can never directly determine grammaticality, but they can give us evidence we can use to infer it.

In what follows, then, we will be looking for super-additive effects in our factorial designs. When we conduct statistical tests, a super-additive effect is indicated by a significant interaction between the factors. Significant main effects of the factors are less revelatory, so we will focus largely on interpreting the interactions. Super-additive effects can also be observed visually in the plots of the data. Simple (additive) effects will take the form of two roughly parallel lines because each factor may decrease acceptability somewhat but in a linear way, whereas a super-additive effect will have a characteristic 'alligator mouth' shape, such that one pair of means is further apart than the other pair, resulting in non-parallel lines. Finally, we will also look for super-additive effects by calculating a *differences-in-differences* (DD) score, which expresses in one number how much larger the difference between one pair of means is than the difference between the other pair of means.

To calculate those scores and to perform our statistical tests, we follow Schütze and Sprouse's (2013) advice and z-score transform the raw ratings by participant. This transformation accounts for scale compression and skew, which is of special concern for code-switching data (Badiola et al. 2018), and expresses the ratings in standardized units. We largely focus on z-scores in reporting the results. However, the raw scores can be interpretable as well, so we report those alongside the z-scores, although they are not our focus.

For each 2 × 2 design, we fit a linear mixed-effects model (LMM) to the data. As Meteyard and Davies (2020) point out, LMMs are powerful analytical tools that can be conceptualized in different ways and whose use is not standardized in the field, so it is incumbent on authors to report their decisions in model fitting and interpretation in as much detail as possible. For every LMM we carried out, the fixed factors were the two grammatical factors at play, which were effect-coded (−0.5, +0.5), plus their interaction. The dependent variable was the z-score of the ratings. We included by-participant and by-item random effects in each model to account for repeated measures. For the by-item effects, 'item' refers to the lexicalization or token set a given token is drawn from, to control for characteristics of the particular words employed. The random effects structure was determined in each case using the top-down method that Barr et al. (2013) employed for their simulations. We began with the maximal model containing all by-participant and by-item slopes and intercepts. If it did not converge, we removed the random slope

that accounted for the least variance. If more than one random slope accounted for zero variance, we removed by-item random slopes first. We continued to remove random slopes until convergence was achieved.

We report the *F* test that is the output of a Type III test of fixed effects for each fixed factor in our models, along with its associated *p* value. Although LMMs can also be understood as regressions—and for the sake of completeness, we report the regression coefficients—we conceptualize the purpose of these tests to be more like ANOVA than regression, and we report the results in a way that reflects that. Where we find significant interactions, we explore them with pairwise post hoc comparisons using the Bonferroni correction for multiple comparisons, even though we recognize that it is likely overly conservative (Larson-Hall 2010).

3.2. Preliminary Step 1: Monolingual English Results

Recall that the purpose of this preliminary test was to verify that the variety of English spoken by our participants matched the expected features. It tested their English via a 2 × 2 factorial design: wh-extraction (Subject vs. Object) and realization of C (*That* vs. Null). The raw ratings (scale = 1–7) and estimated marginal mean z-scores appear by condition in Table 7. The z-scores are also shown in the interaction plot in Figure 1.

Table 7. Monolingual English experiment results.

| Wh	C	Example	Mean Rating	EMM Z-Score
Object	*That*	What did the editors confirm <u>that</u> the author had written in only one week?	5.77	0.499
Object	Null	What did the editors confirm ___ the author had written in only one week?	5.71	0.457
Subject	*That*	Who did the editors confirm <u>that</u> had written the book in only one week?	4.12	−0.118
Subject	Null	Who did the editors confirm ___ had written the book in only one week?	5.93	0.535
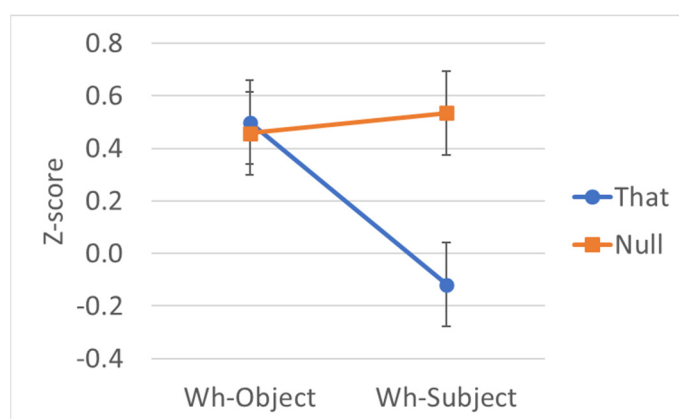
Figure 1. Estimated marginal mean z-scores by wh-extraction and realization of C, English monolingual data (error bars = 95% confidence intervals).

As described in Section 3.1, we fit a linear mixed-effects model to the data and present the full results in Table 8.

Table 8. Linear mixed-effects model, monolingual English experiment.

Fixed Effect	Coefficient (β)	SE	95% CI	*t*	*F*	*p*
C	−0.31	0.07	−0.44–−0.17	−4.55	20.66	<.001
Wh	0.27	0.08	0.11–0.43	3.51	12.31	.002
Wh*C	0.69	0.12	0.44–0.95	5.58	31.18	<.001
Random Effects			**Variance**		**SE**	
By-Subject intercept			0.019		0.021	
By-Subject slope over C			0.011		0.022	
By-Subject slope over Wh			0.017		0.023	
By-Subject slope over Wh*C			0.006		0.029	
By-Item intercept			0.021		0.020	
By-Item slope over Wh			0.021		0.024	
By-Item slope over Wh*C			0.011		0.021	

All three fixed factors produced significant effects, but the result of interest is the interaction (which also likely drives the other two effects). The interaction indicates the presence of a super-additive effect, which is also clear in the characteristic pattern in Figure 1. Bonferroni-corrected post hoc pairwise comparisons confirm that subject extraction over *that* is rated significantly worse than object extraction ($p < .001$) and worse than subject extraction without *that* ($p < .001$). As a measure of effect size, we calculated a *differences-in-differences* (DD) score of 0.695.

The pattern we observe for monolingual English is exactly what is expected from the literature. This finding is noteworthy especially in light of early experimental work that did not find experimental evidence of the *that*-trace effect in some dialects of English (Sobin 1987), supporting instead the emerging experimental consensus that the *that*-trace in English is robust (Cowart and McDaniel 2021).

3.3. Preliminary Step 2: Monolingual Spanish Results

The Spanish monolingual experiment was similarly designed to verify that these participants' grammars resembled what was expected from the literature. Its 2 × 2 factorial design included wh-extraction (Subject vs. Object) and realization of C (*Que* vs. Null). The raw ratings (scale = 1–7) and estimated marginal mean z-scores of the ratings are presented in Table 9 and the z-scores are displayed in Figure 2.

Table 9. Monolingual Spanish experiment results.

| Wh	C	Example	Mean Rating	EMM Z-Score
Object	*Que*	¿Qué confirmaron las editoras que el autor había escrito en tan solo una semana?	5.82	0.371
Object	Null	¿Qué confirmaron las editoras ___ el autor había escrito en tan solo una semana?	3.28	−0.675
Subject	*Que*	¿Quién confirmaron las editoras que había escrito el libro en tan solo una semana?	4.60	−0.113
Subject	Null	¿Quién confirmaron las editoras ___ había escrito el libro en tan solo una semana?	3.90	−0.427
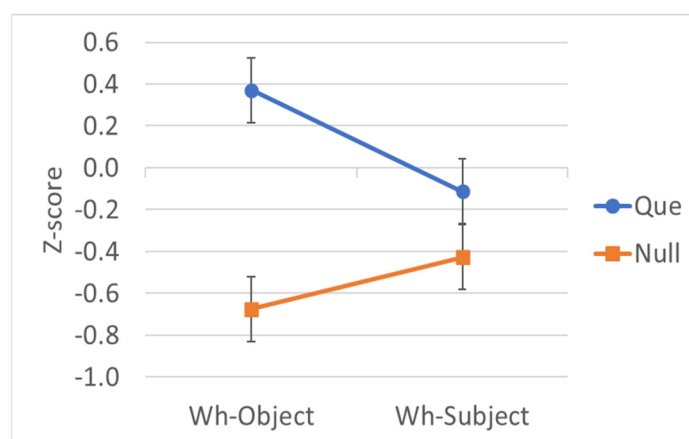
Figure 2. Estimated marginal mean z-scores by wh-extraction and realization of C, Spanish monolingual data (error bars = 95% confidence intervals).

As described in Section 3.1, we fit a linear mixed-effects model to the data and present the full results in Table 10.

Table 10. Linear mixed-effects model, monolingual Spanish experiment.

Fixed Effect	Coefficient (β)	SE	95% CI	*t*	*F*	*p*
C	0.68	0.07	0.54–0.82	9.25	85.61	<.001
Wh	0.12	0.07	−0.03–0.26	1.61	2.59	.108
Wh*C	0.73	0.15	0.44–1.02	4.98	24.79	<.001

Random Effects	Variance	SE
By-Subject intercept	0.020	0.017
By-Item intercept	0.004	0.012

Significant effects were found for realization of C and for the interaction, which indicates a super-additive effect. Bonferroni-corrected post hoc pairwise comparisons revealed that both types of wh-extraction are rated significantly worse when C is null (*p* < .001 for objects, *p* = .003 for subjects), although the ratings are much closer for subject extraction—the 95% confidence intervals of the estimated z-scores nearly touch (although they do not overlap), as seen in Figure 2. Post hoc tests also revealed that wh-subject extraction was rated worse than wh-object extraction when C was *que* (*p* < .001), but, surprisingly, that wh-subject extraction was rated *better* than wh-objects when C was null (*p* = .017). As a measure of effect size, we calculated a *differences-in-differences* (DD) score of 0.732.

At first blush, these monolingual results appear to suggest that our participants do not have the necessary syntactic features in their Spanish to test our hypotheses; however, we conducted two follow-up experiments to understand this puzzling outcome, and that additional evidence suggests that the issue is a task effect due to an unanticipated confound in the sentences that omit *que* while extracting the subject. Because understanding our participants' Spanish is crucial to our central purpose, it is worth a brief excursion to contextualize and explain the two follow-up experiments, to which we now turn.

To begin, recall that, given the heterogeneity of bilingual language acquisition, these stimuli test whether participants' linguistic systems contain the key properties under study for our language dyad. This includes two properties for Spanish: (i) an obligatory complementizer and (ii) no subject/object extraction asymmetry. We found a consistent penalty to acceptability for omitting the complementizer (despite the unexpected small effect by extraction type within the omission cases), which we take to indicate the expected ungrammaticality of *que*-omission. Yet, we also found an unexpected subject/object asymmetry when the complementizer was overt.

What is the cause of this puzzling finding? Perhaps our bilingual speakers simply have different linguistic knowledge than monolinguals (after all, that's why we're checking this in the first place). However, the only other experimental study of the *that*-trace in Spanish of which we are aware (Chacón et al. 2015) found the exact same pattern in (presumably monolingual) Spanish, which suggests that our results are not due to bilingualism but instead accurately represent the facts about Spanish.

To verify this supposition, we conducted a first follow-up experiment with the same materials but with 20 monolingual speakers of Mexican Spanish. Other than the participants, everything about the experiment was the same. We present the full results in Appendix A, but the important fact is that we, like Chacón et al., found that the monolingual speakers displayed a very similar pattern to that of the bilingual speakers: monolingual speakers rated subject extraction over an overt complementizer significantly worse than object extraction and rated omitted *que* very low across the board.

Therefore, our bilingual speakers look just like monolinguals, suggesting that the unexpected bilingual results are not caused by influence from English or some other aspect of bilingualism. However, the mystery remains: why do both groups rate subject extraction low?

Chacón et al. (2015, p. 12) speculate "that Spanish has an independent bias against subject extraction that is revealed in the complementizer conditions, but is overshadowed in the no complementizer conditions by the degradedness of missing complementizers." Is this bias because subject extraction is in fact ungrammatical, despite claims in the literature, or might it be something else? To check, we tested simpler examples of subject extraction with an overt complementizer, such as in (9), in a second follow-up experiment, this one a small-scale, informal task with just 20 tokens, judged by 36 native Spanish speakers (not controlled for dialect, only for acquisition of Spanish in childhood). We found all examples of subject extraction to be acceptable, as predicted.

9. Simple, acceptable subject extractions

   a.
¿Quién	crees	que	lo	hizo?
who	believe.2SG	that	it.ACC	did.3SG

   'Who do you think did it?'

   b.
¿Quién	no	crees	que	llegue	a tiempo?
who	no	believe.2SG	that	arrive.SUBJ.3SG	on-time

   'Who do not you think will arrive on time?'

This second follow-up (and the consensus of other work on Spanish) suggests that subject extraction is allowed in Spanish. Perhaps, then, the "independent bias against subject extraction" that Chacón et al. observe has an extragrammatical cause (perhaps it is unusual, rare, or disfavored for discourse/pragmatic reasons); as we discussed in Section 3.1, such confounds are endemic to acceptability judgments.

If subject extraction is grammatical but generally penalized in acceptability this way, then the question becomes not "Why is subject extraction with *que* worse than expected?" but rather "Why is subject extraction with null C *better* than expected?" That is, why aren't those sentences even worse, producing a simple additive effect (two parallel, downward-sloping lines)? Chacón et al. speculate that the ungrammaticality of *que*-omission obscures the effect. Perhaps what they find is a floor effect: these sentences receive judgments that cluster at the low end of the scale, so there's no room to rate subject extraction with null C worse. That explanation is reasonable, and may also be playing a role in our data, but we raise another possibility as well. Consider the examples in (10). The acceptability of sentences like (10a) can be improved substantially if the matrix verb and subject *confirmaron las editoras* 'the editors confirmed' are instead interpreted as a parenthetical, converting the sentence into a monoclausal question, as in (10b).

10. Parenthetical possibility

a. *

¿Quién	confirmaron	las	editoras	había	escrito	el	libro
who	confirmed.3PL	the	editors	had.3SG	written	the	book
en	tan solo	una	semana?				
in	only	one	week				

'Who did the editors confirm that had written the book in only one week?'

b.

¿Quién,	confirmaron	las	editoras,	había	escrito
who	confirmed.3PL	the	editors	had.3SG	written
el	libro	en	tan solo	una	semana?
the	book	in	only	one	week

'Who, the editors confirmed, had written the book in only one week?'

If some participants were interpreting these sentences this way, it may have improved the acceptability for subject extraction for these cases. Tellingly, such a reading would improve the subject-extraction sentences alone. Object extraction would still be ungrammatical with a parenthetical reading of the intended main clause material, as shown in (11). Spanish requires subject-verb inversion in matrix questions, yet a parenthetical reading of (11a) as (11b) leaves *el autor* 'the author' as a matrix-clause subject without inversion.

11. Parenthetical impossibility

a. *

¿Qué	confirmaron	las	editoras	el	autor	había	escrito
what	confirmed.3PL	the	editors	the	author	had.3SG	written
en	tan solo	una	semana?				
in	only	one	week				

'What did the editors confirm that the author had written in only one week?'

b. *

¿Qué,	confirmaron	las	editoras	el	autor	había	escrito
what	confirmed.3PL	the	editors	the	author	had.3SG	written
en	tan solo	una	semana?				
in	only	one	week				

'What, the editors confirmed, the author had written in only one week?'

If our speculation is on the right track, it would predict that only the subject-extraction sentences would be improved by the parenthetical reading, yielding precisely the interaction we observed. Interestingly, a similar issue with a potential parenthetical reading arises in work on the *that*-trace effect in German (Cowart and McDaniel 2021; Kiziak 2010; Reis 1996), suggesting our conjecture is plausible. If so, then part of the issue here may be a task effect.[10]

What implications does this discussion have for interpreting the code-switching stimuli? First, we stress that the bilingual group's Spanish results look like the results from both our monolingual follow-up and from Chacón et al. (2015). Therefore, we have been able to verify that their Spanish has the properties that any other Spanish speaker's grammar would have, including obligatory complementizers. Regarding the apparent subject/object asymmetry, we contend that the evidence reviewed above suggests that divergences from expectations stem from extra-syntactic factors. More research is clearly needed on Spanish alone to understand these puzzling facts better, but our follow-up work with monolingual Spanish has demonstrated that our bilingual speakers have similar grammars to those of monolinguals, and it is clear that subject extraction is grammatical in Spanish. We therefore conclude that these participants still have the syntactic properties of interest in their Spanish.[11] Additionally, recall that the factorial design of our code-switching experiment controls to some extent for extra-grammatical effects like this one (see Schütze and Sprouse 2013; Sprouse et al. 2016 for discussion).

3.4. Code-Switching Results by Type

We turn now to the results of the code-switching experiment. We conducted two 2×2 comparisons because interpreting two-way interactions is more straightforward than interpreting three-way interactions, and we consider each comparison separately for ease of exposition. Thus, we first hold C constant as English *that* while varying the language of

T and the extracted argument in Section 3.4.1, then do the same while holding Spanish *que* constant (Section 3.4.2). In each case, we are testing whether a particular syntactic head is responsible for determining when extraction over an overt complementizer is acceptable. If a given head is responsible for the processes that ultimately drive the *that*-trace effect, then changing the language of that head should also change the relative acceptability of subject extraction.

3.4.1. Extraction over *That*

The first set of code-switching results we consider are those in which C is English *that*. Holding C constant, we have a 2 × 2 factorial design: language of T (English, Spanish) and wh-extraction (Subject, Object). Table 11 presents the mean raw ratings (scale = 1–7) and the estimated marginal mean z-scores of the ratings; Figure 3 plots the z-scores.

Table 11. CS experiment results: Extraction over *that*.

| Wh	T	Example	Mean Rating	EMM Z-Score
Object	Eng	*Qué asumieron los maestros* <u>that</u> *the child had read before the test?*	5.48	0.421
Object	Span	What did the teachers assume <u>that</u> *el niño había leído antes del examen*?	5.41	0.414
Subject	Eng	*Quién asumieron los maestros* <u>that</u> *had read the text before the test?*	4.38	−0.017
Subject	Span	Who did the teachers assume <u>that</u> *había leído el texto antes del examen*?	4.44	0.001
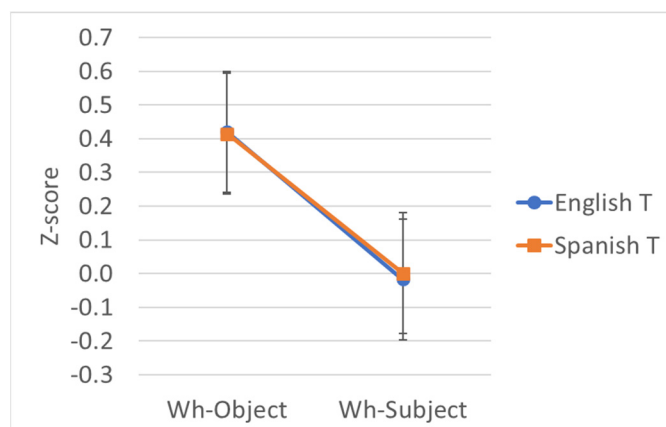
Figure 3. Estimated marginal mean z-scores by wh-extraction and language of T, code-switching data with C = *that* (error bars = 95% confidence intervals).

We fit a linear mixed-effects model to the data, as described in Section 3.1, and present the results in Table 12.

Table 12. Linear mixed-effects model, CS experiment: Extraction over *that*.

Fixed Effect	Coefficient (*β*)	SE	95% CI	*t*	*F*	*p*
T	0.01	0.09	−0.16–0.17	0.06	0.00	.949
Wh	0.43	0.09	0.25–0.60	4.94	24.41	<.001
Wh*T	−0.02	0.17	−0.37–0.32	−0.14	0.02	.888

Random Effects	Variance	SE
By-Subject intercept	0.006	0.019
By-Subject slope over Wh*T	0.094	0.037
By-Item intercept	0.034	0.024
By-Item slope over Wh*T	0.004	0.030

We found a significant main effect for wh-extraction, such that extracting a wh-subject was always worse than extracting an object. We do not find an effect for language of T, which indicates that the ratings for these sentences were not penalized for a code-switch between C and T, nor do we find an interaction, which suggests that there is no super-additive effect. As a measure of effect size, we calculated a *differences-in-differences* (DD) score of 0.025. The parallel results with Spanish and English T suggest that having T in Spanish is not sufficient to change the relative acceptability of subject extraction.

3.4.2. Extraction over *Que*

The second set of code-switching results we consider are those in which C is Spanish *que*. Holding C constant, we again have a 2 × 2 factorial design: language of T (English, Spanish) and wh-extraction (Subject, Object), Table 13 presents the mean raw ratings (scale = 1–7) and the estimated marginal mean z-scores of the ratings; Figure 4 plots the z-scores.

Table 13. CS experiment results: Extraction over *que*.

Wh	T	Example	Mean Rating	EMM Z-Score
Object	Eng	*Qué asumieron los maestros que* the child had read before the test?	4.93	0.203
Object	Span	What did the teachers assume *que el niño había leído antes del examen*?	5.26	0.368
Subject	Eng	*Quién asumieron los maestros que* had read the text before the test?	3.24	−0.501
Subject	Span	Who did the teachers assume *que había leído el texto antes del examen*?	5.93	0.647
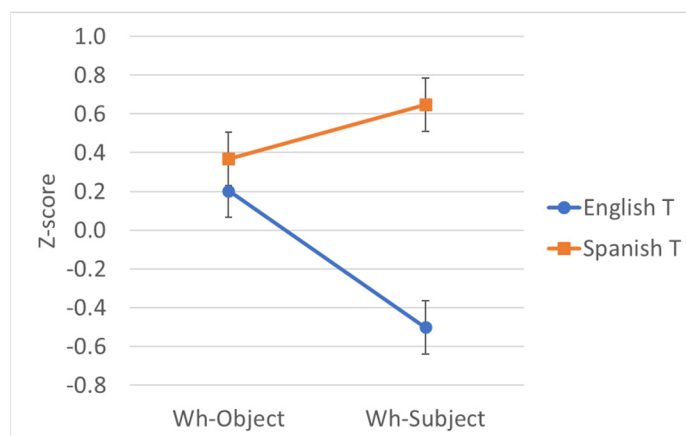
Figure 4. Estimated marginal mean z-scores by wh-extraction and language of T, code-switching data with C = *que* (error bars = 95% confidence intervals).

We fit a linear mixed-effects model to the data, as described in Section 3.1 and present the results in Table 14.

Table 14. Linear mixed-effects model, CS experiment: Extraction over *que*.

Fixed Effect	Coefficient (β)	SE	95% CI	t	F	p
T	0.66	0.07	0.53–0.79	9.96	99.17	<.001
Wh	0.21	0.07	0.08–0.35	3.16	9.95	.003
Wh*T	−0.98	0.13	−1.25−−0.72	−7.37	54.31	<.001
Random Effects			**Variance**		**SE**	
By-Subject intercept			0.005		0.011	
By-Item intercept			0.012		0.018	
By-Item slope over Wh			0.002		0.024	

All three fixed factors show significant effects, but the result of interest is the interaction, which indicates a super-additive effect. Bonferroni-corrected post hoc pairwise comparisons did not find a difference between English T and Spanish T when the extracted argument is an object ($p = .079$) but did find a difference when the extracted argument is a subject ($p < .001$). As a measure of effect size, we calculated a *differences-in-differences* (DD) score of 0.983. Notice that the interaction suggests that the complementizer alone does not change the relative acceptability of subject extraction.

3.4.3. Summary of Code-Switching Results

To visualize the code-switching data all together, we plotted the mean z-scores for extracting a wh-object (blue line with circles) and a wh-subject (orange line with squares) across the four possible combinations of C and T (*x*-axis) in Figure 5. On the same *y*-axis we also represent the difference between the score for object extraction and subject extraction with the gray bars, where a larger bar indicates a larger penalty for extracting the subject.
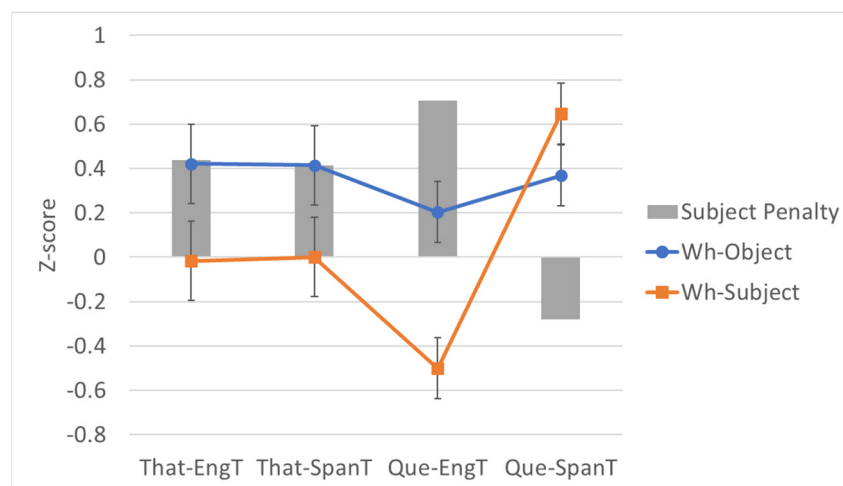
Figure 5. Comparison of subject extraction penalty across combinations of C and T, code-switching data (error bars = 95% confidence intervals).

Figure 5 makes clear that subject extraction is penalized in each of the combinations of C and T except when both C and T are in Spanish.

Additionally, compiling the results in Figure 5 shows—as an anonymous reviewer points out—that, while wh-subject extraction has low acceptability for both cases with English T, the largest subject penalty appears to come when C is *que* (12), not when it is *that* (13), suggesting an additional problem with extracting a subject over *que*, a sort of '*que*-trace effect.'

12. *Quién   asumieron   los   maestros   que*   had read the text before the test?
    who   assumed.3PL the   teachers   that
    'Who did the teachers assume that had read the text before the test?'

13. *Quién asumieron los maestros* that had read the text before the test?

Nevertheless, we also noticed an apparent general penalty for this switch between *que* and English T, which may well be an independent effect. To test whether the penalty was indeed greater for sentences like (12) than those like (13), we fit a linear mixed-effects model to the data for the four sentences with English T, making a 2×2 comparison across realization of C (*que*, *that*) and wh-extraction (Subject, Object); we present the results in Table 15.

Table 15. Linear mixed-effects model, CS experiment: Extraction with English T.

Fixed Effect	Coefficient (β)	SE	95% CI	t	F	p
C	−0.36	0.07	−0.50−−0.21	−4.93	24.34	<.001
Wh	0.59	0.07	0.44–0.73	8.10	65.59	<.001
Wh*C	0.26	0.14	−0.03–0.54	1.77	3.148	.077

Random Effects	Variance	SE
By-Subject intercept	0.014	0.015

The interaction, indicating a potential super-additive effect, is trending toward but fails to reach significance. As a measure of effect size, we calculated a *differences-in-differences* (DD) score of 0.256, which is smaller than the significant effects we discussed above. These borderline results are suggestive when paired with the visual inspection of the differences in Figure 5 but also warrant caution. We could be observing the effect of some independent, extra-grammatical fact not considered in our design; as described in Section 3.1, such confounds are endemic to acceptability judgment experiments. We therefore reserve judgment at present. We must note, however, that Chacón et al.'s (2015) findings that subject extraction over *que* was significantly worse than object extraction in monolingual Spanish add some support to the notion of a '*que*-trace effect,' and that we found identical results in our initial monolingual follow-up experiment, although, importantly, we did not find the same subject–object contrast with simpler sentences. Together with the borderline code-switching results, it seems clear that some mysteries remain regarding extraction over *que*, which could be fruitful avenues for future research.

4. Discussion

4.1. Empirical Findings

Before turning to our research questions, we would like to highlight one salient finding in our monolingual data: we find robust evidence for the classic *that*-trace effect in our participants' Midwestern US English (*contra* Sobin 1987), which further supports the emerging experimental consensus (Cowart and McDaniel 2021). On a similar note, our monolingual Spanish results, despite the task effect we explained in Section 3.3, demonstrate that the heritage speakers of Spanish who comprise our experimental group showed the same pattern of judgments as monolingual Spanish speakers, a fact which is potentially relevant to studies interested in bilingual divergence from the baseline input (see Montrul 2008; Polinsky 2018); here we find no such divergence.

Let us turn now to our first research question: What combinations of C and T permit subject extraction over an overt complementizer in Spanish–English code-switching? Recall

that we assume that the realization of C or T in a given language reflects a set of syntactic features. Consequently, if a given language permits subject extraction over an overt complementizer, then the syntactic head which matches that language must be the head with the relevant features for permitting subject extraction in that context.

We began by comparing subject and object extraction over the overt English complementizer *that*. We found that subject extraction was always worse than object extraction, whether T was in English or in Spanish, exactly parallel to the contrast observed in English. The acceptability of subject extraction did not improve with Spanish T, suggesting that Spanish T alone is not sufficient to obviate the *that*-trace effect. We then compared subject and object extraction over the overt Spanish complementizer *que*. Here, we found a contrast between English and Spanish T. When T was in English, subject extraction was significantly less acceptable than object extraction despite the complementizer being in Spanish, while subject extraction with Spanish T was more acceptable than object extraction.

Together, these results suggest that subject extraction over an overt complementizer only becomes acceptable when both C and T are in Spanish. Our finding aligns with evidence suggesting that null subjects (Sande 2018), post-verbal subjects (Ebert and Hoot 2018; Hoot and Ebert 2021), and possibly V2 (Vanden Wyngaerd 2020) are also conditioned by C and T together, pointing away from the proposal that the phase head determines the properties of its complement in code-switching (González-Vilbazo and López 2011, 2012, 2013). That said, perhaps González-Vilbazo and López's Phase Head Hypothesis could be incorporated into an account of our findings if it were coupled with Feature Inheritance (Chomsky 2008), under which T inherits (some of) its features from C, meaning that C would still be fundamentally running the show. Vanden Wyngaerd (2020) suggested something along these lines, as did we in an earlier study (Ebert and Hoot 2018), although we subsequently examined the idea in more detail and found the evidence mixed (Hoot and Ebert 2021). One barrier to applying Feature Inheritance to our data is that the field has not yet settled on its precise details, making its extension to code-switching and to experimental work challenging (but see Shim 2013, 2016). Nevertheless, our data clearly point to a tight relationship between C and T that is at least suggestive of an interaction between the two heads in need of further explanation.

Overall, the picture emerging from the code-switching evidence aligns with the cross-linguistic evidence showing a correlation between post-verbal or null subjects and the lack of the *that*-trace effect. Such a finding was by no means preordained; it was certainly possible to imagine that subject position might involve some feature of T (like the EPP or 'rich agreement') while the *that*-trace effect might come down to a feature of C. Yet the pattern of judgments in both the present study and our previous experiment (Hoot and Ebert 2021) are exactly parallel. To our eyes, this is a striking piece of evidence in favor of the link between the *that*-trace effect and a particular subject position.

4.2. Testing Theoretical Accounts

We now turn to our second research question: Which theoretical proposals are supported by the code-switching evidence? Table 16 compares our findings to the predictions of each of the five theoretical accounts we tested.

Table 16. Support for accounts of the *that*-trace effect.

Account	Basic Claim	Spanish Is Different Because . . .	Ban on Extraction from Spec-TP Is . . .	Prediction for CS	Account Supported?
Anti-Locality (Douglas 2017; Erlewine 2016, 2020)	Movement from Spec-TP to Spec-CP is too short; extraction from Spec-TP thus universally barred.	It does not have the EPP; Spanish allows extraction from post-verbal position, so movement is not too short.	Universal.	Extraction only from post-verbal position in CS, so whatever determines subject position (C + T?) determines extraction in CS.	Yes. Recent experimental work suggest C and T together permit post-verbal subjects.
Criterial Freezing (Rizzi 2006, 2015; Rizzi and Shlonsky 2007)	Positions with interpretive properties (like subjects) are frozen; extraction from Spec-TP thus universally barred.	Null expletive fills the subject position in Spanish, so subject can be extracted from lower position.	Universal.	Null expletives permit extraction, so whatever determines null subject availability (C + T?) determines extraction in CS.	Yes. Recent experimental work suggest C and T together permit null subjects.
Prosodic Alignment (Kandybowicz 2006, 2009; McFadden and Sundaresan 2018; Sato and Dobashi 2016)	Empty Spec-TP cannot align with left edge of intonational phrase (or cannot form phrase with C) so syntax/prosody matching fails; extraction from Spec-TP thus universally barred.	V-to-T movement means V is highest head in intonational phrase and therefore at left edge, which is thus not empty.	Universal.	Assuming V-to-T is a property of T, language of T determines CS behavior.	No. T alone does not obviate the *that*-trace effect.
Labeling (Chomsky 2013, 2015)	T is deficient and cannot be labeled alone; can only be labeled with subject, so subject is frozen.	T is not deficient (because of 'rich agreement').	Language specific.	Language of T determines CS behavior.	No. T alone does not obviate the *that*-trace effect.
T-to-C (Pesetsky and Torrego 2001)	T raised to C surfaces as *that*; extracting a subject is more economical and blocks T raising, so *that*-t.	Spanish C is a true complementizer, not an instance of T in C.	Language specific.	Language of C determines CS behavior.	No. C alone does not obviate the *that*-trace effect.

We find the most support for the Anti-Locality and Criterial Freezing accounts, both of which posit (a) a universal syntactic prohibition on extraction from Spec-TP and (b) that cross-linguistic differences emerge from other processes which can avoid that position. However, our study only addresses the second of those claims. Although we find a correlation between the availability of the *that*-trace effect and previous evidence from post-verbal/null subjects, as these accounts predict, our evidence does not allow us to test the *reason* subjects cannot be extracted from Spec-TP—either because the movement to do so is too short (Anti-Locality) or because anything in Spec-TP is frozen (Criterial Freezing).

The Prosodic Alignment account with testable predictions for Spanish (McFadden and Sundaresan 2018) similarly posits a universal prohibition on subject extraction (because it would create an illicit prosodic structure). Under the assumption that the syntax/prosody mismatch is avoided in Spanish because V-to-T movement ensures the relevant prosodic structure is not empty, we predicted that the language of T would correspond to the availability of subject extraction: Spanish verbs fill the T head, and therefore fill the relevant prosodic structure. However, we found that Spanish T was not enough to license subject extraction. Nevertheless, different assumptions about cross-linguistic variation under a prosodic account might produce different predictions for code-switching, so our results do not necessarily rule out this family of approaches. Indeed, research on the prosody of code-switching is still quite limited, and it would be valuable to examine direct intonation evidence to truly put prosodic accounts to the test.

Finally, we do not find support for either proposal (Labeling or T-to-C) identifying language-specific features of either C or T alone as the barrier to subject extraction.

4.3. Narrowing the Range of Hypotheses

In addition to testing specific theoretical accounts, we can take a step back and ask whether our results can also serve to narrow the range of *possible* approaches to accounting for the *that*-trace effect. What distribution of syntactic features[12] would be, in principle, compatible with our findings?

To explain the difference between Spanish and English, we must presumably either posit (a) some feature in English that *blocks* subject extraction or (b) some feature in Spanish that *allows* subject extraction. If we take our results to show that the relevant features involve both C and T, we thus have two possibilities. If the relevant syntactic features *ban* subject extraction, as in English, then those features must be able to appear on either head, since English C or English T alone was enough to prevent subject extraction. This implies some redundancy, since the presence of either head is sufficient to ban extraction. If, instead, the nature of the features *permits* subject extraction, as in Spanish, they must be such that their effect is only visible when both heads appear together, since extraction was permitted only when both heads were in Spanish.

If our suggestion is on the right track, future accounts of the *that*-trace effect need to explain either how C and T in Spanish work together to permit subject extraction (and why neither can do it alone, as well as why it is generally banned otherwise) or how C and T in English can each separately bring about a ban on subject extraction (and why such a redundancy exists). The same explanation is needed for cross-linguistic variation in subject position and null subjects, which, as we have seen, are clearly related.

5. Conclusions

Our results suggest that subject extraction over an overt complementizer is only possible in Spanish–English code-switching when both C and T are in Spanish. This finding is in line with previous work in code-switching that has found that both C and T play a role in determining subject position as well. We therefore provide novel evidence in favor of a connection between a specific, pre-verbal subject position and a ban on extraction. Although such a connection is already accepted by many theorists, it is noteworthy to find further empirical confirmation of it from a heretofore unexplored data source.

Our second contribution is that we put the predictions of existing theoretical accounts to the test directly. We found some indirect support for those accounts that postulate a universal restriction on extraction from Spec-TP, but, of course, more research is needed, and code-switching experiments could continue to test these theories. For example, Anti-locality accounts could be further tested by repeating our experiment with an additional manipulation of the internal structure of CP to produce shorter or longer movement opportunities. Code-switching sentences that include different sizes of CP (as with recomplementation or topicalization structures in Spanish, for instance), could perhaps more directly test the claim that the barrier to movement is due to anti-locality. Moreover, it would be valuable to test different language pairs with the same paradigm, perhaps drawing on Gilligan's (1987) typology of languages to find combinations of features not yet tested. Finally, work directly measuring prosody is clearly needed.

Our final contribution is to recommend some ways to narrow the range of possible future accounts. The pattern of acceptability in code-switching suggests one of two possible paths for syntactic accounts of the *that*-trace effect and its cross-linguistic variation. This implication of our results can serve to help future theorists constrain the set of possible hypotheses to explain the *that*-trace effect. In this way, we hope also to have demonstrated the utility of applying code-switching data to core questions in linguistic theory.

Appendix A. Follow-Up Monolingual Results

To contextualize the findings reported in Section 3.3, we conducted a follow-up study with 20 monolingual speakers of Mexican Spanish. We chose Mexican Spanish because it is by far the most common variety in our area (as in much of the US), so we expected it to be closest to the variety to which our bilinguals were exposed. We recruited (via the Internet) 20 participants born and living in Mexico who reported being monolingual Spanish speakers and having graduated high school. Their mean age was 29.5 (SD 11.7, range 18–58). Other than the participants, everything about the experiment was the same as for the bilingual group.

The results are displayed in Table A1 and Figure A1. Results of the LMM are in Table A2.

Table A1. Monolingual Spanish experiment redux with monolingual speakers.

| Wh	C	Example	Mean Rating	EMM Z-Score
Object	*Que*	¿**Qué** confirmaron las editoras que el autor había escrito en tan solo una semana?	4.84	0.175
Object	Null	¿**Qué** confirmaron las editoras ___ el autor había escrito en tan solo una semana?	2.70	−0.701
Subject	*Que*	¿**Quién** confirmaron las editoras que había escrito el libro en tan solo una semana?	2.88	−0.630
Subject	Null	¿**Quién** confirmaron las editoras ___ había escrito el libro en tan solo una semana?	2.59	−0.734
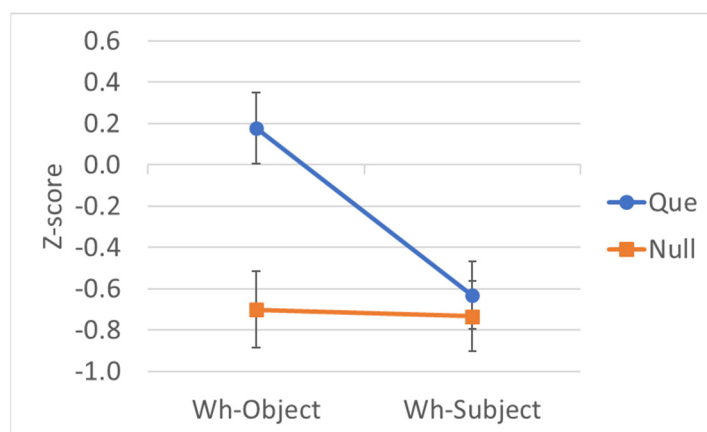
Figure A1. Estimated marginal mean z-scores by wh-extraction and realization of C, Spanish monolingual data redux, judged by monolingual speakers (error bars = 95% confidence intervals).

Table A2. Linear mixed-effects model, monolingual Spanish experiment redux with monolingual speakers.

Fixed Effect	Coefficient (β)	SE	95% CI	t	F	p
C	0.49	0.10	0.27–0.71	4.69	21.98	<.001
Wh	0.42	0.10	0.20–0.63	4.00	16.01	<.001
Wh*C	0.77	0.21	0.35–1.19	3.75	14.04	.001

Random Effects	Variance	SE
By-Subject intercept	0.002	0.026
By-Subject slope over C	0.003	0.043
By-Subject slope over Wh*C	0.083	0.050
By-Item intercept	0.009	0.022
By-Item slope over Wh	0.006	0.039
By-Item slope over Wh*C	0.003	0.046

It is clear that these speakers gave consistently low ratings for omitting *que* and puzzlingly low ratings of subject extraction over *que*, in line with what Chacón et al. (2015) found, and the same as the results for our bilingual group.

Notes

[1] Some may call this projection Infl, and some posit projections just for subject agreement, like AgrS or Rizzi's Subj projection. We have chosen to always refer to the highest verbal projection as T and the subject position as Spec-TP for the sake of consistency.

[2] Post-verbal subjects also widely correlate with null subjects, so it could be that it is not the availability of *post-verbal* subjects that is crucial, but rather the availability of null subjects, especially of particular types. Nicolis (2008) presents evidence from eight creole languages suggesting that the crucial feature is the existence of null expletives in the language, rather than subject position per se. Nonetheless, the essential insight is the same—something allows the extraction to proceed from post-verbal position—and for Spanish (like Italian) that extraction may well proceed from a post-verbal subject position that exists independently, which is why we focus our description on that fact.

[3] As mentioned previously, Rizzi's system has a dedicated functional head Subj that hosts subjects in its specifier, distinct from TP, but we will continue to refer to Spec-TP as the position of all subjects for consistency.

[4] Nicolis (2008) presents evidence that what matters is whether null *expletives*, specifically, are possible, not pro-drop more generally, but it seems fair to assume they go together in the case of Spanish.

[5] An anonymous reviewer points out that this same reasoning cannot be directly applied to their account of English, however. If we assume an English auxiliary occupies the T head, the notion that a phonetically filled T renders extraction possible predicts that **Who do you think that will leave?* or **Who do you think that has left?* should be grammatical, contrary to fact. McFadden and Sundaresan do not consider this point, but perhaps they could address it by following Sato and Dobashi (2016) in appealing to Truckenbrodt's (1999) Lexical Category Condition. Under this view, even an English T head filled with an overt auxiliary might be invisible to PF, whereas Spanish T would also contain the lexical verb due to V-to-T raising, yielding the same cross-linguistic

contrast in sentences with and without auxiliaries. We will tentatively assume this contrast holds for the sake of the present work, but we agree more research into these details would be valuable. Intriguingly, some evidence from Dutch (den Dikken 2007) shows that the linear order of the auxiliary and verb (T-V vs. V-T) can affect the availability of extraction, suggesting that even within a single language there could be some effect worth exploring of the contents of the T head.

[6]   We choose a relatively low minimum proficiency in part because this task penalizes incorrectly guessing 'yes' on a false word, and heritage speakers display a well-known '*yes*-bias;' they are reluctant to reject unfamiliar linguistic structures due to linguistic insecurity (Polinsky 2018), which pushes their scores on this task lower than they might otherwise be.

[7]   Spanish examples are direct translations of the English sentences in Table 3 which is why they are not glossed.

[8]   All the code-switched examples are versions of the sentences *What did the teachers assume that the child had read before the test?* or *Who did the teachers assume that had read the text before the test?*. By convention, one language (here, Spanish) is italicized for readers' convenience, but nothing was italicized for participants.

[9]   The same experiments found that sentences with post-verbal Spanish subjects, as in (i), were rated very low when T was in English, further suggesting that subject position has more to do with T than with features of the subject itself.

(i)   *No   recuerdo   cuánto   dinero*   have stolen   *esos   delincuentes*   in the last decade.
      no   remember.1SG   how.much   money   those   criminals
      'I do not remember how much money those criminals have stolen in the last decade.'

[10] An anonymous reviewer suggests a clever way to test our conjecture: many parentheticals cannot take negative quantifiers as subjects. For example, Collins and Postal (2014, p. 196) note that sentences like **Cathy will not, nobody asserted/proved/reported/said/wrote, divorce Frank* are not possible, while it is certainly possible for a negative quantifier to be a matrix clause subject, as in *Nobody asserted Cathy will not divorce Frank*. Assuming the same holds for Spanish, we could potentially force a non-parenthetical reading by including a negative quantifier, as in **¿Quién no confirmó nadie ___ había escrito el libro en tan solo una semana?* 'Who did no one confirm had written the book in only one week?'. Because the parenthetical reading is not available, this sentence must be interpreted as subordination with a missing complementizer, our intended reading. We agree this would be a valuable follow-up study, but for the moment we must leave it for future research.

[11] Further evidence for this interpretation comes from the code-switching results themselves. If these speakers in fact have a *that*-trace effect in their Spanish just like in English, then we would expect to observe a subject/object asymmetry in all code-switching sentences as well, since both languages would be the same. Instead, we see in Section 3.4 that, when C and T are both in Spanish, subject extraction is at least as acceptable as object extraction. This finding would be puzzling if their Spanish truly prohibited subject extraction, suggesting that the code-switching evidence indeed surfaces the relevant contrast between Spanish and English.

[12] Importantly, syntactic features still play a role even for accounts that ultimately attribute the *that*-trace effect to prosodic requirements. Because the syntactic features of a given sentence determine its structure and that syntactic structure must be mapped onto a prosodic structure, those features necessarily play a role in the prosodic structure of the sentence.

References

Alexiadou, Artemis, and Elena Anagnostopoulou. 1998. Parametrizing AGR: Word Order, V-Movement and EPP-Checking. *Natural Language & Linguistic Theory* 16: 491–539. [CrossRef]

Bader, Markus, and Jana Häussler. 2010. Toward a model of grammaticality judgments. *Journal of Linguistics* 46: 273–330. [CrossRef]

Badiola, Lucia, Rodrigo Delgado, Ariane Sande, and Sara Stefanich. 2018. Code-switching attitudes and their effects on acceptability judgment tasks. *Linguistic Approaches to Bilingualism* 8: 5–24. [CrossRef]

Barbosa, Pilar P. 2011. Pro-drop and Theories of pro in the Minimalist Program Part 1: Consistent Null Subject Languages and the Pronominal-Agr Hypothesis: Null Subject Languages and the Pronominal-Agr Hypothesis. *Language and Linguistics Compass* 5: 551–70. [CrossRef]

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]

Bošković, Željko. 2016. On the timing of labeling: Deducing Comp-trace effects, the Subject Condition, the Adjunct Condition, and tucking in from labeling. *The Linguistic Review* 33: 17–66. [CrossRef]

Brandi, Luciana, and Patrizia Cordin. 1989. Two Italian dialects and the null subject parameter. In *The Null Subject Parameter*. Studies in Natural Language and Linguistic Theory 15. Edited by Osvaldo Jaeggli and Kenneth J. Safir. Dordrecht: Kluwer, pp. 111–42. [CrossRef]

Bresnan, Joan. 1977. Variables in the theory of transformations. In *Formal Syntax*. Edited by Peter W. Culicover, Thomas Wasow and Adrian Akmajian. New York: Academic Press, pp. 157–96.

Camacho, José. 2013. *Null Subjects*. Cambridge: Cambridge University Press.

Chacón, Dustin Alfonso, Michael Fetters, Margaret Kandel, Eric Pelzl, and Colin Phillips. 2015. Indirect learning and language variation: Reassessing the that-trace effect. Unpublished manuscript.

Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge: MIT Press.

Chomsky, Noam. 2008. On phases. In *Foundational Issues in Linguistic Theory: Essays in Honor of Jean-Roger Vergnaud*. Edited by Robert Freidin, Carlos Otero and Maria Luisa Zubizarreta. Cambridge: MIT Press, pp. 134–66.

Chomsky, Noam. 2013. Problems of projection. *Lingua* 130: 33–49. [CrossRef]

Chomsky, Noam. 2015. Problems of projection: Extensions. In *Structures, Strategies, and Beyond*. Edited by Elisa Di Domenico, Cornelia Hamann and Simona Matteini. Amsterdam: John Benjamins, pp. 1–16. [CrossRef]

Collins, Chris, and Paul M. Postal. 2014. *Classical NEG Raising: An Essay on the Syntax of Negation*. Cambridge: MIT Press.

Cowart, Wayne, and Dana McDaniel. 2021. The that-trace effect. In *The Cambridge Handbook of Experimental Syntax*. Edited by Grant Goodall. Cambridge: Cambridge University Press, pp. 258–77.

Davies, Mark. 2006. *A Frequency Dictionary of Spanish: Core Vocabulary for Learners*. New York and London: Routledge.

Davies, Mark. 2008. The Corpus of Contemporary American English (COCA): 520 Million Words, 1990-Present. Available online: http://corpus.byu.edu/coca/ (accessed on 6 May 2018).

den Dikken, Marcel. 2007. Questionnaire Study on Dutch *That*-Trace Effects: Stimuli and Results. *Unpublished Manuscript*. Available online: https://www.gc.cuny.edu/CUNY_GC/media/CUNY-Graduate-Center/PDF/Programs/Linguistics/Dikken/dutch_that_trace_results.pdf (accessed on 19 October 2021).

Douglas, Jamie. 2017. Unifying the that-trace and anti-that-trace effects. *Glossa: A Journal of General Linguistics* 2: 60. [CrossRef]

Drummond, Alex. 2017. Internet Based EXperiments (Ibex). Available online: http://spellout.net/ibexfarm/ (accessed on 8 September 2020).

Ebert, Shane. 2014. *The Morphosyntax of Wh-Questions: Evidence from Spanish-English Code-Switching*. Chicago: University of Illinois at Chicago Dissertation.

Ebert, Shane, and Bradley Hoot. 2018. *That*-trace effects in Spanish-English code-switching. In *Code-Switching—Experimental Answers to Theoretical Questions: In Honor of Kay González-Vilbazo*. Edited by Luis López. Issues in Hispanic and Lusophone Linguistics 19. Amsterdam: John Benjamins, pp. 101–45. [CrossRef]

Ebert, Shane, and Bryan Koronkiewicz. 2018. Monolingual stimuli as a foundation for analyzing code-switching data. *Linguistic Approaches to Bilingualism* 8: 25–66. [CrossRef]

Erlewine, Michael Yoshitaka. 2016. Anti-locality and optimality in Kaqchikel Agent Focus. *Natural Language & Linguistic Theory* 34: 429–79. [CrossRef]

Erlewine, Michael Yoshitaka. 2020. Anti-locality and subject extraction. *Glossa: A Journal of General Linguistics* 5: 84. [CrossRef]

Gilligan, Gary Martin. 1987. *A Cross-Linguistic Approach to the Pro-Drop Parameter*. Los Angeles: University of Southern California Dissertation. Available online: http://www.proquest.com/docview/1664339268/citation/627F44948625453CPQ/1 (accessed on 24 August 2021).

González-Vilbazo, Kay, Laura Bartlett, Sarah Downey, Shane Ebert, Jeanne Heil, Bradley Hoot, Bryan Koronkiewicz, and Sergio Ramos. 2013. Methodological Considerations in Code-Switching Research. *Studies in Hispanic & Lusophone Linguistics* 6: 119–38.

González-Vilbazo, Kay, and Luis López. 2011. Some properties of light verbs in code-switching. *Lingua* 121: 832–50. [CrossRef]

González-Vilbazo, Kay, and Luis López. 2012. Little v and parametric variation. *Natural Language & Linguistic Theory* 30: 33–77. [CrossRef]

González-Vilbazo, Kay, and Luis López. 2013. Phase switching. Paper presented at the Code-Switching in the Bilingual Child: Within and across the Clause, Center for General Linguistics (ZAS), Berlin, Germany, May 14.

Grosjean, François. 1985. The bilingual as a competent but specific speaker-hearer. *Journal of Multilingual and Multicultural Development* 6: 467–77. [CrossRef]

Hoot, Bradley, and Shane Ebert. 2021. On the position of subjects in Spanish: Evidence from code-switching. *Glossa: A journal of general linguistics* 6: 73. [CrossRef]

Izura, Cristina, Fernando Cuetos, and Marc Brysbaert. 2014. Lextale-Esp: A test to rapidly and efficiently assess the Spanish vocabulary size. *Psicologica: International Journal of Methodology and Experimental Psychology* 35: 49–66.

Juzek, Tom. 2016. *Acceptability Judgement Tasks and Grammatical Theory*. Oxford: University of Oxford Dissertation.

Juzek, Tom, and Jana Häussler. 2015. Non-cooperative behaviour & acceptability judgement tasks. Paper presented at the Methods and Linguistic Theories Symposium, Bamberg, Germany, November 27–28.

Kandybowicz, Jason. 2006. Comp-trace effects explained away. In *Proceedings of the 25th West Coast Conference on Formal Linguistics*. Edited by Donald Baumer, David Montero and Michael Scanlon. Somerville: Cascadilla Proceedings Project, pp. 220–28.

Kandybowicz, Jason. 2009. Embracing edges: Syntactic and phono-syntactic edge sensitivity in Nupe. *Natural Language & Linguistic Theory* 27: 305–44. [CrossRef]

Kenstowicz, Michael. 1989. The null subject parameter in modern Arabic dialects. In *The Null Subject Parameter*. Studies in Natural Language and Linguistic Theory 15. Edited by Osvaldo Jaeggli and Kenneth J. Safir. Dordrecht: Kluwer, pp. 263–75. [CrossRef]

Kiziak, Tanja. 2010. *Extraction Asymmetries: Experimental evidence from German*. Amsterdam: John Benjamins.

Larson-Hall, Jenifer. 2010. *A Guide to Doing Statistics in Second Language Research Using SPSS*. New York and London: Routledge.

López, Luis. 2020. *Bilingual Grammar: Toward an Integrated Model*. Cambridge: Cambridge University Press.

MacSwan, Jeff. 1999. *A Minimalist Approach to Intra-Sentential Code-Switching*. New York: Garland.

MacSwan, Jeff. 2004. Code-Switching and Grammatical Theory. In *The Handbook of BILINGUALISM*, 1st ed. Edited by Tej K. Bhatia and William C. Ritchie. Malden: Blackwell, pp. 415–62.

MacSwan, Jeff. 2013. Code-Switching and Grammatical Theory. In *The Handbook of Bilingualism and Multilingualism*, 2nd ed. Edited by Tej K. Bhatia and William C. Ritchie. Chichester and Malden: Wiley-Blackwell, pp. 321–50.

McFadden, Thomas, and Sandhya Sundaresan. 2018. What the EPP and comp-trace effects have in common: Constraining silent elements at the edge. *Glossa: A Journal of General Linguistics* 3: 43.1–43.34. [CrossRef]

Meteyard, Lotte, and Robert A. I. Davies. 2020. Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language* 112: 104092. [CrossRef]

Montrul, Silvina. 2008. *Incomplete Acquisition in Bilingualism: Re-Examining the Age Factor*. Studies in Bilingualism 39. Amsterdam: John Benjamins.

Nicolis, Marco. 2008. The null subject parameter and correlating properties: The case of Creole languages. In *The Limits of Syntactic Variation*. Edited by Theresa Biberauer. Linguistik Aktuell/Linguistics Today 132. Amsterdam: John Benjamins, vol. 132, pp. 271–94. [CrossRef]

Perlmutter, David. 1968. *Deep and Surface Structure Constraints in Syntax*. Cambridge: MIT Dissertation. Available online: http://hdl.handle.net/1721.1/13003 (accessed on 17 August 2017).

Perlmutter, David. 1971. *Deep and Surface Structure Constraints in Syntax*. New York: Holt, Reinhart, and Wilson.

Pesetsky, David. 2017. Complementizer-trace effects. In *The Wiley Blackwell Companion to Syntax*, 2nd ed. Edited by Martin Everaert and Henk van Riemsdijk. London: Wiley Blackwell.

Pesetsky, David, and Esther Torrego. 2001. T-to-C movement: Causes and consequences. In *Ken Hale: A Life in Language*. Edited by Michael Kenstowicz. Cambridge: MIT Press.

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

Reis, Marga. 1996. Extractions from verb-second clauses in German? In *On Extraction and Extraposition in German*. Linguistik Aktuell/Linguistics Today 11. Edited by Uli Lutz and Jürgen Pafel. Amsterdam: John Benjamins, pp. 45–88. [CrossRef]

Ritchart, Amanda, Grant Goodall, and Marc Garellek. 2016. Prosody and the that-trace effect: An experimental study. In *Proceedings of the 33rd West Coast Conference on Formal Linguistics*. Edited by Kyeong-min Kim, Pocholo Umbal, Trevor Block, Queenie Chan, Tanie Cheng, Kelli Finney, Mara Katz, Sophie Nickel-Thompson and Lisa Shorten. Somerville: Cascadilla Proceedings Project, pp. 320–28. Available online: http://www.lingref.com/cpp/wccfl/33/paper3251.pdf (accessed on 19 October 2021).

Rizzi, Luigi. 1982. *Issues in Italian Syntax*. Studies in Generative Grammar 11. Dordrecht: Foris.

Rizzi, Luigi. 2006. On the form of chains: Criterial positions and ECP effects. In *Wh-Movement: Moving on*. Edited by Lisa Lai Shen Chen and Norbert Corver. Current Studies in Linguistics 42. Cambridge: MIT Press, pp. 97–134.

Rizzi, Luigi. 2015. Notes on labeling and subject positions. In *Structures, Strategies and Beyond: Studies in Honour of Adriana Belletti*. Edited by Elisa Di Domenico, Cornelia Hamann and Simona Matteini. Linguistik Aktuell/Linguistics Today 223. Amsterdam: John Benjamins, pp. 17–46. [CrossRef]

Rizzi, Luigi, and Ur Shlonsky. 2007. Strategies of subject extraction. In *Interfaces + Recursion = Language? Chomsky's Minimalism and the View from Syntax-Semantics*. Edited by Uli Sauerland and Hans-Martin Gärtner. Studies in Generative Grammar [SGG] 89. Berlin: Mouton de Gruyter, pp. 115–60.

Rothman, Jason, and Jeanine Treffers-Daller. 2014. A prolegomenon to the construct of the native speaker: Heritage speaker bilinguals are natives too! *Applied Linguistics* 35: 93–98. [CrossRef]

Sande, Ariane. 2018. *C Plus T as a Necessary Condition for Pro-Drop: Evidence from Code-Switching*. Chicago: University of Illinois at Chicago Dissertation.

Sato, Yosuke, and Yoshihito Dobashi. 2016. Prosodic Phrasing and the *That*-Trace Effect. *Linguistic Inquiry* 47: 333–49. [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.

Shim, Ji Young. 2013. *Deriving Word Order in Code-Switching: Feature Inheritance and Light Verbs*. New York: City University of New York Dissertation.

Shim, Ji Young. 2016. Mixed verbs in code-switching: The syntax of light verbs. *Languages* 1: 8. [CrossRef]

Sobin, Nicholas. 1987. The variable status of COMP-trace phenomena. *Natural Language & Linguistic Theory* 5: 33–60. [CrossRef]

Sprouse, Jon. 2011. A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior Research Methods* 43: 155–67. [CrossRef]

Sprouse, Jon, Ivano Caponigro, Ciro Greco, and Carlo Cecchetto. 2016. Experimental syntax and the variation of island effects in English and Italian. *Natural Language & Linguistic Theory* 34: 307–44. [CrossRef]

Toquero-Pérez, Luis Miguel. 2021. *That*-Trace Effects: Haplology Is the Answer. Unpublished Manuscript; lingbuzz/005548. Available online: https://ling.auf.net/lingbuzz/005548 (accessed on 8 September 2021).

Truckenbrodt, Hubert. 1999. On the relation between syntactic phrases and phonological phrases. *Linguistic Inquiry* 30: 219–55. [CrossRef]

Vanden Wyngaerd, Emma. 2020. C0 and Dutch-English code-switching. *Ampersand* 7: 100060. [CrossRef]

Villa-García, Julio. 2018. Properties of the extended verb phrase: Agreement, the structure of INFL, and subjects. In *The Cambridge Handbook of Spanish Linguistics*. Edited by Kimberly L. Geeslin. Cambridge: Cambridge University Press, pp. 329–50. [CrossRef]

Woolford, Ellen. 1983. Bilingual code-switching and syntactic theory. *Linguistic Inquiry* 14: 520–36. [CrossRef]