Resumption ameliorates different islands differentially:

Acceptability data from Modern Standard Arabic[0]

Matthew A. Tucker[†,‡]**, Ali Idrissi**[§]**, Jon Sprouse**[**] **& Diogo Almeida**[†]

[†]Psychology Program, NYU Abu Dhabi

[‡]Department of Linguistics, Oakland University

[§]Department of English Literature and Linguistics, Qatar University

[**]Department of Linguistics, University of Connecticut

Abstract

Two acceptability experiments are presented which assess whether resumptive pronouns freely alternate with gaps and/or ameliorate island violation effects in *wh*-questions in Modern Standard Arabic. Both experiments test Complex Noun Phrase Constraint violations, adjunct island violations, and *whether* island violations. The results indicate that resumption is largely only acceptable with structurally complex DP fillers (*which*-NP) and is generally preferred to gapped structures in long-distance dependencies. Resumption is also shown to ameliorate the grammatical component of some island violations (adjunct and *whether* islands), but in different quantitative amounts across different islands. The overall picture which emerges is one in which resumption is quantitatively, but perhaps not qualitatively, helpful in repairing grammatical constraint violations in Modern Standard Arabic.

[0]Thanks to Souad Al Helou for assistance with stimuli creation and experimental procedure. Thanks to Salam Khalifa and Anas Shahrour for help with stimuli preparation. Thanks to the members of the United Arab Emirates University, Tommi Leung, and `arabic-L` email list for assistance with participant recruitment. Thanks to Jim McCloskey, Dustin Alfonso Chacón, two anonymous reviewers, and audiences at the 30th Annual Symposium on Arabic Linguistics and University of Minnesota for comments on previous versions of this work. All remaining faults are ours alone.

1 Introduction

A hallmark of natural language is the ability for sentences to contain relationships which hold between words at a potentially infinite distance, as in the case of the English *wh*-question in (1):[1]

(1) *Which sabre fencer*$_i$ was US Olympian and New York City resident Daryl Homer said by the commentators to be widely expected to beat t_i at the 2016 Olympics?

In (1), the syntactic and interpretative relationship between the italicized *wh*-FILLER and the embedded verb *beat* licenses the apparent displacement of *beat*'s direct object and in interpretation of the filler (*which sabre fencer*) in the GAP position after *beat*. Most notably for theories of formal linguistics and language processing, this *wh*-DEPENDENCY can obtain over a potentially infinite distance as the pattern in (1) makes clear (see, *e.g.*, Ross, 1967). This potentially infinite distance and the ubiquity of *wh*-dependencies cross-linguistically makes their study central to answering questions about the relationship between grammar and parsing.

1.1 The Grammatical Status of Resumption

While most languages share the property of having overt long-distance dependencies, one dimension along which languages vary concerns whether there is widespread availability of another, related construction involving a RESUMPTIVE PRONOUN (for an overview, see McCloskey, 2006 and references therein). Examples of resumptive pronouns appear in (2) for both English and a spoken variety

[1] Throughout this paper we use *italics* to indicate *wh*-fillers, **bold** to indicate resumptive pronouns, and *t* with coindexing to indicate gap positions of filler-gap dependencies. The use of this notation should not be construed as implicit support for a particular formal theory of the gap position (*e.g.,* traces, silent copies, deleted NPs, *etc.*).

of Arabic (Levantine):[2]

(2) a. That's the kind of murderer *who* Columbo usually catches **them** before the end of the episode.

b. *miin/ʔajja$_i$ mariid^ɛ* zeerit-**o**$_i$ naadia.
who/which patient$_i$ visited.3FS-**him**$_i$ Naadia
"*Who/which patient* did Nadia visit (**him**$_i$)?"
(Aoun, Benmamoun, & Choueiri, 2010:132)

Following McCloskey (2006), we take a resumptive pronoun to be any pronoun obligatorily coreferent with a sentence-internal filler which appears in the tail position of a filler-gap dependency.

Resumption is famously known to vary cross-linguistically in the extent to which it is accepted in distinct grammatical constructions and in distinct languages. While theoretical formulations of this distinction differ, a common strand of thought has taken languages to fall into two classes: (i) LIMITED resumption languages where resumption is generally dispreferred and appears only in limited grammatical constructions (such as English and Brazilian Portuguese) and (ii) PERVASIVE resumption languages where resumption is much more pervasive and even required in some contexts (such as Irish, Hebrew, or Arabic).[3]

While English is often touted as an example of a limited resumption language based upon off-line judgments, the picture is not obviously so clear-cut as grammaticality judgments would lead one to believe. Resumptive pronouns were noticed to be somewhat widespread in spoken English by Kroch (1981) and Prince (1990). Corpus studies of English also suggest a higher rate of resumption that might have been expected by this categorization of English as a limited

[2]In this paper we use the following abbreviations in glossing: 1/2/3 = 1st/2nd/3rd person, M/F = masculine/feminine gender, S/P = singular/plural number, COMP = complementizer, NEG = negation.

[3]Crucially, here we will not distinguish between so-called "intrusive" and "grammaticalized" pronouns based upon the availability of bound readings (Chao & Sells, 1983; Sells, 1984) or "apparent" and "true" resumption based upon the availability of reconstruction (Aoun, Choueiri, & Hornstein, 2001; Aoun et al., 2010; see also the distinction between "Class 1" and "Class 2" languages from Dermidache, 1991). This is done because we will ultimately suggest that this distinction is better viewed as a quantitative, not qualitative, one.

resumption language (Bennett, 2008). Perhaps even more perplexingly, despite their widespread appearance in spontaneous speech and writing, it has not been straightforward to demonstrate that their usage correlates with acceptability in the minds of experimental participants. Zukowski & Larsen (2004) and Ferreira & Swets (2005), for instance, have demonstrated that speakers are not willing to accept resumptives that they themselves have previously produced.

On the other hand, languages such as Irish, Hebrew, or Arabic are typically characterized as pervasive resumption languages — so much so that resumption can be shown to be required in particular constructions.[4] Just how pervasive the acceptability of resumptive pronouns is in these languages is also unclear, however, as differences exist among Hebrew and Arabic (Shlonsky, 1992) and among regional varieties of Arabic (Aoun et al., 2001; Malkawi & Guilliot, 2007) as to which constructions allow resumptive pronouns, suggesting that these languages do not form a unitary class on distributional evidence alone. Experimentally, the empirical state of affairs is no clearer for these languages. In Hebrew, Meltzer-Asscher, Fadlon, Goldstein, & Holan (2015) showed that speakers are often unwilling to accept written resumptive pronouns (though they will accept auditorily presented resumptives). Similarly, Farby, Danon, Walters, & Ben-Shachar (2010) show that resumption is dispreferred in Hebrew relative clauses and slightly improved compared to gaps in island violations, though overall ratings for sentences with resumptive pronouns in all constructions were quite low. Both of these findings are quite striking from the point of view of theoretical studies such as Shlonsky (1992), which take resumptives to be freely varying in these same structural positions in Hebrew and required in Arabic.[5]

[4] These languages are all frequently the focus of studies on resumption. For Irish, see especially McCloskey (1979; 1990); for Hebrew, see especially Borer (1984) and Dermidache (1991); and for Arabic, see especially Aoun (1981); Aoun et al. (2001); and Aoun et al. (2010). For both Hebrew and Arabic together, see Shlonsky (1992).

[5] As a reviewer points out, one issue which is raised by studies which ask for retrospective judgments on resumption is whether linguistically naive speakers employ prescriptive behavior

The gestalt which emerges from this literature is that while theoretical accounts of resumption have posited good reasons to believe that there is some difference between pervasive and limited resumption languages, what remains less clear is whether this difference is one of quality or quantity, especially given the results from Hebrew.[6] Moreover, it remains unclear what the relationship is between the grammaticality judgments which underpin the theoretical work on resumption and the acceptability measures used in experimental studies which call the generalizations of this theoretical work into question.

1.2 Resumption and Island Amelioration

One place where this heterogeneous literature on resumption coalesces around a consensus involves the interaction of resumption with constructions which independently disallow filler-gap dependencies, such as SYNTACTIC ISLANDS (ROSS, 1967, *et seq.*). Attempts to construe a *wh*-filler with a gap inside of these structures usually results in extreme degradation or wholesale ungrammaticality (3):[7]

(3) * *What*$_i$ did Poirot investigate the cause of [because he couldn't understand the reason for t_i]?

While the bracketed structure in (3) is acceptable without a *wh*-dependency, as is the independent creation of long-distance *wh*-filler-gap dependencies, sentences like that in (3) are degraded. The typical explanation for these effects, stemming

when rejecting a resumptive in favor of a gap. In English, for instance, resumption is certainly prescriptively derided in general. This is potentially an issue in Modern Standard Arabic, a language with a robust classical grammatical tradition which informs modern prescriptive norms. Since our aim is to investigate the interaction of resumption with grammatical islands (see below), we have no choice but to freely consider gaps and resumptive pronouns together in order to provide a comparison between the two options for expressing the tail of the filler-gap dependency.

[6]Note that this issue becomes more acute if we also consider the data discussed in note 3 concerning bound readings of resumptive pronouns, which some English speakers from the Midwest (including the first author) do allow (Sells, 1984).

[7]Where relevant, we use [brackets] to denote island structures while remaining agnostic to constituency within that structure.

from the work of Ross (1967), is that the grammatical process responsible for the creation of filler-gap dependencies cannot operate out of these structures.

Many researchers have hypothesized that a resumptive pronoun in the place of the gap might ameliorate or even fully rescue island violations (Kroch, 1981, *et seq.*), a logical theoretical conclusion given the hypothesis that island unacceptability is somehow driven by constraints on the creation or representation of the gap itself. Amelioration is also generally expected in lines of inquiry which take resumptive pronouns to facilitate sentence processing, a procedure which is arguably more difficult in the face of a syntactic island.[8] That resumptives might indeed ameliorate islands is seemingly confirmed by judgments such as those in (4), where the native speaker intuitions indicate that resumption ignores island constraints or improves the resulting grammaticality:[9]

(4) a. *ʔajja sˁaħn*$_i$ baddkun taʕrfo [ʔəza tˁalabit***(-o**$_i$**)** laila
which dish$_i$ want.2P know.2P [whether ordered.3FS***(-it**$_i$**)** Laila
b-l-matˁʕam]?
in-the-restaurant]
"*Which dish*$_i$ do you want to know [whether Laila ordered ***(it**$_i$**)** in the restaurant]?"

b. ? *Who*$_i$ did McNulty aggressively question [the rumor that Lester knew **them**$_i$]?

However, experimental confirmation of this intuition has been notoriously hard to come by in the literature, at least from the previous studies on English. The vast majority of studies suggest that resumption does not significantly rescue an island violation (Alexopoulou & Keller, 2007; Heestand, Xiang, & Polinsky, 2011; Clemens, Scontras, & Polinsky, 2012 and Polinsky, Clemens, Morgan, Xi-

[8] See Kroch (1981); Prince (1990); Erteschik-Shir (1992); Asudeh (2004; 2011; 2012); and Hofmeister & Norcliffe (2013); among many others. For the claim that islands are difficult to process, see Sprouse, Wagers, & Phillips (2012) and references therein.

[9] (4a) is from Lebanese Arabic (Aoun et al., 2010:146). The strongest formulation of this claim is that resumption is required as a LAST RESORT when filler-gap dependency formation is blocked and dispreferred otherwise (Kroch, 1981; Shlonsky, 1992; Fox, 1999; and Aoun et al., 2001).

ang, & Heestand, 2013).[10] This is not a ubiquitous finding, however, as Ackerman, Frazier, & Yoshida (2015) have shown that forced-choice tasks do reveal some measure of acceptability judgment amelioration in islands, a result which aligns with the theoretical literature's conclusion that gaps are strongly unavailable inside islands. Finally, Beltrama & Xiang (2017) have suggested that grammaticality or acceptability may not be improved by resumption in an island, but comprehensibility may be, insofar as speakers reported sentences with a resumptive pronoun as more interpretable than those with gaps. Beltrama & Xiang (2017) conclude that there is a processing benefit to resumption inside islands, but that benefit is due to increased understanding of intended meaning, not acceptability.

However, as noted above, English may not be the ideal target language for experimental studies of resumption owing to its status as a limited resumption language. Only one study that we are aware of that directly addresses the relationship between resumption and islands in pervasive resumption languages. In two experiments, Farby et al. (2010) examined whether Hebrew resumption was preferred to gapping in structures with and without islands and demonstrated that resumption was actually dispreferred in non-island contexts but marginally preferred in island-violating filler-gap dependencies. This dovetails with results from Meltzer-Asscher et al. (2015), who also showed that resumption is disfavored by Hebrew speakers in written presentation, even in non-island contexts.

Stepping back, we can observe a disconnect between the theoretical and experimental literature on resumption: whereas the theoretical literature suggests that island violations should be greatly improved by the presence of a resumptive pronoun, the core finding from each of the English studies is that acceptability does not meet grammaticality threshold criteria even when a resumptive pronoun

[10] Similar conclusions can be drawn from the results reported in McDaniel & Cowart (1999); McKee & McDaniel (2001); Keffala & Goodall (2011); Keffala (2011); and Han et al. (2012), if one draws a distinction between subject and object gaps in island contexts.

appears in an island. However, one must ask whether this could be due to the unacceptable nature of resumptive pronouns more generally in limited resumption languages. Against this backdrop it is somewhat curious that previous studies on Hebrew show only a marginal preference for resumption in island contexts, calling into question the characterization of pervasive and limited resumption languages as a qualitative distinction. However, literature from pervasive resumption languages is not the norm in psycholinguistic work and has only begun to emerge in recent years, leaving the vast preponderance of experimental evidence coming from English or other limited resumption languages.

Here, we aim to add to this emerging literature by examining the behavior of resumption in islands in Modern Standard Arabic, a language which allows resumption in constituent *wh*-questions (5a) and can even employ resumption to amnesty violations of the Empty Category Principle/*that*-Trace Effect (5b):

(5) a. *ʔajja kitaabin*$_i$ ʔiʃtarat-**hu**$_i$ laila min al-maktabati?
which book$_i$ bought.3FS-**it** Laila from the-bookstore
"*Which book*$_i$ did Laila buy **it**$_i$ from the bookstore?"
(Aoun et al., 2010:136)

b. *ʔajja ʔawladin*$_i$ qult-um ʔanna-***(hum**$_i$**)** fi-l-bajti.
which children$_i$ said-2P that-***(they**$_i$**)** in-the-house
"*Which children*$_i$ did you say that **they**$_i$ were in the house?"
(Aoun et al., 2010:137)

Modern Standard Arabic therefore provides a fruitful grammatical context with pervasive resumption which is demonstrably capable of repairing violations of grammatical constraints. All else being equal, one would expect that resumption should have an ameliorating effect on island violations in this language, if such an effect indeed exists.[11]

[11] It is certainly *a priori* possible that spoken dialectal Arabic could differ in its treatment of resumption and islandhood from the variety that forms the object of our study, Modern Standard Arabic. However, our study uses written stimuli, and there are no conventions for writing spoken Arabic in a uniform way. See Aoun & Choueiri (2000); Aoun et al. (2001); Guilliot & Malkawi (2006); Malkawi & Guilliot (2007); and Aoun et al. (2010) for discussions of dialectal differences in the availability of resumption in general.

1.3 The Present Study

The two experiments reported here attempt to add data concerning three major questions to the literature examining the relationship between resumption and island amelioration: (1) Do pervasive resumption languages such as Arabic show differential results concerning island amelioration by pronominal resumption? (2) Does the amelioration effect of resumption — if it exists — appear differentially by island type? and (3) Can the acceptability improvement induced by resumption (relative to gapped structures) be shown to ameliorate the grammatical constraint on island extractions independent of other acceptability costs associated with islands? These questions are examined in two experiments using three islands: *whether* islands, adjunct islands, and the Complex Noun Phrase Constraint.

1.3.1 Methodology

The methodology we use to address these questions, especially the component of acceptability question, is the factorial Likert-acceptability designs reported in Sprouse et al. (2012); Almeida (2014); and Sprouse, Caponigro, Greco, & Cecchetto (2016). We opt to use this design due to its demonstrated ability to separate island effects from other known influences on off-line ratings due to extra-grammatical processing concerns. Sprouse et al. (2012) and other researchers note that there are two key processing constraints that every island-violating extraction also violates (see also Kluender & Kutas, 1993 and Hofmeister & Sag, 2010, among others):

(6) a. LENGTH: There is a cost associated with long-distance filler-gap dependency formation/parsing — longer dependencies are harder to create/parse than shorter dependencies.
b. STRUCTURE: There is a cost associated with island structures — island structures are complex, and building/parsing them requires more work than building/parsing non-island structures.[12]

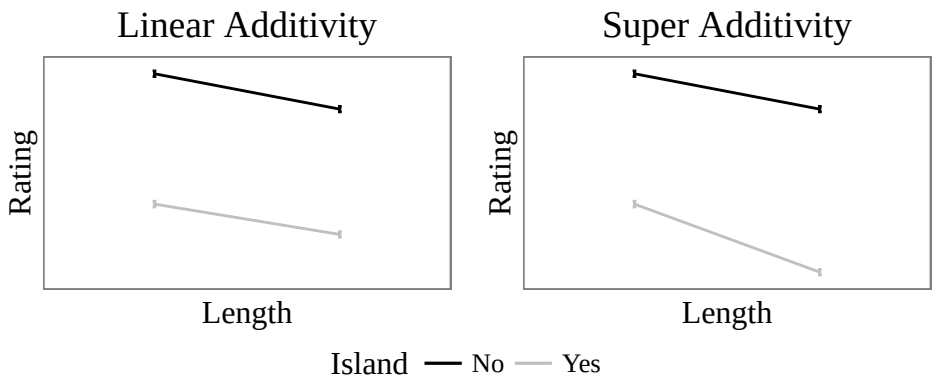

Figure 1: Predicted factorial plots for a linearly additive (no island, left) result and a superadditive (island, right) result.

Starting with Sprouse et al. (2012), it has been shown repeatedly that one can assess the components of length and structure penalties in island violation structures in fully-crossed experimental designs which manipulate both length and structure simultaneously. As Sprouse et al. (2012) demonstrate (and which was replicated for Brazilian Portuguese by Almeida, 2014 and Italian by Sprouse et al., 2016), island violations do not incur ratings which are a simple linear addition of the penalties incurred by long dependency length and island structure alone (see Figure 1). In contrast to what one would expect if island violations were simply the combination of long dependencies from difficult structures (the LINEAR ADDITIVE scenario; left sub-plot), island violations in English, Italian, and Brazilian Portuguese have been shown to be greater than the sum of length and structure penalties alone (the SUPER ADDITIVE scenario; right sub-plot).

While the studies in Sprouse et al. (2012); Almeida (2014); and Sprouse et al. (2016) were designed to assess the predictions of various grammatical and

[12]Throughout this paper, we leave the precise nature of the complexity underlying island structures somewhat vague, since its specification is not necessary for the claims we ultimately make — we simply follow the general assumption that some property of island structures makes them hard to construct and/or process. See Kluender & Kutas (1993) and Hofmeister & Sag (2010) for specific proposals about this complexity in terms of general processing constraints such as a lexical semantic complexity attributed to the members of the island construction. However, these specific complexity effects are sometimes not present in experimental studies which directly assess their presence; see the discussion in Sprouse et al. (2012).

psycholinguistic models of island effects, here we focus on another application of this design and reasoning: its ability to *define* the presence of an island effect in rating studies. If a crossed design yields a superadditive result, then by the logic in Sprouse et al. (2012), *et seq.*, an island is present and being violated in the stimuli. If, on the other hand, a linearly additive pattern results, then one cannot conclude that any observed decrease in rating is attributable to more than processing considerations (defined in terms of length and structure) alone.

Here we employ this line of reasoning to assess whether resumptive pronouns in Arabic ameliorate islands. We begin by attempting to demonstrate that island effects obtain in three structures in Modern Standard Arabic (MSA): (1) *whether*-islands, (2) adjunct islands, and (3) Complex Noun Phrase Constraint (CNPC, Ross, 1967) violations. We then ask whether the addition of a resumptive pronoun in the tail position of the dependency changes the resulting picture from a superadditive to linearly additive one, a picture which is consistent with the grammatical idea that resumption has an ameliorating effect.

2 Experiment 1

Experiment 1 attempts to replicate as closely as possible, for MSA, the constituent question designs in Sprouse et al. (2012) and Sprouse et al. (2016) with the addition of resumptive pronoun conditions which allow assessment of the impact of resumption on superadditivity.

2.1 Participants

Participants were 123 native speakers of Arabic and proficient readers of Modern Standard Arabic by self report (81 female; mean age 22.5 years). Participants were all either students or faculty at the United Arab Emirates University (UAEU), New York University Abu Dhabi (NYUAD), or members of their families recruited by word-of-mouth.[13] Nearly all participants were self-reported native speakers of Emirati Arabic, a variety of the Gulf spoken dialect of Arabic.[14]

All participants provided informed consent and were compensated for 45 minutes of time. This and the subsequent experiment were approved by both the NYUAD Institutional Review Board as well as the UAEU Ethics Committee.

2.2 Materials & Design

For this experiment, three islands were selected from the set created for the experiment in Sprouse et al. (2012) and Sprouse et al. (2016): (i) WHETHER ISLANDS, (ii) ADJUNCT ISLANDS, and (iii) violations of the CNPC. Equivalent sentences with a constituent *wh*-question dependency that spanned these islands were constructed in Modern Standard Arabic:[15] (i) *whether* islands involved complement

[13] The recruitment at the UAEU also explains the discrepancy in gender balance — instruction at the UAEU is conducted on gender-segregated campuses and the majority of recruitment was conducted by a female research assistant on the female campus.

[14] We include this detail because the spoken Arabic dialects are well-known for displaying differences in the acceptability of resumption across constructions (Malkawi & Guilliot, 2007; Aoun et al., 2010), and it is almost certainly the case that a participant's spoken dialect exerts influence on judgments of MSA sentences, especially in rare or marginal constructions. However, since Emirati Arabic resumption has never been the focus of detailed theoretical work, we must leave this matter for future research, noting only here that our data represent the judgments of a largely Emirati population. To our knowledge, the only other dialects represented in our participant pool are Palestinian and Sudanese Arabic.

[15] *wh*-constituent questions were chosen in order to maximize parallels with the Sprouse et al. (2012; 2016) stimuli. We take up the issue of other types of extractions (such as relativization and left dislocation/topicalization) in the conclusion.

question CPs headed by ما إذا/*maa iðaa*, (ii) adjunct islands involved adjunct CPs headed by إذا/*iðaa* ("if"), and (iii) CNPC violations involved a noun with a CP complement. Examples of each of these islands (with resumption) appear in the sentences in (7–9):

(7) *Whether* Islands:
ماذا يتساءل الشرطي ما إذا كان محمود سرقه؟
maaðaa$_i$ jatasaaʔalu aʃ-ʃurtˤii [maa ʔiðaa kaana maħmood saraqa-**hu**$_i$]?
what$_i$ wonders the-policeman [whether had Mahmoud stole-**it**$_i$]?
"*What*$_i$ does the policeman wonder [whether Mahmoud stole (**it**$_i$)]?"

(8) Adjunct Islands:
ماذا تقلق إذا نسيه المحامي في المكتب؟
maaðaa$_i$ taqlaqu [ʔiðaa nasiija-**hu**$_i$ ʔal-muħaamii fii-l-maktab]?
what$_i$ worry.2MS [if forgot.3MS-**it**$_i$ the-lawyer at-the-office]
"*What*$_i$ do you worry [if the lawyer forgot (**it**$_i$) at the office]?"

(9) CNPC Violations:
ماذا أنكرتَ الحقيقة أن احمد أكله؟
maaðaa$_i$ ʔankarta [ʔal-ħaqiiqa ʔanna ʔaaħmed ʔakala-**hu**$_i$]?
what$_i$ denied.2MS [the-fact COMP Ahmed ate.3MS-**it**$_i$]
"*What* did you deny [the fact that Ahmed ate (**it**$_i$)]?"

For each of the six experimental sentences in each island, six conditions were created by varying three experimental factors: (1) LENGTH of the filler-gap dependency (SHORT, where extraction took place from the matrix subject position and LONG, where extraction took place out of the embedded object position), (2) the presence or absence of an ISLAND structure (NOISLAND, ISLAND), and (3) the presence or absence of a RESUMPTIVE pronoun (NORESUMPTION, RESUMPTION). Factors (1–2) were fully crossed in our design. However, since a resumptive pronoun in the Short conditions would involve a *wh*-filler immediately followed by a pronoun which resumes it (*e.g.*, **Who he wonders if Shawn saw Gus*), we elected to use a deficient design and manipulate the RESUMPTION variable only inside the Long conditions. The result is a 2 × 2 + 2 experimental

design where resumptive pronouns alternate with gaps only in the Long extraction conditions where the moved *wh*-word is the thematic direct object.[16] A complete item set for one *whether* island condition appears in (10) and a complete list of all items and conditions appears along with all data and analysis code in a the online supplementary materials.[17]

(10) a. Short/No Island/No Resumption

man jaʕtaqidu ʔanna maħmood saraqa ʔal-miħfaðˁa?
who thinks.3ms comp Mahmoud stole.3ms the-wallet

"Who thinks that Mahmoud stole the wallet?"

b. Long/No Island/No Resumption

maaðaa jaʕtaqidu ʔaʃ-ʃurtˁii ʔanna maħmood saraqa?
what thinks.3ms the-policeman comp Mahmoud stole

"What does the policeman think that Mahmoud stole?"

c. Long/No Island/Resumption

maaðaa jaʕtaqidu ʔaʃ-ʃurtˁii ʔanna maħmood saraqa-hu?
what thinks.3ms the-policeman comp Mahmoud stole-it

"What does the policeman think that Mahmoud stole (it)?"

d. Short/Island/No Resumption

man jatasaaʔalu maa ʔiðaa kaana maħmood saraqa
who wonders whether had.3ms Mahmoud stole.3ms
ʔal-miħfaðˁa?
the-wallet

"Who wonders whether Mahmoud stole the wallet?"

e. Long/Island/No Resumption

maaðaa jatasaaʔalu ʔaʃ-ʃurtˁii maa ʔiðaa kaana maħmood
what wonders the-policeman whether had Mahmoud
saraqa?
stole?

"What does the policeman wonder whether Mahmoud stole?"

f. Long/Island/Resumption

[16] Crucially, this is a position in which resumption is typically characterized as optional in MSA. See Aoun et al. (2010); Alotaibi & Borsley (2013).

[17] Available at the first author's *figshare* account: `https://figshare.com/authors/Matthew_Tucker/700594`.

maaðaa jatasaaʔalu ʔaʃ-ʃurtˤii maa ʔiðaa kaana maħmood
what wonders the-policeman whether had Mahmoud
saraqa-hu?
stole-it?

"What does the policeman wonder whether Mahmoud stole (it)?"

Alongside these items we created 64 filler sentences with a uniform distribution of acceptability ratings (mean rating: 3.4 as rated by four native speakers including the second author). The combined fillers and items were distributed across six lists in a Latin Square design such that a given participant only saw one experimental sentence from each island item set and each list had a 3.5:1 filler-to-item ratio. Each participant therefore saw one observation per condition. Finally, since these items were written in Modern Standard Arabic, it is worth noting that we employed diacritic short vowel/case markers only where needed to disambiguate lexically ambiguous strings in context — exactly the same use of diacritics which is common in everyday written MSA in the Arab world.

2.3 Procedure

Assembly of the experimental surveys was completed by computer software written by the authors.[18] Participants met with an experimenter who provided them with a paper copy of the survey and explained the directions verbally as well as answered any questions the participants had. Participants were then allowed to complete the survey at their leisure and arranged for follow-up appointment with the experimenter to return the survey and receive compensation.

The survey itself consisted of a set of instructions (in MSA) which directed the participant to carefully read the following set of sentences and rate them in terms of their acceptability from 1/١ ("You cannot imagine a speaker of Arabic

[18]This software is available at `https://github.com/matthew-tucker/Likertator`.

saying or writing that sentence”/لا تستطيع التخيل أن ناطقاً باللغة العربية يكتب أو يقول
جملةً كهذه) to 7/٧ (“this sentence is very natural to you”/هذه الجملة تبدو لك طبيعية جداً).
Participants were instructed to take as much time as they needed and to indicate their choice by circling the appropriate number below the sentence. Results of both the acceptability and demography data were digitized by hand.

2.4 Analysis & Predictions

Since our experimental design was a 2 × 2 + 2 reduced design, there was no *a priori* obvious way to analyze all the conditions together in a single statistical model. Since we were primarily interested in the presence or absence of an island effect with or without resumption, we fit two linear mixed effects regression models for each island type to subsets of the conditions with and without resumptive pronouns. Thus data were entered into two linear mixed models, both with LENGTH and ISLAND as fixed effects, one for Resumption conditions (the “resumption model”) and another for NoResumption conditions (the “gap model”). Short dependencies were included in both models. Fixed effects were dummy-coded with the Short, NoIsland, and NoResumption levels as reference. In order to directly compare gaps and resumptive pronouns, we also conducted a planned comparison between the Long/Island conditions with and without resumption.

For both these models it was also not possible to construct a fully-crossed random effects structure with subjects and items as random effects owing to the lack of multiple observations per condition per subject. Both models, therefore, contained only subjects as random effects. In this and the subsequent condition we first fit a random maximal effects structure (Barr, Levy, Scheepers, & Tily, 2013). In order to avoid concerns about interpretability and overparameterization (Bates, Kliegl, Vasishth, & Baayen, 2015), we then compared this model

to a model with only random intercepts for subjects. Since in all cases these models were qualitatively identical, we report here only the models with random intercepts for subjects for all islands and all experiments. Finally, owing to the difficulty in estimating degrees of freedom for *t*-tests of fixed effect significance in mixed effects models, we simply report the *t*-value without an accompanying *p*−value. A fixed effect is considered significant if the absolute value of its *t*-ratio was greater than two, a reasonable heuristic for its 95% confidence interval not including zero (Gelman & Hill, 2006). We report and comment on marginal effects (based upon the 90% confidence interval) when the absolute value of *t* is greater than 1.65, as well. All the data and analysis code for this project is available with the online supplementary materials.

If the grammatical and acceptability patterns in MSA are similar to those studied in English (Sprouse et al., 2012), Portuguese (Almeida, 2014), and Italian (Sprouse et al., 2016), then we expect to find effects of both the LENGTH and ISLAND variables such that long sentences and island-containing sentences should be rated lower than their short and non-island counterparts, respectively. Additionally, the presence of a grammatical island effect should manifest as an interaction of these two variables, such that the Long/Island conditions are much worse than would be predicted by the sum of the Long and Island penalties alone.

For resumption, what we expect depends upon the grammatical and psychological reality of resumption as an ameliorating grammatical process. If resumption is preferred in Arabic, as the grammaticalized resumption hypothesis supposes, then the Resumption conditions should show higher ratings than the corresponding NoResumption conditions, both with and without islands. In island contexts, if it is indeed the case that islands are ameliorated by resumption, then we expect a rating increase in the Long/Island/Resumption contexts which is equal to or greater than the increase seen in non-island long dependencies with resump-

	Island Type		
Condition	*Whether*	*Adjunct*	*CNPC*
Short/NoIsland	0.67 (0.08)	0.35 (0.09)	0.79 (0.08)
Long/NoIsland/NoR	0.65 (0.08)	−0.01 (0.09)	−0.25 (0.08)
Long/NoIsland/R	−0.31 (0.08)	−0.40 (0.07)	−0.53 (0.07)
Short/Island	0.28 (0.08)	0.03 (0.07)	0.36 (0.08)
Long/Island/NoR	−0.45 (0.07)	−0.59 (0.07)	−0.62 (0.07)
Long/Island/R	−0.35 (0.07)	−0.75 (0.07)	−0.59 (0.07)

Table 1: Means and standard errors of standardized ratings over subject rates for each condition and island in Experiment 1 (N = 123).

tion. If either of these hypotheses about Arabic are incorrect, then we expect to find no significant effect of the resumption manipulation in this experiment.

2.5 Results

Before analysis, raw acceptability ratings were z−score transformed. This standardization transformation expresses a given rating in terms of its difference from the by-participant mean in units of standard deviations from that mean. This transformation helps mitigate the effect of individual differences in scale bias from participant to participant.

The normalized (z-transformed) means and standard errors for all three island types appear in Table 1 and the results of the linear mixed effects regression for both gap and resumption models appear in Table 2. In this and the results section of Experiment 2, factorial plots for each of the three islands are presented separately. For textual presentation of the results, we comment on factors with t-values greater than 1.65 ($p \approx .1$) and 2.00 ($p \approx 0.05$) separately.

TERM	ISLAND TYPE		
	Whether	*Adjunct*	*CNPC*
	Gap Model		
Length	−0.61(−5.84)	−0.36(−3.20)	−1.04(−9.74)
Island	−0.40(−3.82)	−0.32(−2.86)	−0.43(−4.06)
Length × *Island*	−0.11(−0.77)	−0.26(−1.66)	0.06(0.43)
	Resumption Model		
Length	−0.98(−9.17)	−0.75(−7.03)	−1.32(−12.63)
Island	−0.40(−3.73)	−0.32(−3.02)	−0.43(−4.12)
Length × *Island*	0.35(2.33)	−0.03(−0.23)	0.37(2.51)

Table 2: Linear mixed effects model coefficient estimates for Experiment 1. Values in parentheses represent the t value against an $H_0 : \beta = 0$.

2.5.1 Whether Islands

A factorial plot of the standardized rating scores for *whether*-island conditions appears in Figure 2.[19] For *whether*-islands, the statistical analysis revealed an effect of LENGTH such that LONG sentences were rated lower than Short sentences in both the gap ($\hat{\beta} = -0.61; s.e. = 0.10; t = -5.84$) and Resumption ($\hat{\beta} = -0.98; s.e. = 0.10; t = -9.17$) models. Similarly, there was also a effect of ISLAND such that Island structures were rated lower than NoIsland structures in both the gap ($\hat{\beta} = -0.40; s.e. = -0.10; t = -3.82$) and Resumption ($\hat{\beta} = -0.40; s.e. = 0.10; t = 3.73$) models. The interaction of LENGTH and ISLAND was not significant in the gap model ($\hat{\beta} = -0.11; s.e. = 0.15; t = -0.77$) but was in the resumption model ($\hat{\beta} = 0.35; s.e. = 0.15; t = 2.33$) such that island status had less of an effect in long conditions than in Short conditions. A planned comparison between the means in the Long/Island/NoResumption and Long/Island/Resumption conditions did not reach significance ($t(121) = -1.17; p = 0.24$).

[19] With this and all subsequent plots, we assume that short conditions are equal in both Resumption and NoResumption conditions for the sake of more coherent plots, despite the absence of a true RESUMPTION contrast in Short conditions.

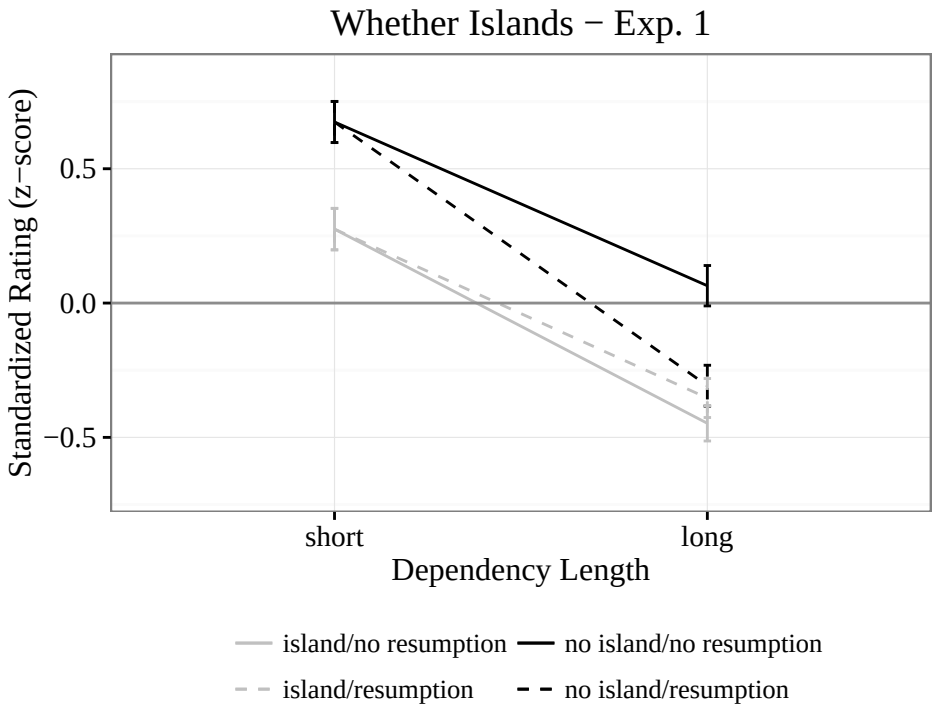

Figure 2: Mean standardized (*z*-transformed) ratings by condition for *whether* islands in Experiment 1. Error bars represent the standard error of the mean across subject ratings.

2.5.2 Adjunct Islands

A factorial plot of the standardized rating scores for the adjunct island conditions appears in Figure 3. For the these islands, the statistical analysis revealed a effect of LENGTH such that Long sentences were rated lower than Short sentences in both the gap ($\hat{\beta} = -0.36; s.e. = 0.11; t = -3.20$) and resumption ($\hat{\beta} = -0.75; s.e. = 0.11; t = -7.03$) models. There was also a effect of ISLAND such that Island sentences were rated lower than NoIsland structures in both the gap ($\hat{\beta} = -0.32; s.e. = 0.11; t = -2.86$) and resumption ($\hat{\beta} = -0.32; s.e. = 0.11; t = -3.02$) models. The interaction of LENGTH and ISLAND status was marginal in the gap model ($\hat{\beta} = -0.26; s.e. = 0.16; t = -1.66$) such that the LENGTH × ISLAND interaction was superadditive, but this did not reach significance in the resumption model ($\hat{\beta} = -0.03; s.e. = 0.15; t = -0.23$). A planned comparison between sentences in the Long/Island/NoResumption condition and Long/Island/Resumption condition revealed that the former were rated significantly higher than the latter ($t(122) = 2.03; p = 0.04$).

2.5.3 CNPC Violations

A factorial plot of the standardized rating scores for the CNPC violations appears in Figure 4. For the CNPC violations, the statistical analysis revealed a effect of LENGTH such that Long sentences were rated lower than Short sentences in both the gap ($\hat{\beta} = -1.04; s.e. = 0.11; t = -9.74$) and the resumption ($\hat{\beta} = -1.32; s.e. = 0.10; t = -12.63$) models. Furthermore, there was also an effect of ISLAND such that sentences containing island structures were rated lower than those without island structures in both the gap ($\hat{\beta} = -0.43; s.e. = 0.11; t = -4.06$) and resumption ($\hat{\beta} = -1.32; s.e. = 0.11; t = -4.12$) models. The interaction of LENGTH and ISLAND was not significant in the gap model ($\hat{\beta} = 0.06; s.e. = 0.15; t = 0.43$) but was significant in the resumption model ($\hat{\beta} = 0.37; s.e. = 0.15; t = 2.51$)

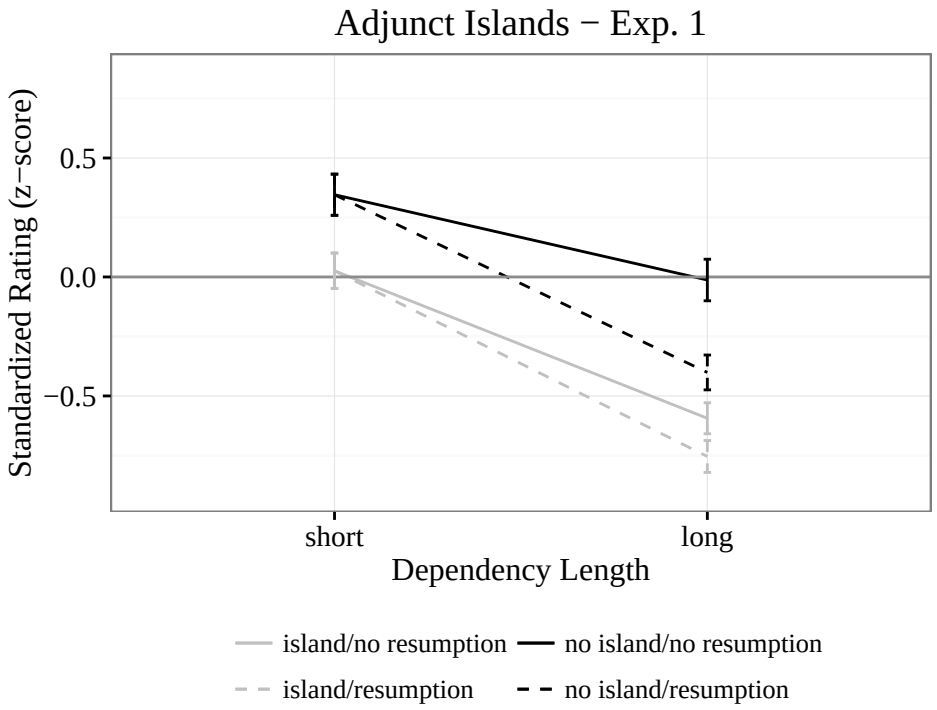

Figure 3: Mean standardized (*z*-transformed) ratings by condition for adjunct islands in Experiment 1. Error bars represent the standard error of the mean across subject ratings.

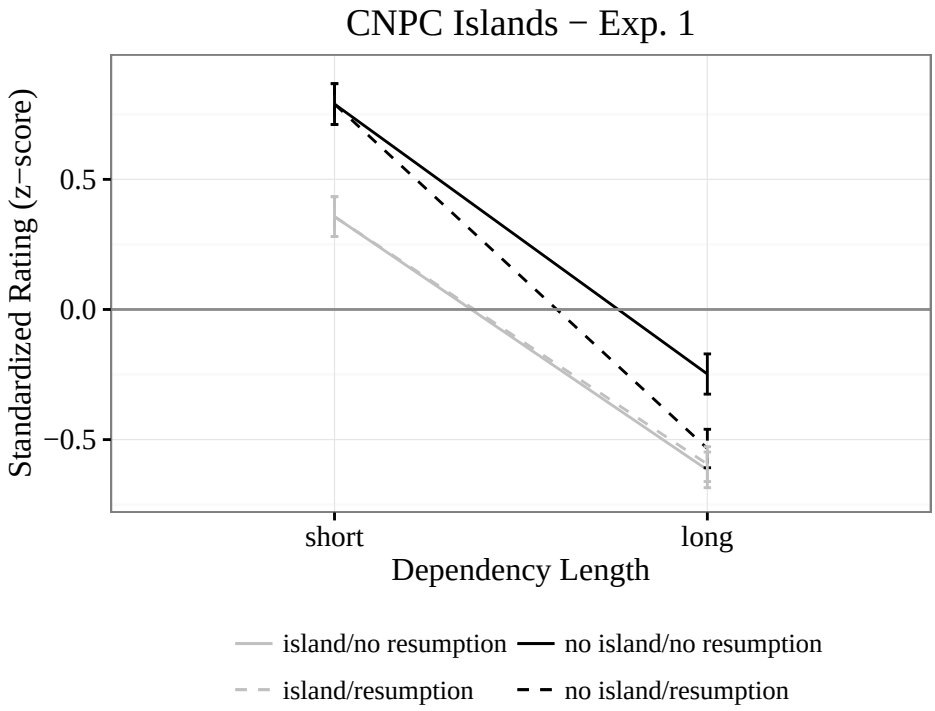

Figure 4: Mean standardized (*z*-transformed) ratings by condition for CNPC violations in Experiment 1. Error bars represent the standard error of the mean across subject ratings.

such that the effect of island structure was diminished in the long conditions relative to the short conditions. A planned comparison between the scores in the Long/Island/NoResumption and Long/Island/Resumption conditions did not reveal a significant difference between sentences with and without a resumptive pronoun in the Long/Island conditions, given that the two were rated very close to equal ($t(122) = -0.22; p = 0.81$).

2.6 Discussion

The results of Experiment 1 are somewhat mixed with respect to the question of island amelioration by resumptive pronouns, and this is in part due to the muddled nature of our acceptability results with respect to both islandhood in MSA and the

acceptability of resumption in these items. We take up each issue separately.

It is clear from our results that not all three of the island constructions — *whether* islands, adjunct islands, and CNPC violations — show superadditivity in either a quantitative or qualitative way. Leaving resumption to the side for a moment, one can observe that all three of our islands show the effects of length and structure on rating scores, but not the superadditive phenomenon associated with islands in Sprouse et al. (2012); Almeida (2014); and Sprouse et al. (2016). While the *whether* and adjunct islands do show the correct sign/qualitative pattern on the interaction of LENGTH and ISLAND status, only the adjunct islands approach qualitative significance and only marginally. It is true that there was a significant interaction of length and structure in the resumption models for adjunct and *whether* islands, but this is due to a massive rating penalty for sentences in the Long/NoIsland/Resumption conditions and as such, the sign is incorrect for a superadditive island effect.

What does this mean for theories of islandhood? The simplest and most radical conclusion here would be that MSA simply lacks these three islands. However, we believe this would miss the point somewhat, as all three structures involve very low ratings ($-0.75 \leq z \leq -0.25$) in long dependency conditions without resumption. Moreover, this conclusion would run counter to every theoretical study of islands in Arabic that we are aware of (see, *e.g.*, Aoun et al., 2010:ch.6 and references therein). What is not clear from our results is whether there is a superadditive component to this unacceptability that is unattributable to length or structure alone. Based on just these results, one might be tempted to conclude that MSA places a premium on short dependency length out of simple structures, and that this just happens to dovetail with structures that are superadditive islands in other languages.

However, we believe that something else is clearly at play in our data, given

the strange behavior of resumption in our results. Nearly every single theoretical and descriptive treatment of MSA filler-gap constructions concludes that resumption is preferred or at least possible in long dependencies. Yet resumption is clearly strongly penalized in our results, regardless of island status. In all three distinct syntactic structures, sentences in the Long/NoIsland/Resumption condition is hugely dispreferred relative to the equivalent gapped structures. Given the size of this penalty, it is reasonable to wonder whether the sentences in the Long/Island/Resumption condition are possibly at floor (a numerical value which is so low as to preclude observation of lower values of any qualitative use), mitigating any ameliorative effect of the pronoun in general.

The experimental items for this study were constructed to be the closest possible MSA analogues of the items used in Sprouse et al. (2012) and Sprouse et al. (2016). However, in doing so, we adopted without discussion a difference in the *wh*-fillers used in those materials. Specifically, all the Long conditions had filler-gap (or filler-pronoun) dependencies whose filler was the *wh*-word ماذا/*maaðaa*, "what," whereas the Short conditions were all with من/*man*, "who."

As Aoun et al. (2010:130–9) point out, however, this is not an innocuous difference. According to their judgments, *maaðaa* is unacceptable as the head of a dependency whose tail is a resumptive pronoun, whereas *man* is much better. Those authors give several sets of judgments on which fillers are acceptable in MSA with resumption, but crucially for our purposes the relevant contrast to *maaðaa* is أيّ/*ʔajja*, "which" combined with an overt NP restrictor (11):[20]

(11) a. *maaðaa*$_i$ ʔiʃtarat(***-hu**$_i$) laila min al-maktabati?
what$_i$ bought.3FS(***-it**$_i$) Laila from the-bookstore
"*What*$_i$ did Laila buy (***it**$_i$) from the bookstore?"

[20]How exactly to characterize this difference is a matter of some importance that we will not take up here. See Aoun et al. (2010) for some discussion of why d(iscourse)-linking (Pesetsky, 1987) and referentiality (in the sense of Cinque, 1990) are not precisely correct notions. We will use the term "complex," in line with Aoun et al.'s (2010) suggestion that the correct cut has to do with whether the DP in question has articulated syntactic structure.

b. *ʔajja kitaabin*$_i$ ʔiʃtarat(**-hu**$_i$) laila min al-maktabati?
which book$_i$ bought.3FS(**-it**$_i$) Laila from the-bookstore
"*Which book*$_i$ did Laila buy (**it**$_i$) from the bookstore?"

While this judgment contrast is not widely reported in the literature on Arabic, the second author and our consultants all share it, meaning that our items from Experiment 1 contain a clear confound in the Long conditions. We therefore can only make limited conclusions from those items concerning the ability of resumption to ameliorate islands. Experiment 2 is designed to test the same thing as Experiment 1 while controlling for the (in)appropriateness of the *wh*-filler associated with the resumptive pronoun.

3 Experiment 2

One of the striking general findings in Experiment 1 was that resumptive pronouns in the long dependency length conditions were quite marked for speakers, even independently of the presence of a syntactic island structure which the filler-pronoun dependency spanned. In fact, the acceptability cost of resumptive pronouns was larger in non-island contexts than in island contexts. This is an odd finding given that resumption is typically characterized as at least optional in filler-gap dependencies in Arabic when grammatical constraints do not preclude a gap and obligatory when such constraints do preclude a gap (Shlonsky, 1992; Aoun et al., 2001; 2010). Experiment 2 addresses the question of whether or not this could have been due to confounds in the design of our experimental materials related to the inclusion of *wh*-fillers that are not easily linked to resumptive pronouns.

Experiment 2 seeks to remedy this confound by replacing all instances of the confounding *maaðaa* with *ʔajja* + NP, fulfilling the *wh*-filler identity requirement

discussed by Aoun et al. (2010) and allowing us to properly assess the ameliorative effect of resumption.

3.1 Participants

Participants were 119 native speakers of Arabic recruited to an online survey via advertisement at the UAEU, NYUAD, and via the `arabic-L` mailing list. Native speaker status was assessed via presentation of the survey entirely in Arabic as well as via demographic questionnaire. Since the survey was presented online, there was no way to ensure that participants finished the entire experiment. Especially given that no compensation was offered, many participants did not finish. Of the 119 participants which began the study, only 53 completed it (44.5%), so we report results from only those subjects who completed the entire experiment. No other subject exclusion criteria were used. All participants provided informed consent via the online survey. None of the recruitment pools used for Experiment 1 were consulted for the recruitment of participants for Experiment 2.

3.2 Materials & Design

Materials for Experiment 2 were constructed by taking the experimental items from Experiment 1 and replacing all the unacceptable instances of ماذا/*maaðaa* with a DP headed by أيّ/*ʔajja* and containing an NP restrictor which was contextually appropriate for the sentence as a whole — typically this was the noun which appeared in the embedded complement position in the Short conditions. All other constrains on experimental filler and item design from Experiment 1 were duplicated. A complete list of experimental items and conditions appears in the accompanying online supplementary materials.

3.3 Procedure

Participants provided informed consent by clicking on a button before being taken to the directions. The directions were exactly the same as Experiment 1 save for the fact that there was no experimenter present to assess understanding. Subjects indicated their responses to acceptability judgment prompts by clicking on a radio button that displayed all the choices between 1 and 7. All other procedural details were identical to Experiment 1.

3.4 Analysis & Predictions

The statistical analysis was identical to Experiment 1. As far as island superadditivity is concerned, our expectations are the same as Experiment 1 — islands should show an acceptability penalty larger in Long/Island conditions not analyzable as the sum of the LENGTH and ISLAND penalties alone (an interaction term). If the confound involving the *wh*-fillers obliterated the ameliorative effect of resumption in Experiment 1, then we also expect to find that resumptive pronouns linked to *ʔajja* + NP will show a larger measure of acceptability increase than was seen for Long conditions in Experiment 1. This should manifest as an improvement in Long/Island/Resumption conditions relative to equivalent sentences without resumption. Moreover, given that the use of *which*-NP fillers often improves the acceptability of long-distance dependencies more generally (Pesetsky, 1987; Cinque, 1990), we expect that Long conditions should now be above floor and allow us to observe whether superadditivity exists and whether it is ameliorated by the presence of a resumptive.

	Island Type		
Condition	*Whether*	*Adjunct*	*CNPC*
Short/NoIsland	0.76 (0.10)	0.47 (0.12)	0.80 (0.12)
Long/NoIsland/NoR	0.11 (0.12)	0.12 (0.12)	0.18 (0.12)
Long/NoIsland/R	0.51 (0.10)	0.65 (0.10)	0.03 (0.11)
Short/Island	0.71 (0.10)	0.34 (0.12)	0.44 (0.11)
Long/Island/NoR	−0.30 (0.09)	−0.80 (0.08)	−0.66 (0.12)
Long/Island/R	−0.19 (0.11)	−0.25 (0.11)	−0.66 (0.10)

Table 3: Means and standard errors of standardized ratings over subject rates for each condition and island in Experiment 2 (N = 53).

	Island Type		
Term	*Whether*	*Adjunct*	*CNPC*
	Gap Model		
Length	−0.65(−4.32)	−0.35(−2.22)	−0.62(−3.86)
Island	−0.05(−0.33)	−0.14(−0.86)	−0.35(−2.19)
Length × *Island*	−0.37(−1.73)	−0.78(−3.44)	−0.48(−2.08)
	Resumption Model		
Length	−0.25(−1.69)	0.18(1.11)	−0.77(−5.03)
Island	−0.05(−0.34)	−0.14(−0.87)	−0.35(−2.29)
Length × *Island*	−0.65(−3.14)	−0.77(−3.43)	−0.34(−1.54)

Table 4: Linear mixed effects model coefficient estimates for Experiment 2. Values in parentheses represent the t value against an $H_0 : \beta = 0$.

3.5 Results

As with Experiment 1, the raw scores were first $z-$transformed before analysis. The standardized mean ratings for each of the three islands appear in Table 3 and the coefficients for the fixed effects in both the gap and resumption linear mixed effect regression models appear in Table 4.

3.5.1 Whether Islands

A factorial plot of the standardized rating scores for the *whether* island conditions appears in Figure 5. For the *whether* islands, the statistical analysis revealed an ef-

fect of LENGTH such that Long sentences were rated lower than Short sentences in both the gap ($\hat{\beta} = -0.65; s.e. = 0.15; t = -4.32$) model and marginal in the resumption ($\hat{\beta} = -0.25; s.e. = -0.15; t = -1.69$) model. There was no effect of ISLAND in either model (all $|t| < 0.40$). There was, however, a marginal interaction of LENGTH and ISLAND in the gap model ($\hat{\beta} = -0.37; s.e. = 0.21; t = -1.73$) and a significant interaction in the resumption model ($\hat{\beta} = -0.65; s.e. = 0.21; t = -3.14$). In both cases this was due to a superadditive pattern in which island status had a greater impact in Long conditions than in Short ones. A planned comparison between sentences without a resumptive in the Long/Island/NoResumption condition and those with a resumptive in the Long/Island/Resumption condition did not reveal a significant difference between the presence or absence of a resumptive pronoun in long dependencies constructed across islands boundaries ($t(44) = 0.48; p = 0.64$).

3.5.2 Adjunct Islands

A factorial plot of the standardized rating scores for the adjunct island conditions appears in Figure 6. For the adjunct islands, the statistical analysis revealed an effect of LENGTH such that Long sentences were rated lower than Short sentences in the gap ($\hat{\beta} = -0.35; s.e. = 0.16; t = -2.22$) but not the resumption ($\hat{\beta} = 0.18; s.e. = 0.16; t = 1.11$) model. Neither model showed a significant effect of ISLAND (all $|t| < 0.90$). However, there was a significant interaction between LENGTH and ISLAND in both the gap ($\hat{\beta} = -0.78; s.e. = 0.21; t = -3.44$) and resumption ($\hat{\beta} = -0.77; s.e. = 0.28; t = -3.43$) models. In both cases this was due to a superadditive pattern, as with the *whether* island conditions. A planned comparison between sentences in the Long/Island/NoResumption condition and Long/Island/Resumption condition revealed that resumption significantly increased ratings relative to the absence of resumption, a result which was

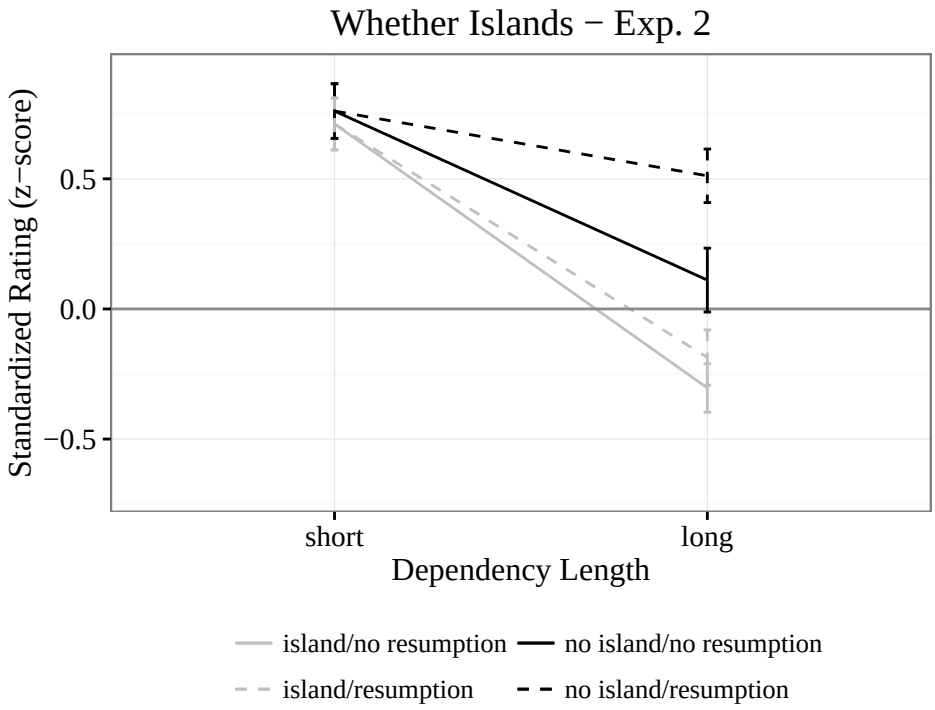

Figure 5: Mean standardized (*z*-transformed) ratings by condition for *whether* islands in Experiment 2. Error bars represent the standard error of the mean across subject ratings.

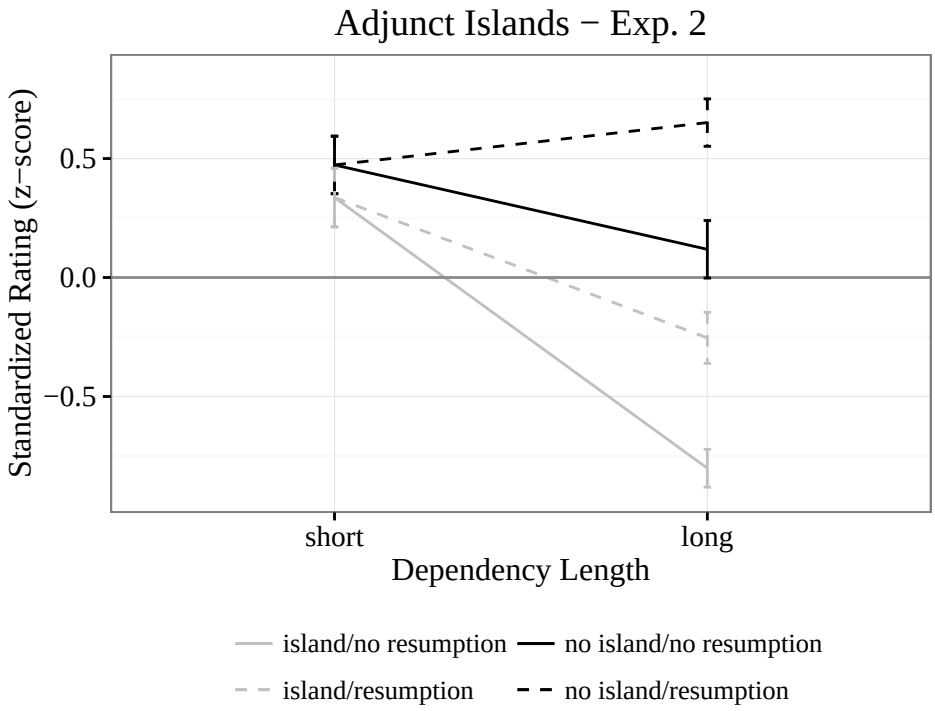

Figure 6: Mean standardized (z-transformed) ratings by condition for adjunct islands in Experiment 2. Error bars represent the standard error of the mean across subject ratings.

quite robust ($t(45) = 4.57; p < 0.0001$).

3.5.3 CNPC Violations

A factorial plot of the standardized rating scores for the CNPC violation conditions appears in Figure 7. For the CNPC violation sentences, the statistical analysis revealed an effect of LENGTH such that Long sentences were rated lower than Short sentences in both the gap ($\hat{\beta} = -0.62; s.e. = 0.16; t = -3.86$) and resumption ($\hat{\beta} = -0.77; s.e. = 0.15; t = -5.03$) models. Furthermore, there was also an effect of ISLAND such that Island sentences were rated lower than NoIsland sentences in both the gap ($\hat{\beta} = -0.35; s.e. = 0.16; t = -2.19$) and resumption ($\hat{\beta} = -0.35; s.e. = 0.16; t = -2.29$) models. Additionally, there was an interaction between LENGTH and ISLAND in the gap model ($\hat{\beta} = -0.48; s.e. = 0.23; t =$

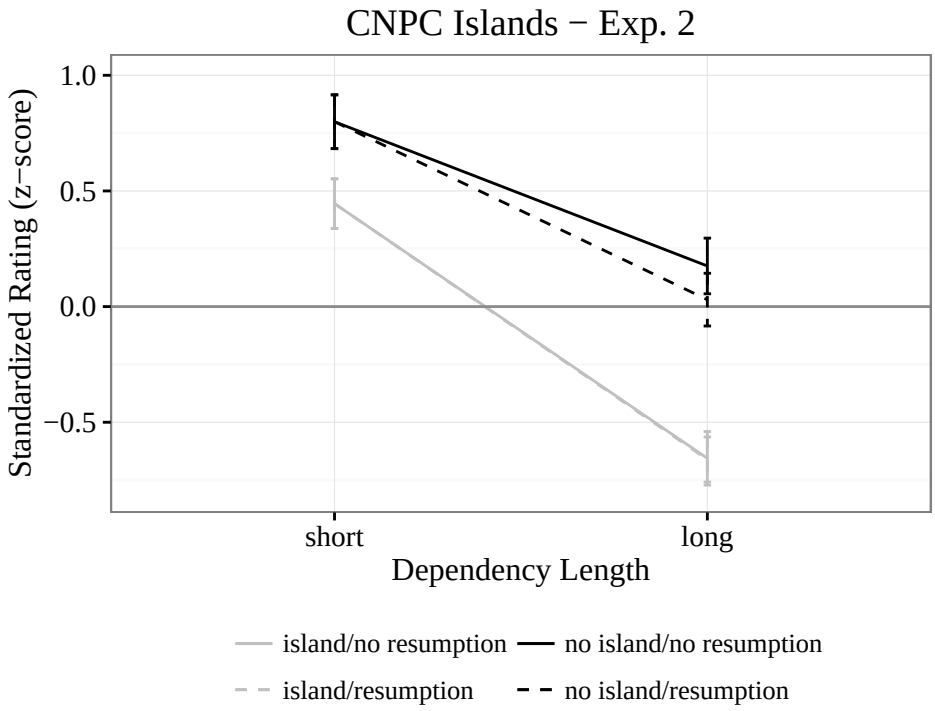

Figure 7: Mean standardized (*z*-transformed) ratings by condition for CNPC violations in Experiment 2. Error bars represent the standard error of the mean across subject ratings.

−2.08) that was not present in the resumption model ($\hat{\beta} = -0.34; s.e. = 0.22; t = -1.54$). A planned comparison between sentences without a resumptive pronoun in the Long/Island/NoResumption condition and those with a pronoun in the Long/Island/Resumption condition did not reach significance ($t(45) = -0.03; p = 0.97$).

3.6 Discussion

The results of Experiment 2 are suggestive of the idea that the confound of inappropriate *wh*-fillers did contaminate the results from Experiment 1. In all three structures — *whether* islands, adjunct islands, and CNPC violations — the data in this experiment show a superadditive effect beyond the contributions of length

and (where it is present) island structure alone. In each of these three structures, the gap model shows a significant interaction of Length and Island such that the effect of structure is larger in the long dependency conditions than in the short dependency conditions — exactly the superadditive effect documented in Sprouse et al. (2012); Almeida (2014); and Sprouse et al. (2016).

What is particularly interesting about these data, however, is that this superadditivity is different in size for each of the three islands. While it is well-known in the theoretical literature that islands do not form a homogeneous class of constructions, our data suggests that the extent to which island violations manifest in acceptability rating studies is a function of individual island identity: adjunct islands display the largest superadditive effect, followed by CNPC violations and *whether* islands. We do not believe this contrast is reducible to known subcategorizations of island structures, since it does not conform to any subcategorization of which we are aware.[21] Since our experiments were not designed to assess differences in island status across these three constructions directly, further study is clearly needed on this point.

Experiment 2 also shows that resumption can ameliorate islands, but again differentially based upon the island in question. For the adjunct islands, resumption in an island-violating context clearly helps a great deal, as evidenced by the significant planned comparison in that structure. CNPC structures, on the other hand, provide an equally clear, but opposite conclusion: resumption appears as acceptable as gapping. With the CNPC violations, however, the strong unacceptability of long-distance dependencies in the island structure without the gap makes one wonder whether we could again be at floor in these data, making the results difficult to interpret — it is very possible MSA simply does not allow filler-gap de-

[21] Crucially, it is not reducible to the notion of "island strength" (see Szabolcsi, 2006; Sprouse et al., 2016; and references therein), since adjunct islands and CNPC islands are both typically taken to be strong islands, yet they show considerable differences in superadditivity in our data.

pendencies to be constructed inside CNPC complements.[22] For *whether* islands, the results seem somewhat in the middle — qualitatively the pronoun appears to help, but the result is not significant.

Finally, focusing on the adjunct islands and *whether* islands, the ameliorative effect of resumption nearly completely wipes out the superadditive effect of the island when assessing the Island/Resumption conditions against the NoIsland/NoResumption conditions — these lines are nearly parallel in the factorial plots. Here we can see the impetus for theoretical researchers to conclude that resumption ameliorates islands, since this would mean that resumption removes the component of the rating penalty associated with extra-processing (and possibly grammatical) concerns, leaving only the penalties associated with length and structure alone.

4 General Discussion

4.1 Differences within Differences

In order to compare our results across islands and experiments, it is helpful to have a numerical estimation of the superadditive component of each island and the effect that resumption has on the superadditivity for a given island. Fortunately, the factorial design we employed throughout both experiments allows for a simple

[22] A reviewer raises one plausible explanation, namely that CNPC island structures are perhaps actually expressed as Construct State nominals, (see Ryding, 2005:ch.8 for a descriptive overview), where the definite determiner is typically ungrammatical on the head noun. Parsing the complex noun phrase as a construct state nominal would suggest a structure where a null copula intervenes between the noun an complementizer (perhaps akin to the "class 2 interrogatives" of Shlonsky, 2002), a parse not available for any of our other islands . We take this point as valid and thank the reviewer for bringing it to our attention — however we are unsure what effect this confound would be expected to have on the CNPC island extractions, so we leave this matter for future theoretical work on the structures underlying complex noun phrase extractions in MSA.

estimation of the superadditive effect using the Differences-in-Differences score (Maxwell & Delaney, 2004; see also the discussion in Sprouse et al., 2012 and Sprouse et al., 2016). The differences-in-differences (DD) score is useful because it provides a clear value of the component of an interaction term not explicable entirely in terms of underlying main or simple effects.

This DD value can be calculated for a two way interaction as follows: first, we calculate the difference between two scores from the four conditions. This value, $D1$, we define as the rating for the Long/NoIsland condition minus the rating for the Long/Island condition. Second, we calculate the difference between the two remaining conditions in the interaction. This value, $D2$, is therefore the rating for the Short/NoIsland condition minus the rating for the Short/Island condition. Straightforwardly, then, $DD = D1 - D2$. This value can be computed for each subject individually and then averaged across subjects to give an estimate of the DD size in a given experiment.

In our study, however, we do not have a 2×2 design with a two-way interaction but a defective $2 \times 2 + 2$ design. We therefore compute two DD scores. The first is computed exactly as described above, using the NoResumption conditions throughout. This DD score gives a straightforward estimation of the superadditivity component of an island *modulo* resumption. The second DD score is designed to assess a change in the superadditivity effect engendered by resuming the gap position with a resumptive pronoun. To that end, it is identical to the first DD score, expect that $D1$ is computed by taking the difference of both the Long/NoIsland/NoResumption and Long/Island/Resumption conditions. This allows us to assess whether there is a superadditive component in resumed island filler-gap dependencies when compared to a long distance dependency with a gap and, if it is present, to compare it directly to the identical DD score for a dependency with gaps across the board. These DD values for all three islands in both

experiments are shown in Table 5.

	Experiment 1		Experiment 2	
Island	No RP	RP	No RP	RP
Whether	0.13 (0.14)	0.05 (0.15)	0.20 (0.21)	0.19 (0.19)
Adjunct	0.25 (0.14)	0.40 (0.15)	0.78 (0.18)	0.21 (0.21)
CNPC	-0.05 (0.14)	-0.06 (0.15)	0.39 (0.23)	0.40 (0.20)

Table 5: Differences within differences scores for all three islands in both experiments, separated by the presence or absence of a resumptive pronoun (RP). Values in parentheses represent the standard error of the mean.

When the DD value is positive, this implies that a superadditive pattern is present and trends in the expected direction for a grammatical island. A number close to zero or negative implies that no superadditive component appears in the ratings for that island/experiment pair. We report standard errors for these DD values but interpret them cautiously, as each subject saw only one item per condition per island in each experiment. We use these DD values in drawing conclusions about the magnitude of island and amelioration effects in what follows.

4.2 Islands in MSA

One of the striking facts in our results is the variability of the presence and magnitude of superadditive island effects across the three constructions and both experiments in our results. In Experiment 1, only *whether* and adjunct islands show superadditive effects, and *whether* islands do so with a smaller magnitude (0.13) than adjunct islands (0.25). In Experiment 2, this general pattern of larger effects for adjunct islands (0.78) than *whether* islands (0.20) remains, but the CNPC violations also showed a superadditive component (0.39).

In theoretical work, which syntactic constructions qualify as islands are often cited as a point of crosslinguistic variation. For instance, Rizzi (1982) notes

that relative clause dependencies can show crosslinguistic variation in islandhood across Italian and English, a fact which is confirmed by the quantitative studies in Sprouse et al. (2016). If superadditivity in quantitative acceptability studies is used as definitional of island effects in the grammar (Almeida, 2014), then our results could be taken to indicate that there is some crosslinguistic variability in the expression of islands when MSA is compared to other languages.

This is an important result since no formal studies that we are aware of discuss the relationship between island acceptability and grammaticality in MSA. We can entertain two hypotheses to account for this heterogeneity in our data. The first would hold that *whether* islands are simply not barriers to the formation of *wh*-filler gap dependencies in Arabic. This would be a somewhat surprising result given the general property that MSA holds many constituents to be barriers to dependency formation with gaps (such as overtly headed CPs and PPs) which are not barriers in, for instance, Indo-European languages. It strikes us as somewhat odd, therefore, to maintain that MSA is more conservative in *wh*-dependency formation in most cases, except in adjunct islands. The other option maintains a more quantitative approach to dealing with the differences between adjunct and *whether* islands. Given that we did not have any *a priori* reason to doubt that Arabic contains the same inventory of island constructions as English, the somewhat weaker status of *whether* islands in our results could be taken as indicative of a quantitative heterogeneity among distinct types of islands in MSA.

However, in our results this issue is possibly confounded by one of population and proficiency in MSA. Experiment 1 demonstrated that only adjunct islands showed a superadditive effect and they only did so marginally. This is true even in the gap statistical model, where the issue of linking a bare *wh*-filler to a resumptive pronoun should not be an issue. We believe that this discrepancy is at least partially attributable to differences in population across the two studies. Ex-

periment 1 mostly surveyed a sample of university students in the United Arab Emirates and their families, whereas Experiment 2 surveyed the `arabic-L` email list, which is comprised of people who, on the whole, have some formal training in the grammar of MSA. The diglossic situation between MSA and the spoken Arabic varieties entails that proficiency in MSA can be an issue in places where other languages are commonly spoken; the UAE is such a place given the prevalence of English. Since English makes much heavier use of the gap strategy relative to MSA, a speaker who is bilingual in spoken Arabic and English with little instruction in MSA beyond primary school may be more inclined to reject resumption *in general* relative to monolingual speakers of a spoken variety of Arabic who have more familiarity with MSA. We think it possible that at least subsets of participants from Experiment 1 may differ in their MSA proficiency from the speakers sampled in Experiment 2, given the possible discrepancy in MSA proficiencies between the two populations. This concern itself underscores the urgent need for work on the psycholinguistics of the spoken Arabic varieties.

4.3 The Grammaticality of Resumption and Gapping

One general issue raised by our study is the grammatical status of resumption and gapping in MSA, even *modulo* the presence of an island. As noted in the introduction, MSA — and Arabic more generally — is generally described as a pervasive resumption language insofar as it makes widespread grammatical use of resumption. Our results suggest that reifying this quantitative difference from English into a qualitative one may be too simplistic, as resumption can appear quite unacceptable even in MSA. In addition to the constraint resumption places on the content of the *wh*-filler, we also see that resumption is dispreferred in certain long-distance dependencies, such as the Long/NoIsland/Resumption conditions in the

CNPC islands in our data. It therefore appears that in certain grammatical corners, a language such as Arabic which otherwise makes widespread use of resumption can behave like English, a language where resumption is generally dispreferred. More generally, this suggests a conception of resumption where typological differences are situated in the representational particulars of the context in which a resumptive pronoun finds itself, not in core grammatical differences between languages.[23] This conception of resumption dovetails with theoretical work that argues that resumption is not a unitary phenomenon, even inside Arabic (Aoun & Choueiri, 2000; Aoun et al., 2001; 2010; Guilliot & Malkawi, 2006; Malkawi & Guilliot, 2007).

However, there is clearly a core validity to the claim that resumption is somehow special in Arabic, as our results in Experiment 2 from *whether* and adjunct structures confirm. There, even in the absence of an island, resumption is preferred to gapping, a result which can be seen as dovetailing with theoretical claims that gapping is generally marked in MSA, but in contrast to the results for Hebrew reported by Meltzer-Asscher et al. (2015). Our results suggest that a long dependency with a gap in a non-island context reduces acceptability more or less to the mean acceptability for a speaker. However, we are hesitant to conclude that this means gapping is unacceptable in MSA given that a sufficiently articulated theory does not exist that specifies the linking between acceptability and grammaticality in such a way that we can assess the status of a mean acceptability rating. Again here, further work is necessary, and we believe much could be gained from assessing these same patterns in relative clauses in addition to constituent *wh*-questions, since those constructions formed the empirical basis of initial studies

[23] This conclusion is no doubt supported by the existence of languages such as Swedish, Vata, and Gbadi, where resumptive pronouns appear in contexts which do not align with the "intrusive" versus "grammaticalized" resumption typology (see McCloskey, 2006 for discussion and references). Note that Egyptian Arabic is also claimed to allow Swedish-type resumption in limited contexts (Wahba, 1984).

into the grammatical status of resumption in Semitic (Borer, 1984; Sells, 1984; Shlonsky, 1992; Aoun et al., 2010).

4.4 Amelioration in Islands

Finally, we return to the larger question which motivated this study: does resumption ameliorate an island violation in MSA? The answer appears to be, "It depends." It depends, firstly, upon the nature of the *wh*-filler which is co-construed with the resumptive pronoun — if this is a bare *wh*-item such as "what," then resumption is simply unacceptable. Provided that a complex *wh*-phrase is used, then resumption can be supportive of otherwise very unacceptable filler-gap dependencies. It depends, also however, upon the island out of which the dependency is formed. Dependencies spanning the CNPCs employed in our study receive no help from resumption (0.39 with a gap versus 0.40 with a pronoun), a fact we have tentatively linked to the unacceptability of any filler-gap dependency with CNPC constructions more generally.

However, even within the remaining adjunct and *whether* islands, it also depends. The amelioration effect is clear in adjunct islands (no resumption at 0.78 versus resumption at 0.20), and not so clear in *whether* islands (no resumption at 0.20 versus resumption at 0.19). It therefore seems as though the kind of island is relevant to the ameliorative effect of resumptive pronouns in ways which largely mimic the extent to which the type of island affects superadditivity in general. However, we must note that even with a resumptive pronoun, adjunct island violations are still rated below the mean acceptability for these speakers. It is therefore possible to describe the ameliorative effect as "mak[ing] the best of a bad job" (Langendoen, 1970). Whether this qualifies as grammatical amelioration depends upon one's syntactic theory, but we can note that this might be the

principle underlying the occasional report that resumption ameliorates islands in English (Ackerman et al., 2015), since those studies involved force-choice tasks between gapping and resumption structures.

4.5 Conclusions

We have demonstrated that superadditivity can be observed in islands in Arabic, both with and without resumptive pronouns in the tail position of a filler-gap dependency. The amount to which quantitative superadditive acceptability penalties appear with islands in MSA depends on the type of island and possibly the proficiency with MSA of the subject population. Where resumption is present, our results show that the acceptability of resumption more generally depends on the type of filler associated with the dependency. Where these constraints are respected, resumption has been shown to increase the acceptability of dependencies spanning island boundaries, especially for adjunct islands. Where amelioration does occur, the resulting improvement nearly obliterates the superadditive component of the island violation, meaning that resumption can be seen to improve the non-processing component of acceptability penalties in Arabic.

These results help to clarify the contribution of various known components of difficult sentence processing at play in filler-gap dependency formation inside island structures, allowing us to understand the theoretical import of claims that resumption facilitates understanding and acceptability in island violations. At the same time, the resulting sentences are still well below average acceptability to speakers, leading to the conclusion that, even in a grammaticalized resumption language such as MSA, resumption still makes the best of two very bad situations — both dependencies in an island with a gap and dependencies in an island with a resumptive pronoun. The result is a picture of the grammar-processing interface

which takes islands to be a multifaceted phenomenon made up of both grammatical and processing concerns, each part of which can be manipulated independently by changes in the nature of the filler-gap or filler-pronoun dependency.

References

Ackerman, L., Frazier, M., & Yoshida, M. (2015). *Resumptive pronouns can ameliorate illicit island extractions.* (Ms., Northwestern University)

Alexopoulou, T., & Keller, F. (2007). Locality, cyclicity, and resumption: At the interface between the grammar and the human sentence processor. *Language*, *83*(1).

Almeida, D. (2014). Subliminal wh-islands in Brazilian Portuguese and the consequences for syntactic theory. *Revista de ABRALIN*, *13*(2), 55–93.

Alotaibi, M., & Borsley, R. D. (2013). Gaps and resumptive pronouns in Modern Standard Arabic. In S. Müller (Ed.), *Proceedings of the 20th international conference on head-driven phrase structure grammar* (pp. 6–23). Stanford, CA: CSLI Publications.

Aoun, J. (1981). *The formal nature of anaphoric relations* (Unpublished doctoral dissertation). Massachusetts Institute of Technology, Cambridge, MA.

Aoun, J., Benmamoun, E., & Choueiri, L. (2010). *The syntax of Arabic*. Cambridge: Cambridge University Press.

Aoun, J., & Choueiri, L. (2000). Epithets. *Natural Language and Linguistic Theory*, *18*(1), 1–39.

Aoun, J., Choueiri, L., & Hornstein, N. (2001). Resumption, movement, and derivational economy. *Linguistic Inquiry*, *32*(3), 371–403.

Asudeh, A. (2004). *Resumption as resource management* (Unpublished doctoral

dissertation). Stanford University, Palo Alto, CA.

Asudeh, A. (2011). Local grammaticality in syntactic production. In E. Bender & J. Arnold (Eds.), *Language from a cognitive perspective: Grammar, usage, and processing: Studies in honor of Thomas Wasow.* Stanford, CA: CSLI Publications.

Asudeh, A. (2012). *The logic of pronominal resumption*. Oxford: Oxford University Press.

Ayoub, G. (1981). *Structure de la Phrase en Arabe Standard* (Unpublished doctoral dissertation). Universite de Paris VII, France.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Bates, D. M., Kliegl, R., Vasishth, S., & Baayen, R. H. (2015, June). *Parsimonious mixed models.* (Ms., arXiv)

Beltrama, A., & Xiang, M. (2017). Ungrammatical but comprehensible: The facilitation effect of resumptive pronouns. *Glossa*, *1*(1).

Bennett, R. (2008). *English resumptive pronouns and the highest subject restriction.* (Handout of a talk given at the Trilateral Linguistics Weekend (TREND), University of California, Santa Cruz, 11 May 2008)

Borer, H. (1984). Restrictive relatives in modern hebrew. *Natural Language & Linguistic Theory*, *2*(2), 219–260.

Chao, W., & Sells, P. (1983). On the interpretation of resumptive pronouns. In *Proceedings of nels* (Vol. 13, pp. 47–61).

Cinque, G. (1990). *Types of A′-dependencies*. Cambridge, Mass.: MIT Press.

Clemens, L., Scontras, G., & Polinsky, M. (2012). *English does not have resumptive pronouns: A cross-sentential account of English resumption.* (Talk presented at the 87th Annual Meeting of the Linguistic Society of America)

Dermidache, H. K. (1991). *Resumptive chains in restrictive relatives, appositives, and dislocation structures* (Unpublished doctoral dissertation). Massachusetts Institute of Technology, Cambridge, MA.

Erteschik-Shir, N. (1992). Resumptive pronouns in islands. In H. Goodluck & M. Rochemont (Eds.), *Island constriants: Theory, acquisition, and processing* (pp. 89–108). Dordrecht: Kluwer.

Farby, S., Danon, G., Walters, J., & Ben-Shachar, M. (2010). The acceptability of resumptive pronouns in Hebrew. In Y. N. Falk (Ed.), *Proceedings of IATL 26.* Israeli Association for Theoretical Linguistics.

Ferreira, F., & Swets, B. (2005). The production and comprehension of resumptive pronouns in relative clause island contexts. In A. Cutler (Ed.), *Twenty-first century psycholinguistics: Four cornerstones* (pp. 101–122). Lawrence Erlbaum Associates.

Fox, D. (1999). *Relative clauses and resumptive pronouns in Hebrew: An optimality theoretic approach.* (Ms., Massachusetts Institute of Technology)

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge, UK: Cambridge University Press.

Guilliot, N., & Malkawi, N. (2006). When resumption determines reconstruction. In D. Baumer, D. Montero, & M. Scanlon (Eds.), *Proceedings of the 25th West Coast Conference on Formal Linguistics* (pp. 168–176). Somerville, MA: Cascadilla Proceedings Project.

Han, C.-H., Elouazizi, N., Galeano, C., Gorgulu, E., Helberg, N., Hinnell, J., … Kirby, S. (2012). Processing strategies and resumptive pronouns in English. In N. Arnett & R. Bennett (Eds.), *Proceedings of the 30th West Coast Conference on Formal Linguistics (WCCFL 30).* Somerville, MA: Cascadilla Proceedings Project.

Heestand, D., Xiang, M., & Polinsky, M. (2011). Resumption still does not rescue

islands. *Linguistic Inquiry*, *42*(1), 138–152.

Hofmeister, P., & Norcliffe, E. (2013). Does resumption facilitate sentence comprehension? In P. Hofmeister & E. Norcliffe (Eds.), *The core and the periphery: Data-drive perspectives on syntax inspired by Ivan A. Sag* (pp. 225–246). Stanford, CA: CSLI Publications.

Hofmeister, P., & Sag, I. A. (2010). Cognitive constraints on syntactic islands. *Language*, *86*(2), 366–415.

Keffala, B. (2011). Resumption and gaps in English relative clauses: Relative acceptability creates an illusion of 'saving'. In C. Cathcart, I. H. Chen, G. Finley, S. Kang, C. S. Sandy, & E. Stickles (Eds.), *Proceedings of the 37th annual meeting of the berkeley linguistics society* (pp. 140–154). Berkeley, California: University of California Press.

Keffala, B., & Goodall, G. (2011). *Do resumptive pronouns ever rescue illicit gaps in English?* (Poster presented at CUNY 2011 Conference on Human Sentence Processing)

Kluender, R., & Kutas, M. (1993). Subjacency as a processing phenomenon. *Language and Cognitive Processes*, *8*(4), 573–633.

Kroch, A. S. (1981). On the role of resumptive pronouns in amnestying island constraint violations. In *Papers from the 17th regional meeting of the Chicago Linguistic Society* (pp. 125–135). University of Chicago.

Langendoen, D. T. (1970). The accessibility of deep structures. In R. A. Jacobs & P. S. Rosenbaum (Eds.), *Readings in English transformational grammar* (pp. 99–104). Waltham, MA: Ginn and Company.

Malkawi, N., & Guilliot, N. (2007). Reconstruction and islandhood in Jordanian Arabic. In M. Mughazy (Ed.), *Perspectives on Arabic linguistics XX* (pp. 87–104). Amsterdam: John Benjamins.

Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing*

data: A model comparison perspective. New York: Psychology Press.

McCloskey, J. (1979). *Transformational grammar and model theoretical semantics*. Dordrecht: Reidel.

McCloskey, J. (1990). Resumption, successive cyclicity, and the locality of operations. In S. Epstein & T. D. Seely (Eds.), *Derivation and explanation* (pp. 184–226). Oxford: Blackwell.

McCloskey, J. (2006). Resumption. In M. Everaert & H. van Riemsdijk (Eds.), *The Blackwell companion to syntax* (pp. 84–117). Blackwell.

McDaniel, D., & Cowart, W. (1999). Experimental evidence for a minimalist account of English resumptive pronouns. *Cognition*, *70*(2), B15–B24.

McKee, C., & McDaniel, D. (2001). Resumptive pronouns in English relative clauses. *Language Acquisition*, *9*(2), 113–156.

Meltzer-Asscher, A., Fadlon, J., Goldstein, K., & Holan, A. (2015). Direct object resumption in Hebrew: How modality of presentation and relative clause position affect acceptability. *Lingua*, *166*, 65–79.

Pesetsky, D. (1987). Wh-in-situ: Movement and unselective binding. In E. J. Reuland & A. G. B. t. Meulen (Eds.), *The representation of (in)definiteness.* Cambridge, Massachusetts: MIT Press.

Polinsky, M., Clemens, L. E., Morgan, A. M., Xiang, M., & Heestand, D. (2013). Resumption in English. In J. Sprouse (Ed.), *Experimental syntax and island effects* (pp. 341–360). Cambridge, UK: Cambridge University Press.

Prince, E. F. (1990). Syntax and discourse: A look at resumptive pronouns. In *Proceedings of the sixteenth annual meeting of the Berkeley Linguistics Society: General session and parasession on the legacy of Grice* (pp. 482–497). Berkeley, California: Department of Linguistics, Unversity of California, Berkeley.

Rizzi, L. (1982). *Issues in Italian syntax*. Dordrecht, Holland: Foris Publications.

Ross, J. R. (1967). *Constraints on variables in syntax* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.

Ryding, K. C. (2005). *A Reference Grammar of Modern Standard Arabic*. Cambridge, UK: Cambridge UP.

Sells, P. (1984). *Syntax and semantics of resumptive pronouns* (Unpublished doctoral dissertation). University of Massachusetts, Amherst.

Shlonsky, U. (1992). Resumptive pronouns as a last resort. *Linguistic Inquiry*, *23*(3), 443–468.

Shlonsky, U. (2002). Constituent questions in Palestinian Arabic. In J. Ouhalla & U. Shlonsky (Eds.), *Themes in Arabic and Hebrew syntax* (pp. 137–160). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Sprouse, J., Caponigro, I., Greco, C., & Cecchetto, C. (2016). Experimental syntax and the variation of island effects in English and Italian. *Natural Language & Linguistic Theory*, *34*(1), 307–44.

Sprouse, J., Wagers, M., & Phillips, C. (2012). A test of the relation between working memory capacity and island effects. *Language*, *88*(1), 82–123.

Szabolcsi, A. (2006). Strong vs. weak islands. In M. Everaert & H. van Riemsdijk (Eds.), *The Blackwell companion to syntax* (Vol. Four). Malden, MA: Blackwell Publishing.

Wahba, W. A.-F. B. (1984). *Wh-constructions in Egyptian Arabic* (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign.

Zukowski, A., & Larsen, J. (2004). *The production of sentences that we fill their gaps.* (Poster Presented at the 2004 CUNY Conference on Human Sentence Processing, University of Maryland College Park)

Supplemental Materials For Tucker, Idrissi, Sprous & Almeida:

Resumption ameliorates different islands differentially:

Acceptability data from Modern Standard Arabic

A Appendix: Experimental Materials

This appendix contains the complete lists of experimental materials used in Experiments 1–2 in the text. They are presented in the following condition order for each experiment:

(1) a. SHORT, NO ISLAND
b. LONG, NO ISLAND
c. SHORT, ISLAND
d. LONG, ISLAND

Additionally, the RESUMPTION condition was created by adding a clitic pronoun to the embedded verb in each of the LONG conditions only.

A.1 Experiment 1

A.1.1 Whether Islands

(2) a. Who thinks that Mahmoud stole the wallet?
من يعتقد أن محمود سرق المحفظة؟
b. What does the policeman think Mahmoud stole?
ماذا يعتقد الشرطي أن محمود سرق؟
c. Who wonders whether Mahmoud stole the wallet?
من يتساءل ما إذا كان محمود سرق المحفظة؟
d. What does the policeman wonder whether Mahmoud stole?
ماذا يتساءل الشرطي ما إذا كان محمود سرق؟

(3) a. Who thinks that Marwa followed after the bus?
من يظنّ أن مروة لاحقت الحافلة؟
b. What does the detective think that Marwa chased after?
ماذا تظنّ المحققة أن مروة لاحقت؟
c. Who knows whether Marwa chased after the bus?
من يعرف ما إذا كانت مروة لاحقت الحافلة؟

d. What does the detective know whether Marwa chased after?
ماذا تعرف المحققة ما إذا كانت مروة لاحقت؟

(4) a. Who thinks that Muna sold the television?
من يعتقد أن منى باعت التلفاز؟
b. What does the administrator think that Muna sold?
ماذا يعتقد المدير أن منى باعت؟
c. Who wonders whether Muna sold the television?
من يتساءل ما إذا كانت منى باعت التلفاز؟
d. What does the administrator wonder whether Muna sold?
ماذا يتساءل المدير ما إذا كانت منى باعت؟

(5) a. Who thinks that Mustafa wrote the letter?
من يظنّ أن مصطفى كتب الرسالة؟
b. What does the soldier know that Mustafa wrote?
ماذا يظنّ الجندي أن مصطفى كتب؟
c. Who knows whether Mustafa wrote the letter?
من يعرف ما إذا كان مصطفى كتب الرسالة؟
d. What does the solider know whether Mustafa wrote?
ماذا يعرف الجندي ما إذا كان مصطفى كتب؟

(6) a. Who thinks that Naser bought the house?
من يعتقد أن ناصر اشترى البيت؟
b. What does the banker think that Naser bought?
ماذا يعتقد المصرفي أن ناصر اشترى؟
c. Who wonders whether Naser bought the house?
من يتساءل ما إذا كان ناصر اشترى البيت؟
d. What does the banker wonder whether Naser bought?
ماذا يتساءل المصرفي ما إذا كان ناصر اشترى؟

(7) a. Who thinks that Noor read the book?
من يظنّ أن نور قرأت الكتاب؟
b. What does the teacher think that Noor read?
ماذا يظنّ المعلم أن نور قرأت؟
c. Who knows whether Noor read the book?
من يعرف ما إذا كانت نور قرأت الكتاب؟
d. What does the teacher wonder whether Noor read?
ماذا يعرف المعلم ما إذا كانت نور قرأت؟

A.1.2 Adjunct Islands

(8) a. Who believes that the lawyer forgot his briefcase at the office?
من يعتقد أن المحامي نسي حقيبته في المكتب؟
b. What do you think that the lawyer forgot at the office?
ماذا تعتقد أن المحامي نسي في المكتب؟
c. Who worries if the lawyer forgot his briefcase at the office?
من يقلق إذا نسي المحامي حقيبته في المكتب؟

d. What do you worry if the lawyer forgot at the office?
ماذا تقلق إذا نسي المحامي في المكتب؟

(9) a. Who wishes that the prince bought a carpet at the market?
من يتمنى أن الامير اشترى سجادة في السوق؟

b. What do you wish that the prince bought at the market?
ماذا تتمنى أن الامير اشترى في السوق؟

c. Who laughs if the prince bought a carpet at the market?
من يضحك إذا اشترى الامير سجادة في السوق؟

d. What do you laugh if the prince bought at the market?
ماذا تضحك إذا اشترى الامير في السوق؟

(10) a. Who believes that the artist will paint a painting today?
من يعتقد أن الفنان سيرسم لوحة اليوم؟

b. What do you believe that the artist will paint today?
ماذا تعتقد أن الفنان سيرسم اليوم؟

c. Who blushes if the artist painted a painting today?
من يخجل إذا رسم الفنان لوحة اليوم؟

d. What do you blush if the artist painted today?
ماذا تخجل إذا رسم الفنان اليوم؟

(11) a. Who hopes that the musician will play his new song at the concert?
من يأمل أن الموسيقيّ سيعزف أغنيته الجديدة في الحفل؟

b. What do you hope that the musician will play at the concert?
ماذا تأمل أن الموسيقيّ سيعزف في الحفل؟

c. Who groans if the musician played his new song at the concert?
من يشتكي إذا عزف الموسيقيّ أغنيته الجديدة في الحفل؟

d. What do you groan if the musician played at the concert?
ماذا تشتكي إذا عزف الموسيقيّ في الحفل؟

(12) a. Who doubts that the king wrote a long speech yesterday?
من يشكّ أن الملك كتب خطابا طويلا بالأمس؟

b. What do you doubt that the king wrote yesterday?
ماذا تشكّ أن الملك كتب بالأمس؟

c. Who laughs if the king wrote a long speech yesterday?
من يضحك إذا كتب الملك خطابا طويلا بالأمس؟

d. What do you laugh if the king wrote yesterday?
ماذا تضحك إذا كتب الملك بالأمس؟

(13) a. Who claims that the friends exhanged greetings at the market?
من يدّعى أن الأصدقاء تبادلوا التحية في السوق؟

b. What do you claim that the friends exchanged at the market?
ماذا تدّعي أن الأصدقاء تبادلوا في السوق؟

c. Who becomes happy if the friends exhanged greetings at the market?
من يفرح إذا تبادل الأصدقاء التحية في السوق؟

d. What do you become happy if the friends exchanged at the market?
ماذا تفرح إذا تبادل الأصدقاء في السوق؟

A.1.3 CNPC Violations

(14) a. Who heard that Meera prepared the bread?
من سمع أن ميرا حضّرت الخبز؟
b. What did you hear that Meera prepared?
ماذا سمعتَ أن ميرا حضّرت؟
c. Who heard the rumor that Meera prepared the bread?
من سمع الإشاعة أن ميرا حضّرت الخبز؟
d. What did you hear the rumor that Meera prepared?
ماذا سمعتَ الإشاعة أن ميرا حضّرت؟

(15) a. Who denied that Ahmed ate the meat?
من أنكر أن احمد أكل اللحم؟
b. What did you deny that Ahmed ate?
ماذا أنكرتَ أن احمد أكل؟
c. Who denied the fact that Ahmed ate the meat?
من أنكر الحقيقة أن احمد أكل اللحم؟
d. What did you deny the fact that Ahmed ate?
ماذا أنكرتَ الحقيقة أن احمد أكل؟

(16) a. Who announced that Essam won the prize?
من أعلن أن عصام ربح الجائزة؟
b. What did you announce that Essam won?
ماذا أعلنتَ أن عصام ربح؟
c. Who announced the news that Essam won the prize?
من أعلن الخبر أن عصام ربح الجائزة؟
d. What did you announce the news that Essam won?
ماذا أعلنتَ الخبر أن عصام ربح؟

(17) a. Who announced that Fatima lost the game?
من أعلن أن فاطمة خسرت المباراة؟
b. What did you announce that Fatima lost?
ماذا أعلنتَ أن فاطمة خسرت؟
c. Who announced the news that Fatima lost the game?
من أعلن الخبر أن فاطمة خسرت المباراة؟
d. What did you announce the news that Fatima lost?
ماذا أعلنتَ الخبر أن فاطمة خسرت؟

(18) a. Who claimed that Habib stole the money?
من زعم أن علي سرق المال؟
b. What did you claim that Habib stole?
ماذا زعمتَ أن علي سرق؟
c. Who claimed the claim that Habib stole the money?
من نشر الادعاء أن علي سرق المال؟
d. What did you claim the claim that Ali stole?
ماذا نشرتَ الادعاء أن علي سرق؟

(19) a. Who heard that Ilyas received the award?
من سمع أن إلياس أخذ الجائزة؟

b. What did you hear that Ilyas received?
ماذا سمعتَ أن إلياس أخذ؟
c. Who heard the rumor that Ilyas received the award?
من سمع الإشاعة أن إلياس أخذ الجائزة؟
d. What did you hear the rumor that Ilyas received?
ماذا سمعتَ الإشاعة أن إلياس أخذ؟

A.2 Experiment 2

A.2.1 Whether Islands

(20) a. Who thinks that Mahmoud stole the wallet?
أيّ محفظة يعتقد الشرطي أن محموداً سرق؟
b. Which wallet does the policeman think Mahmoud stole?
أيّ محفظة يعتقد الشرطي أن محموداً سرق؟
c. Who wonders whether Mahmoud stole the wallet?
من يتساءل ما إذا كان محمود سرق المحفظة؟
d. Which wallet does the policeman wonder whether Mahmoud stole?
أيّ محفظة يتساءل الشرطي ما إذا كان محمود سرق؟

(21) a. Who thinks that Marwa followed after the bus?
من يظنّ أن مروة لاحقت الحافلة؟
b. Which bus does the detective think that Marwa chased after?
أيّ حافلة تظنّ المحققة أن مروة لاحقت؟
c. Who knows whether Marwa chased after the bus?
من يعلم ما إذا كانت مروة لاحقت الحافلة؟
d. Which bus does the detective know whether Marwa chased after?
أيّ حافلة تعلم المحققة ما إذا كانت مروة لاحقت؟

(22) a. Who thinks that Muna sold the television?
من يعتقد أن منى باعت التلفاز؟
b. Which television does the administrator think that Muna sold?
أيّ تلفاز يعتقد المدير أن منى باعت؟
c. Who wonders whether Muna sold the television?
من يتساءل ما إذا كانت منى باعت التلفاز؟
d. Which television does the administrator wonder whether Muna sold?
أيّ تلفاز يتساءل المدير ما إذا كانت منى باعت؟

(23) a. Who thinks that Mustafa wrote the letter?
من يظنّ أن مصطفى كتب الرسالة؟
b. Which letter does the soldier know that Mustafa wrote?
أيّ رسالة يظنّ الجندي أن مصطفى كتب؟
c. Who knows whether Mustafa wrote the letter?
من يعلم ما إذا كان مصطفى كتب الرسالة؟

d. Which letter does the solider know whether Mustafa wrote?
أيّ رسالة يعلم الجندي ما إذا كان مصطفى كتب؟

(24) a. Who thinks that Naser bought the house?
من يعتقد أن ناصراً اشترى البيت؟

b. Which house does the banker think that Naser bought?
أيّ بيت يعتقد المصرفي أن ناصراً اشترى؟

c. Who wonders whether Naser bought the house?
من يتساءل ما إذا كان ناصر اشترى البيت؟

d. Which house does the banker wonder whether Naser bought?
أيّ بيت يتساءل المصرفي ما إذا كان ناصر اشترى؟

(25) a. Who thinks that Noor read the book?
من يظنّ أن نوراً قرأت الكتاب؟

b. Which book does the teacher think that Noor read?
أيّ كتاب يظنّ المعلم أن نوراً قرأت؟

c. Who knows whether Noor read the book?
من يعلم ما إذا كانت نور قرأت الكتاب؟

d. Which book does the teacher wonder whether Noor read?
أيّ كتاب يعلم المعلم ما إذا كانت نور قرأت؟

A.2.2 Adjunct Islands

(26) a. Who believes that the lawyer forgot his briefcase at the office?
من يعتقد أن المحامي نسي حقيبته في المكتب؟

b. Which briefcase do you think that the lawyer forgot at the office?
أيّ حقيبة تعتقد أن المحامي نسي في المكتب؟

c. Who worries if the lawyer forgot his briefcase at the office?
من يقلق إذا نسي المحامي حقيبته في المكتب؟

d. Which briefcase do you worry if the lawyer forgot at the office?
أيّ حقيبة تقلق إذا نسي المحامي في المكتب؟

(27) a. Who wishes that the prince bought a carpet at the market?
من يتمنى أن الأمير اشترى سجادة في السوق؟

b. Which carpet do you wish that the prince bought at the market?
أيّ سجادة تتمنى أن الأمير اشترى في السوق؟

c. Who surprises if the prince bought a carpet at the market?
من يستغرب إذا اشترى الأمير سجادة في السوق؟

d. Which carpet do you surprise if the prince bought at the market?
أيّ سجادة تستغرب إذا اشترى الأمير في السوق؟

(28) a. Who believes that the artist will paint a painting today?
من يعتقد أن الفنان سيرسم لوحة اليوم؟

b. Which painting do you believe that the artist will paint today?
أيّ لوحة تعتقد أن الفنان سيرسم اليوم؟

c. Who blushes if the artist painted a painting today?
من يخجل إذا رسم الفنان لوحة اليوم؟

d. Which painting do you blush if the artist painted today?
أيّ لوحة تخجل إذا رسم الفنان اليوم؟

(29) a. Who hopes that the musician will play his new song at the concert?
من يأمل أن الموسيقيّ سيعزف أغنيته الجديدة في الحفل؟

b. Which new song do you hope that the musician will play at the concert?
أيّ أغنية جديدة تأمل أن الموسيقيّ سيعزف في الحفل؟

c. Who feels angry if the musician played his new song at the concert?
من يغضب إذا عزف الموسيقيّ أغنيته الجديدة في الحفل؟

d. Which new song do you feel angry if the musician played at the concert?
أيّ أغنية جديدة تغضب إذا عزف الموسيقيّ في الحفل؟

(30) a. Which long speech do you doubt that the king wrote yesterday?
أيّ خطاب طويل تشكّ أن الملك كتب بالأمس؟

b. Which long speech do you doubt that the king wrote it yesterday?
أيّ خطاب طويل تشكّ أن الملك كتب بالأمس؟

c. Who feels happy if the king wrote a long speech yesterday?
من يفرح إذا كتب الملك خطابا طويلا بالأمس؟

d. Which long speech do you feel happy if the king wrote yesterday?
أيّ خطاب طويل تفرح إذا كتب الملك بالأمس؟

(31) a. Who claims that the friends exhanged the gift at the market?
من يدّعى أن الأصدقاء تبادلوا الهدية في السوق؟

b. Which gift do you claim that the friends exchanged at the market?
أيّ هدية تدّعي أن الأصدقاء تبادلوا في السوق؟

c. Who becomes happy if the friends exhanged the gift at the market?
من يفرح إذا تبادل الأصدقاء الهدية في السوق؟

d. Which gift do you become happy if the friends exchanged at the market?
أيّ هدية تفرح إذا تبادل الأصدقاء في السوق؟

A.2.3 CNPC Violations

(32) a. Who heard that Meera prepared the bread?
من سمع أن ميرا حضّرت الخبز؟

b. Which bread did you hear that Meera prepared?
أيّ خبز سمعتَ أن ميرا حضّرت؟

c. Who heard the rumor that Meera prepared the bread?
من سمع الإشاعة أن ميرا حضّرت الخبز؟

d. Which bread did you hear the rumor that Meera prepared?
أيّ خبز سمعتَ الإشاعة أن ميرا حضّرت؟

(33) a. Who denied that Ahmed ate the meat?
من أخفى أن أحمداً أكل اللحم؟

b. Which meat did you deny that Ahmed ate?
أيّ لحم أخفيتَ أن أحمداً أكل؟
c. Who hid the fact that Ahmed ate the meat?
من أخفى الحقيقة أن أحمداً أكل اللحم؟
d. Which meat did you hide the fact that Ahmed ate?
أيّ لحم أخفيتَ الحقيقة أن أحمداً أكل؟

(34) a. Who announced that Essam won the prize?
من أعلن أن عصاماً ربح الجائزة؟
b. Which prize did you announce that Essam won?
أيّ جائزة أعلنتَ أن عصاماً ربح؟
c. Who announced the news that Essam won the prize?
من أعلن الخبر أن عصاماً ربح الجائزة؟
d. Which prize did you announce the news that Essam won?
أيّ جائزة أعلنتَ الخبر أن عصاماً ربح؟

(35) a. Who announced that Fatima lost the game?
من أعلن أن فاطمة خسرت المباراة؟
b. Which game did you announce that Fatima lost?
أيّ مباراة أعلنتَ أن فاطمة خسرت؟
c. Who announced the news that Fatima lost the game?
من أعلن الخبر أن فاطمة خسرت المباراة؟
d. Which game did you announce the news that Fatima lost?
أيّ مباراة أعلنتَ الخبر أن فاطمة خسرت؟

(36) a. Who claimed that Ali stole the money?
من زعم أن علياً سرق المال؟
b. Which money did you claim that Ali stole?
أيّ مال زعمتَ أن علياً سرق؟
c. Who claimed the claim that Ali stole the money?
من نشر الإدعاء أن علياً سرق المال؟
d. Which money did you claim the claim that Ali stole?
أيّ مال نشرتَ الإدعاء أن علياً سرق؟

(37) a. Who heard that Ilyas received the award?
من سمع أن إلياساً أخذ الجائزة؟
b. Which award did you hear that Ilyas received?
أيّ جائزة سمعتَ أن إلياساً أخذ؟
c. Who heard the rumor that Ilyas received the award?
من سمع الإشاعة أن إلياساً أخذ الجائزة؟
d. Which award did you hear the rumor that Ilyas received?
أيّ جائزة سمعتَ الإشاعة أن إلياساً أخذ؟