Locality, Cyclicity and Resumption: At the Interface between the Grammar and the Human Sentence Processor

Theodora Alexopoulou

UMR 8163 ‘Savoirs, Textes, Langage’

Université de Lille III

59653 Villeneuve d’Ascq Cedex, France

phone: +33-3-20-41-65-12, fax +33-3-20-41-67-14

email: theodora.alexopoulou@univ-lille3.fr

Frank Keller

School of Informatics

University of Edinburgh

2 Buccleuch Place, Edinburgh EH8 9LW, UK

phone: +44-131-650-4407, fax: +44-131-650-6626

email: keller@inf.ed.ac.uk

Language 83:1, 110–160, 2007.

Locality, Cyclicity and Resumption: At the Interface between the Grammar and the Human Sentence Processor

Abstract. We present an experimental investigation of the role of resumptive pronouns. We investigate object extraction in wh-questions for a range of syntactic configurations (non-islands, weak islands, strong islands) and for multiple levels of embedding (single, double, and triple). In order to establish the crosslinguistic properties of resumption, parallel experiments are conducted in three languages, viz. English, Greek, and German. Three main experimental results are reported. First, resumption does not remedy island violations: resumptive pronominals are at most as acceptable as gaps, but not more acceptable. This result disconfirms claims in the literature that resumptives can 'save' island violations. Second, embedding reduces acceptability even in extraction out of non-islands and declaratives, structures standardly assumed as fully grammatical. Third, non-islands and weak islands pattern together, and contrast with strong islands, in terms of the effect of resumption and embedding. Our experimental findings show a remarkable consistency across the three languages we investigate; crosslinguistic variation appears confined to quantitative differences in crosslinguistically identical principles. We argue that these experimental results can be explained by the interaction of grammatical principles with resource limitations of the human parser. In particular, extraction from non-islands and weak islands imposes increased demands on the computational resources of the parser. We extend Gibson's (1998) Syntactic Prediction Locality Theory in order to formalize this intuition and account for the processing complexity of A-bar dependencies.*

1. Introduction. Acceptability judgments form the empirical basis of linguistic theory (in particular in syntax and semantics), at least in the generative tradition. A crucial feature of some theoretically important judgments is their gradient nature, i.e. they fall somewhere between fully acceptable and fully unacceptable. This observation can be traced back to the early days of generative grammar. Indeed, an account of gradience has been viewed as a theoretical desideratum by Chomsky (1975), who argues that 'an adequate linguistic theory will have to recognize degrees of grammaticalness' (ibid., p. 131). Recent years have witnessed an increased interest in the gradient nature of acceptability judgments, mainly due to the emergence of an experimental paradigm that makes it possible to obtain reliable gradient acceptability judgments. Magnitude Estimation (Bard et al. 1996) allows subjects to indicate as many degrees of acceptability as they perceive, thus overcoming restrictions imposed by standard ordinal scales. Over the last decade, this methodology has been fruitfully applied to the study of a wide range of phenomena (see Sorace and Keller 2005 for an overview). In this article we take this approach a step further by exploring the potential of Magnitude Estimation for reliable crosslinguistic comparisons, in line with recent work by Keller and Sorace (2003), Featherston (2005), and Meyer (2003).

The phenomenon under investigation is the interaction between islands and resumption in questions. Since Ross (1967), indirect questions and relative clauses are known as 'islands', i.e. environments that cannot contain the gap of a long-distance filler-gap dependency. Indeed, in the generative tradition, sensitivity to such islands is one of the main diagnostics of movement. Resumption, i.e. the involvement of a pronominal in place of the illicit gap, has been viewed as a 'last resort' device that can 'save' island violations, by restoring structures containing illicit gaps to full acceptability. The locality conditions underlying island effects and their interaction with resumption are of central theoretical importance, given that the availability of otherwise unbounded non-local dependencies is viewed as a crucial feature of natural language. However, such conditions still resist satisfactory theoretical understanding, and their interaction with resumption has remained largely mysterious. In this article we approach these theoretical questions through a systematic experimental investigation of the interaction between locality and resumption in wh-questions. Much of the existing literature

has focused on resumptives in relative clauses, or has made no explicit distinction between questions and relative clauses.

The chosen experimental methodology allows us to quantify distinctions between 'weak' and 'strong' islands and compare them with grammatical extractions out of non-islands (e.g. non-factive that-complements). The Magnitude Estimation paradigm allows us to measure the effect of resumption in its interaction with each distinct factor, and so obtain a richer and more systematic picture of its 'saving' nature. Crucially, the crosslinguistic aspect of the experimental results reported here allow a new understanding of the nature of the crosslinguistic variation involved in these phenomena.

In what follows we first introduce the basic phenomena and the rationale behind our experimental setup. We then present the results of three experiments on object extractions in English, German, and Greek. A follow-up study on Greek object extractions is then described, investigating interactions between islands, resumption and case mismatches, as well as comparisons with declarative sentences and questions involving embedded clauses. Finally, we summarize the main results of these experiments, and then develop a theoretical analysis of the experimental data.

2. Background: locality restrictions and resumption.

2.1. Strong and weak islands. As is well known, the acceptability of the wh-questions in 1 is sensitive to the type of clause containing the gap that corresponds to the initial wh-phrase (Ross 1967). Thus, a wh-question involving a gap within a (non-factive) that-complement as in 1a is considered fully acceptable, but acceptability degrades when the same configuration involves an indirect question as in 1b; 1c, where the gap is contained in a relative clause is strongly unacceptable. Since Ross (1967), indirect questions and relative clauses (Complex NPs) are considered 'islands' for movement, i.e. environments from which movement is illicit. In fact, sensitivity to islands is taken as one of the primary diagnostics of movement. The contrast in the acceptability of 1b and 1c is acknowledged informally by referring to indirect questions as 'weak' islands and to relative clauses as 'strong' ones.

(1) a. Who does John think Mary will choose t?

b. ?*Who did Mary wonder whether they will fire t?

c. *Who did John meet the girl who will marry t?

As in 1, combinations of stars and question marks are standardly used to indicate varying degrees of acceptability. The shortcomings of this practice are discussed in detail by Bard et al. (1996), Cowart (1997), and Schütze (1996). One key problem is that the various combinations of diacritics are not systematic. Thus, for cases like 1, there is no clear understanding of the 'distance' in acceptability between 1b and 1a. For example, it seems that for authors like Chung and McCloskey (1983) the contrast between 1b and 1c is much stronger than indicated here. In fact, it would appear to be essentially equivalent to the contrast between 1a and 1c, since they judge 2 as follows (from Chung and McCloskey 1983, ex. 3a and 4d).[1]

(2) a. *Which dialects can you find speakers that linguists would agree know well?

b. Who were you wondering if we should see?

A further complication is that it is not obvious that the absence of a diacritic in 2b indicates full acceptability (on a par with 1a), rather than the absence of ungrammaticality. In fact, for Chung and McCloskey (1983), it appears that the absence of a star in 2b indicates that the sentence does not violate any grammatical principle, not that it has full acceptability.[2]

Either way, there is an important question underlying such discrepancies. Are they an artifact of the absence of an unambiguous notational system and the lack of a systematic way of quantifying linguistic intuitions, or do they represent real disagreements about the acceptability of the structures in question? We address this question here by conducting a series of Magnitude Estimation studies designed to investigate the nature of the contrasts between weak islands, strong islands and non-islands, and by quantifying the difference in acceptability between these configurations.

If establishing the acceptability status of weak islands within a language is problematic, the task becomes significantly more difficult for crosslinguistic comparisons. For example, at least two authors judge Greek examples involving extraction out of an indirect question as in 3 as grammatical (Tsimpli 1995, Alexopoulou 1999). (3b is an instance of focus-movement where the extracted phrase ta vivlia bears sentential stress, as indicated by the small capitals.) The question then is are weak island effects completely absent from Greek, i.e. are extractions

out of indirect questions as acceptable as extractions from that-complements? Or do Greek 'weak' islands involve only a mild decrease in acceptability, leading the authors to the conclusion that no grammatical violation is involved in 3? Is there any sense in which 'weak' islands in Greek are 'weaker' than in English, thus explaining the contrast between 3 and 1b?

(3) a. Pion anarotiθikes an θa apolisune?
who.acc wondered.2sg whether/if will fire.3pl
'Who did you wonder whether they will fire?'

b. Me rotise ta vivlia an epestreψa
me.acc asked.3sg the books if returned.1sg
'He asked me if I returned the books.'

2.2. Islands and resumptive pronouns. Crosslinguistically, pronouns may appear in relative clauses and questions in the place of a gap as in the Greek relative clause in 4. Such (operator/A-bar bound) pronouns are known as resumptive pronouns (Sells 1984).

(4) mia istoria pu tin eleγe i γiaγia mu otan imun mikri
a story that her said.3sg the.nom grandma my when was.1sg young
'a story my grandma used to tell when I was young'

Resumptive pronouns are excluded from simple, unembedded questions crosslinguistically, and in particular in English, Greek, and German.[3]

(5) a. Who did you fire ∅/*him?

b. Pion ∅/*ton apelises?
who.acc ∅/him fire.2sg

c. Wen entlassen wir ∅/*ihn?
who.acc fire.1pl we ∅/him

An important question is whether the star diacritic has the same meaning in all three languages. On some level it does, since it indicates that in all three languages, speakers judge such sentences as unacceptable, and linguists consider them to involve a grammatical violation. However, Greek allows resumption in embedded questions and other structures more freely than English and German (see examples 8 and section 4). The question then is are

sentences like 5 equally unacceptable in all languages?

While (non-d-linked) questions generally resist resumption, the presence of a pronominal is often viewed as a 'last resort' device 'saving' island violations (Ross 1967, Kroch 1981). The island violating examples below are 'saved' by the presence of a pronominal (examples from Haegeman 1994).

(6) a. This is the man whom$_i$ Emsworth told me when he will invite him$_i$.

b. This is the man whom$_i$ Emsworth made the claim that he will invite him$_i$.

Similar claims have been made for Greek; for instance by Merchant (2004), who offers the example in 7.

(7) O Γianis ine o adras pu i Maria efiɣe apo to parti
the.nom Giannis.nom is the.nom man.nom that the.nom Maria left.3sg from the party
otan ton iδe
when him saw.3sg
'Giannis is the man Maria left from the party when she saw him.'

While authors agree that there is an effect of resumption, the exact acceptability status of such sentences is not clear. The absence of any diacritics from the above sentences should indicate full acceptability. However, it is not obvious that this is so. The experimental research reported in subsequent sections quantifies the 'saving' effect of resumption and investigates the interaction between island violations and resumption.

Furthermore, the present study aims at a thorough crosslinguistic investigation of the issue. While resumption has a remedying effect crosslinguistically, some crosslinguistic differences exist. For example, a pronominal is acceptable in 8b in Greek, but not in 8a. Our aim, thus, is to also establish the status of crosslinguistic contrasts such as in 8 and investigate how such contrasts might relate to the general acceptability of weak islands and resumption in these languages.

(8) a. Which student did you wonder whether we'll invite *him?

b. Pion fititi anarotiθikes an θa ton kalesume?
who.acc student.acc wondered.2sg whether will him invite.1pl

Before continuing, a clarification is in order. Sells (1984) distinguishes 'intrusive' from 'true' resumption. The former is the type discussed here, i.e. resumption in island configurations appearing as a 'last resort/saving' device in place of an illicit gap/trace. True resumption involves pronominals that can be freely bound by operators, in the absence of any apparent grammatical violation. True resumption is widespread in Semitic and Celtic relative clauses.[4] We adopt this distinction here, and in the remainder the term resumption is used to indicate intrusive resumption, unless otherwise specified. (We also often use the term 'resumptive' as a short for 'resumptive pronoun'.)

2.3. Resumption and embedding. Finally, a much less discussed case of interaction between locality and resumption is pointed out by Erteschik-Shir (1992). She argues that a resumptive pronoun becomes more acceptable as the extraction site becomes more deeply embedded, a claim illustrated with the examples in 9. Dickey (1996) provides experimental evidence confirming this observation.

(9) a. This is the girl that John likes t/*her.
b. This is the girl that Peter said that John likes t/??her.
c. This is the girl that Peter said that John thinks that Bob likes t/?her.
d. This is the girl that Peter said that John thinks that yesterday his mother had given some cakes to ?t/her.

Similarly, Tsimpli (1999) argues that in Greek, a pronominal is acceptable when embedded at least one that-clause away from the matrix (compare 10 with 5b).

(10) Pion ipoptefθike i Maria oti θa ton kalesume?
who.acc suspected.3sg the.nom Maria that will him.acc invite.1pl
'Who did Maria suspect we will fire?'

Here, we investigate such effects experimentally in English, Greek, and German object extractions.

2.4. Questions vs. relative clauses. Before we turn to the presentation of the experimental results, a note is in order on why questions rather than relative clauses were chosen for this investigation, in particular since the majority of relevant observations in the theoretical

literature involve relative clauses rather than questions.

Under standard syntactic assumptions, questions and relative clauses share the same syntax (e.g. instances of A-bar dependencies, possibly both involving a quantificational operator in the sense of Lasnik and Stowell 1991). But there are some important empirical differences, not always accommodated theoretically, which make questions more suitable for an initial investigation.

First, and perhaps most importantly, certain corpus studies have indicated the very real existence of resumption in English relative clauses outside island environments (Prince 1990, 1997, Cann et al. 2005). We are not aware of any such studies for questions. The general unavailability of resumption in questions allows us to better isolate interactions between resumption and embedding/islands.

Second, there are a number of factors interacting with resumption in relative clauses that are absent from questions. For instance, Prince (1990) found an interaction between resumption and the definiteness of the head (see also Suñer 1998): of 100 relative clauses in her corpus,[5] 84 appeared in relative clauses headed by an indefinite (gap relative clauses headed by an indefinite were 76). At the same time, the theoretical literature on Greek takes pronominals to be unacceptable in definite relative clauses but optional in indefinite ones (Alexiadou and Anagnostopoulou 2000, Tsimpli 1999). While worthy of investigation, such effects make it harder to pin down the interaction between resumption and islands in relative clauses.[6]

Finally, resumption interacts with whether or not a relative clause is introduced by a relative pronoun or a complementizer (Joseph 1980b, 1983); thus, in Greek pu-relative clauses resumption is obligatory in oblique positions (indirect object, possessor).[7] Such interactions are absent from questions, which therefore make it possible to focus on interactions of resumption with locality.

In sum, while the study of relative clauses (as well as d-linked questions) is important for understanding interactions of resumption and locality, questions are more suitable for an initial investigation due to the general unavailability of resumption in these structures. Indeed there is evidence that the results reported here do not extend to relative clauses, as indicated by

McDaniel and Cowart's (1999) Magnitude Estimation study investigating the interaction between that-trace violations, weak islands and resumption in English relative clauses. We discuss this study in more detail section 8.4.

2.5. Magnitude Estimation. The present study relies on subtle linguistic intuitions regarding the relative acceptability of gaps and resumptive pronouns in extraction. As discussed in section 2, the standard practice of collecting informal judgments has yielded contradictory data for the phenomenon at hand, and is inadequate for crosslinguistic comparisons. To address this situation, we have employed Magnitude Estimation (ME), an experimental paradigm, designed to overcome the shortcomings of informal collection of judgments (Schütze 1996, Bard et al. 1996, Cowart 1997). In this section we briefly present the main features of ME.

Magnitude Estimation is a technique developed in psychophysics to measure judgments of sensory stimuli (Stevens 1975). In particular, ME was developed to determine to which extent subjects can reliably indicate proportional judgments corresponding to degrees of magnitude in perceived physical stimuli, e.g. whether subjects can reliably indicate not just that a light is more or less bright than a reference stimulus, but how many times brighter. The key feature of ME is that it makes it possible to investigate proportional judgments by employing a continuous numerical scale. This allows subjects to indicate as many sensory distinctions as they perceive; in this respect it overcomes problems associated with 5- or 7-point ordinal scales conventionally used to measure human intuitions. ME studies have yielded highly reliable judgments for a whole range of sensory modalities, such as brightness, loudness, or tactile stimulation, demonstrating that generally equal ratios on the physical dimensions give rise to equal ratios of judgments; for example, in judgments of brightness, every time the stimulus energy doubles, the subjective brightness becomes 1.5 times larger; in judgments of line length, doubling physical line length doubles subjective line length (psychophysical relationships can then be viewed as a set of mathematical functions). The ME paradigm has been extended successfully to the psychosocial domain (Lodge 1981), e.g. for investigating notions like social prestige. Bard et al. (1996) and Cowart (1997) have shown that ME can also be applied to linguistic judgments, and that it provides fine-grained

measurements of acceptability, which are robust enough to yield statistically significant results, while being highly replicable both within and across speakers.

In linguistic ME, subjects are asked to assign numbers to a series of linguistic stimuli proportional to the acceptability they perceive, according to the following procedure. First, they see a set of instructions that explain the concept of numerical ME using line length. Several example lines and corresponding numerical estimates are provided to illustrate the concept of proportionality. After reading the instructions, subjects take part in a training phase designed to familiarize them with the task. In the training phase, subjects are asked to use ME to judge the length of a set of lines. In particular, they see a reference line and are told to give it an arbitrary number; then, they are asked to assign a number to each following line representing how long the line is in proportion to the reference line; if the line is twice as long as or half the reference line, the number they assign should be twice or half the number of the reference line, etc. Subjects are told that linguistic acceptability can be judged in the same way as line length. A set of practice items involving examples of sentences of varying acceptability is used to illustrate the task. Finally, subjects judge the experimental items. Subjects first see a modulus (reference item), to which they assign an arbitrary number. Then, all other stimuli are rated proportional to the modulus, i.e. if a sentence is three times as acceptable as the modulus, it gets three times the modulus number, etc. Subjects are randomly assigned to stimulus sets, and the stimuli in a given stimulus set are presented in random order (a new order is generated for each subject).

All ME data presented here were normalized and log-transformed before being graphed or subjected to statistical tests. Normalization is necessary as experimental subjects can freely choose the number they assign to the modulus sentence. Dividing each numeric judgment by the modulus judgment therefore creates a common scale for all subjects. Then the data are transformed by taking the decadic logarithm. This transformation ensures that the judgments are normally distributed and is standard practice for ME data (Bard et al. 1996).

Normalized, log-transformed data can be interpreted as follows: the modulus sentence receives a judgment of zero. Sentences with a positive judgment are more acceptable than the modulus, while sentences with a negative judgment are less acceptable than the modulus. Note

that the modulus (in line with recommendations in the ME literature) was selected to be a sentence of intermediate acceptability.

It is important to note that the absolute judgments obtained in a ME experiment are not meaningful. All judgments have to be interpreted relative to the modulus (or relative to the judgments for another sentence in the same experiment). The reason for this is that a range of irrelevant factors can vary between experiments (e.g. the number and acceptability of filler sentences). It follows that it is not possible to compare numeric judgments across experiments, or to assign labels such as 'grammatical' or 'ungrammatical' to sentences based on the absolute judgments obtained in a magnitude estimation experiment.

In the experiments here, we deal with this by including control conditions in all our experiments. These conditions consist of sentences of known acceptability status (e.g. fully acceptable or fully unacceptable items). We then compare the other sentences in the experiment to the controls and hence indirectly draw conclusions about their acceptability status.

3. Experiment 1: resumptives and object extraction in English. The first experiment investigated how embedding and island constraints interact with resumption in English. Four different configurations were used: complement clause without that (non-island), non-factive complement clause with that (non-island), complement clause with whether (weak island), relative clause (strong island). Two levels of embedding were tested: single embedding (one complement clause or relative clause) and double embedding (one that-complement clause intervening between the wh-phrase and the that- or whether-complement clause or a relative clause). To have a standard of comparison, we also included sentences without embedding (control condition, zero embedding). Example sentences are given in 11–14.

(11) Non-island condition (bare clause)

a. Who will we fire ∅/him? (zero embedding)

b. Who does Mary claim we will fire ∅/him? (single)

c. Who does Jane think Mary claims we will fire ∅/him? (double)

(12) Non-island condition (that-clause)

a. Who does Mary claim that we will fire ∅/him? (single)

b. Who does Jane think that Mary claims that we will fire ∅/him? (double)

(13) Weak island condition (whether-clause)

a. Who does Mary wonder whether we will fire ∅/him? (single)

b. Who does Jane think that Mary wonders whether we will fire ∅/him? (double)

(14) Strong island condition (relative clause)

a. Who does Mary meet the people that will fire ∅/him? (single)

b. Who does Jane think that Mary meets the people that will fire ∅/him? (double)

3.1. Method.

Subjects. Fifty-five subjects were recruited over the Internet by postings to newsgroups and mailing lists. All subjects were self-reported native speakers of English. Linguists and students of linguistics were excluded from the sample.

Materials. The design crossed the following factors: Embedding (single or double embedding), Island (complement clause without that, complement clause with that, complement clause with whether, relative clause), and Resumption (gap or resumptive). This resulted in Embedding $\times$ Island $\times$ Resumption $= 2 \times 4 \times 2 = 16$ cells. As controls, we included stimuli without embedding (gap or resumptive), increasing the number of cells to 18. Nine lexicalizations were used for each cell, yielding a total of 162 stimuli.

The stimulus set was divided into nine subsets of 18 stimuli by placing the items in a Latin square. A set of 18 fillers was used, covering the whole acceptability range.

Procedure. The method used was Magnitude Estimation as proposed by Bard et al. (1996) and Cowart (1997) and described in section 2.5.

After reading the instructions and before proceeding to the training phase, subjects had to fill in a short demographic questionnaire, which included age, sex, handedness, and language region, which was defined as the place where the subject learned his or her first language. After the training and practice phase, each subject judged one set of 18 experimental stimuli and all 18 fillers, i.e. a total of 36 items. Subjects were randomly assigned to stimulus sets; this assignment was slightly unbalanced, as the number of subjects was not a multiple of the

number of stimulus sets. The stimuli in a given stimulus set were presented in random order; a new order was generated for each subject.

Keller and Alexopoulou (2001) present a detailed discussion of the safeguards that WebExp puts in place to ensure the authenticity and validity of the data collected, and also present a validation study comparing web-based and lab-based judgment data (for other validation studies, see Keller and Asudeh 2001, Corley and Scheepers 2002).

3.2. Results. All data were normalized and log-transformed as described in section 2.5. Figure 1 graphs the mean judgments with standard errors for all four configurations. In the following, we only report the qualitative results for this experiment (i.e. the significant differences obtained). The details of the statistical analyses can be found in Appendix A.

[Figure 1 about here.]

An Analysis of Variance (Anova) was carried out to determine which of the experimental factors had a significant effect on acceptability, and to establish any significant interactions between the factors. The Anova yielded significant main effects of Embedding, Island, and Resumption. The interaction of Island and Resumption was also significant. This is the most relevant interaction in this study, as it indicates that the acceptability of resumptives is sensitive to island violations. We also found a significant interaction of Embedding and Resumption; the other interactions were only significant by subjects.

A Tukey post-hoc test was conducted to further investigate the interaction of Island and Resumption. This test allows us to determine in which of the island conditions a gap is more acceptable than a resumptive. The results show that gaps are significantly more acceptable than resumptives for bare clauses, that-clauses, and whether-clauses. For relative clauses, no significant difference was found.

As a next step, a series of Dunnett tests was carried out to determine if the single and double embedding conditions were significantly different from the unembedded control condition. Table 1 lists the results of the Dunnett tests.

[Table 1 about here.]

3.3. Discussion. The first important finding is that embedding per se has an effect on the acceptability of wh-questions even in the non-island condition (extraction out of that-clauses). This effect is only a tendency in the first level of embedding, but reaches statistical significance in the case of doubly embedded questions. This is unexpected from a grammar point of view, since no grammatical violation is associated with these structures. However, this finding is consistent with results in the psycholinguistic literature. For example, Frazier and Clifton (1989) report results from a self-paced reading task of similar sentences indicating processing difficulty; similar findings have been reported more recently by Dickey (1996) and Kluender (1998). We discuss these studies in section 8.2.

As expected, extraction from weak islands (whether-clauses) was less acceptable than extraction from non-islands (bare clauses and that-clauses): we found a significant difference between zero embedding and single embedding for whether-clauses, but not for bare clauses and that-clauses. For all three types of extraction, we also found that double embedding was significantly worse than zero embedding. For relative clauses, on the other hand, we found that single embedding was seriously unacceptable, on a par with resumptive pronouns; double embedding did not reduce acceptability further.

Turning finally to the issue of resumption, pronominals were found to be significantly worse than gaps in all conditions (with the exception of extraction from relative clauses). However, we found that resumption appears to reverse the effect of embedding: in bare clauses doubly embedded resumptives were significantly more acceptable than the unembedded control (by subjects only). In that-clauses, singly embedded resumptives were significantly more acceptable than the control (by subjects only). As we see in subsequent sections, this effect is also present in Greek and German.

4. Experiment 2: resumptives and object extraction in Greek. The purpose of the present experiment was to test the crosslinguistic validity of our findings for English and the potential influence of some structural differences between English and Greek. In particular, Greek differs from English in the following ways: (i) indirect questions in Greek are not considered islands (see section 2.1); (ii) a pronominal is acceptable in wh-questions when embedded at least one that-clause away from the matrix (see section 2.3) or in an embedded question (see

section 2.2); (iii) unlike English, Greek exhibits productive resumption in Clitic Left Dislocation (clld) as in 15.

(15) to Γiani to sinadisame stin aγora
the.acc Γiani.acc him met.1pl at-the market
'We met Γianis at the market.'

The stimuli used were analogous to the ones in Experiment 1 for English. However, there was no bare clause condition, as the complementizer is obligatory in Greek complement clauses of this type.[8] The following are example materials that illustrate the types of sentences used. (In Greek the pronominal is a clitic, attached to the left of the verb. Gaps, marked by ∅ below, are indicated preverbally only for convenience. No claims are made for the position of the trace corresponding to the gap.)

(16) Non-island condition (that-clause)

a. Pion θa ∅/ton apolisume? (zero embedding)
who.acc will ∅/him fire.1pl
'Who will we fire?'

b. Pion isxirizete i Anna oti θa ∅/ton apolisume? (single)
who.acc claim.3sg the.nom Anna that will ∅/him fire.1pl
'Who does Anna claim that we will fire?'

c. Pion nomizi o Petros oti isxirizete i Eleni oti θa ∅/ton apolisume? (double)
who.acc think.3sg the.nom Petros.nom that claim.3sg the.nom Eleni that will ∅/him fire.1pl
'Who does Petros think that Eleni claims that we will fire?'

(17) Weak island condition (whether-clause)

a. Pion anarotiete i Maria an θa ∅/ton apolisume? (single)
who.acc wonder.3sg the.nom Maria whether will ∅/him fire.1pl
'Who does Maria wonder whether we will fire?'

b. Pion nomizi o Petros oti anarotiete i Maria an
who.acc think.3sg the.nom Petros.nom that wonder.3sg the.nom Maria whether

θa ∅/ton apolisume? (double)

will ∅/him fire.3pl

'Who does Petros think that Maria wonder whether we will fire?'

(18) Strong island condition (relative clause)

a. Pion sinadai i Eleni tus tipus pu θa ∅/ton

who.acc meet.3sg the.nom Eleni the.acc guys.acc that will ∅/him

apolisun? (single)

fire.3pl

'Who does Eleni meet the guys that will fire?'

b. Pion nomizi o Petros oti sinadai i Eleni tus

who.acc think.3sg the.nom Petros.nom that meet.3sg the.nom Eleni the.acc

tipus pu θa ∅/ton apolisun? (double)

guys.acc that will ∅/him fire.3pl

'Who does Petros think that Eleni meets the guys that will fire?'

Our focus here is on intrusive resumption. However, given examples like 15, it's worth asking whether intrusive and non-intrusive resumption[9] can be reliably distinguished in Greek, since d-linking can improve the acceptability of resumption in unembedded questions.[10] We discuss this issue in more detail in section 8.2. For the moment we present our basic assumptions: (i) simple unembedded questions with a pronominal involve ungrammatical cases of putative true resumption; the main reason for this assumption is the grammaticality contrast between examples 15 and 16a; crucially, resumption is acceptable in clld in non-embedded contexts; (ii) examples like 16b–17 involve intrusive resumption, again due to the contrast with unembedded questions, which is absent from clld, i.e. at least under standard assumptions, there is no acceptability contrast between embedded and unembedded examples of clld as in 19a and 19b. If anything, acceptability degrades when a weak island is involved as in 19c. Unlike clld, pronominals embedded in questions are better than unembedded ones according to the theoretical literature and the results to be reported below.[11]

(19) a. to Γiani to sinadisame stin aγora

the.acc Γiani.acc him met.1pl at-the market

‘We met Γianis at the market.’

b. to Γiani ipame sti Maria oti to sinadisame stin aγora
the.acc Γiani.acc said.1pl to-the Maria that him met.1pl at-the market
‘We told Maria that we met Γianis at the market.’

c. to Γiani mas rotise i Maria an to sinadisame stin aγora
the.acc Γiani.acc us asked.3sg the Maria.nom if him met.1pl at-the market
‘Maria asked us whether we met Γianis at the market.’

4.1. Method. Fifty-nine subjects were recruited over the Internet by postings to newsgroups and mailing lists. All subjects were self-reported native speakers of Greek. Linguists and students of linguistics were excluded from the sample.

The design mirrored the one for English, but left out the no-that-clause condition (the complementizer is obligatory in Greek). This resulted in Embedding $\times$ Island $\times$ Resumption $= 2 \times 3 \times 2 = 12$ cells. As controls, we included stimuli without embedding (gap or resumptive), increasing the number of cells to 14. Seven lexicalizations were used for each cell, yielding a total of 98 stimuli. The stimulus set was divided into seven subsets of 14 stimuli by placing the items in a Latin square. A set of 14 fillers was used, covering the whole acceptability range.

The same procedure as in Experiment 1 was used. Instructions were presented in Greek.

4.2. Results. The data were normalized and log-transformed as in Experiment 1. Figure 2 graphs the mean judgments for all three configurations. For details of the statistical analyses, see Appendix A.

[Figure 2 about here.]

An Anova yielded significant main effects of Embedding, Island, and Resumption. The interaction of Island and Resumption was also significant, which indicates that the acceptability of resumptives is sensitive to island violations. The interactions Island/Embedding, Embedding/Resumption, and Island/Embedding/Resumption were only significant by subjects.

A post-hoc Tukey test was conducted to further investigate the interaction of Island and Resumption. Its results show that gaps are significantly more acceptable than resumptives for that-clauses, but not for whether- and relative clauses. A series of Dunnett tests were conducted to compare the single and double embedding conditions with unembedded control conditions; see Table 2.

[Table 2 about here.]

4.3. Discussion. As in English, we found that embedding has an effect in that-clauses, which in the case of Greek is significant for both levels of embedding (see Figure 2(a)). Furthermore, as in English, weak island violations (extraction out of whether-clauses) led to a significant drop in acceptability (see Figure 2(b)). Greek and English are not different in this respect, a finding contrasting with the theoretical literature where weak island violations such as 3 are considered fully acceptable. In particular, gaps singly embedded in a whether-clause are worse than gaps singly embedded in a that-clause (as in English). Strong island violations (extraction out of a relative clause) were found to lead to strong unacceptability (as expected, and like in English). No effect of double embedding was detected in relative clauses.

As in English, pronominals are unacceptable in simple Greek questions. However, this violation appears to be less serious in Greek. Greek unembedded pronominals, while significantly worse than gaps, are still significantly better than strong island violations (see Figure 2(c)). By contrast, unembedded pronominals in English questions are as unacceptable as strong island violations (see Figure 1(d)). Under the assumption that extraction out of relative clauses is equally unacceptable in both English and Greek, the higher acceptability of Greek unembedded pronominals compared to relative clause islands indicates that pronominals are more tolerated in Greek than English.

For Greek that-clauses, we failed to find a significant difference between singly or doubly embedded resumptives and the unembedded control condition (zero embedding). However, for whether-clauses, pronominals were significantly more acceptable in the embedded conditions (single and double) compared to the control condition. We also found (as in English), that resumption fails to interact with (i.e. reverse the effect of) strong island violations; unlike embedded pronominals in that- and whether-clauses, pronominals in relative clauses were

worse than in the control condition.

Finally, it is worth noting that doubly embedded pronominals in that-clauses and singly and doubly embedded pronominals in whether-clauses are as acceptable as gaps. This optionality between the gap and the pronominal has been noted in the theoretical literature on Greek, which, however, fails to note their degraded status in comparison to unembedded controls (see sections 2.2 and 2.3).[12] Note finally that Greek differs from English in that pronominals remain significantly worse than gaps in the latter in all conditions (except strong island violations).

5. Experiment 3: resumptives and object extraction in German. The aim of the present experiment was to test the crosslinguistic validity of Experiments 1 and 2 by investigating the interaction of the factors embedding, resumption, and island violation in German. In 20–22, we list an example stimulus for each experimental condition. These stimuli are closely parallel to the English ones in 11–14.

(20) Non-island condition (that-clause)

a. Wen entlassen wir ∅/ihn? (zero embedding)
who.acc fire.1pl we ∅/him
'Who will we fire?'

b. Wen behauptet Petra, dass wir ∅/ihn entlassen? (single)
who.acc claim.3sg Petra that we ∅/him fire.1pl
'Who does Petra claim that we will fire?'

c. Wen denkt Barbara, dass Petra behauptet, dass wir ∅/ihn entlassen? (double)
who.acc think.3sg Barbara that Petra claim.3sg that we ∅/him fire.1pl
'Who does Barbara think that Petra claims that we will fire?'

(21) Weak island condition (whether-clause)

a. Wen überlegt Petra, ob wir ∅/ihn entlassen? (single)
who.acc ponder.3sg Petra whether we ∅/him fire.1pl
'Who does Petra ponder whether we will fire?'

b. Wen denkt Barbara, dass Petra überlegt, ob wir ∅/ihn
who.acc think.3sg Barbara that Petra ponder.3sg whether we ∅/him
entlassen? (double)
fire.1pl
'Who does Barbara think that Petra ponders whether we will fire?'

(22) Strong island condition (relative clause)

a. Wen trifft Petra die Leute, die ∅/ihn entlassen? (single)
who.acc meet.3sg Petra the people that ∅/him fire.3pl
'Who does Petra meet the people that will fire?'

b. Wen denkt Barbara, dass Petra die Leute trifft, die ∅/ihn
who.acc think.3sg Barbara that Petra the people meet.3sg that ∅/him
entlassen? (double)
fire.3pl
'Who does Barbara think that Petra meets the people that will fire?'

5.1. Method. Thirty-seven subjects were recruited over the Internet by postings to newsgroups and mailing lists. All subjects were self-reported native speakers of German. Linguists and students of linguistics were excluded from the sample.

The experimental design and distribution of materials was identical to that in Experiment 2. The same procedure as in Experiments 1 and 2 was used. Instructions were presented in German.

5.2. Results. The data were normalized and log-transformed as in Experiments 1 and 2. Figure 3 graphs the mean judgments for all three configurations. For details of the statistical analyses, see Appendix A.

[Figure 3 about here.]

An Anova yielded a significant main effect of Island, but the main effects of Embedding, and Resumption were not significant. The interaction of Island and Resumption was also significant, which indicates that the acceptability of resumptives is sensitive to island

violations. The interaction Island/Embedding was also significant; all other interactions failed to reach significance.

A post-hoc Tukey test was conducted to further investigate the interaction of Island and Resumption. No significant difference between the acceptability of gaps and resumptives was found for any of the three clause types (that-clause, whether-clause, relative clauses). Furthermore, we carried out a series of Dunnett tests to compare the single and double embedding conditions with unembedded control conditions; see Table 3.

[Table 3 about here.]

5.3. Discussion. The experimental results for German[13] show that embedding reduces the acceptability of gapped clauses. This is true for all three clause types, i.e. even for that-clauses. We also found that resumption can reverse the effect of embedding; in that- and whether-clauses, embedded resumptives were more acceptable than the unembedded control. No such effect was found for relative clauses. Crucially, however, resumptives were never more acceptable than gaps; they were at most equally acceptable. All these experimental findings replicate what we have found for English and Greek in the previous two experiments.

There are some crosslinguistic differences, however. In English and Greek, we found that extraction from that-clauses was more acceptable than extraction from whether-clauses. For German, the two clause types are similar in acceptability (see Figure 3(a) and 3(b)). Furthermore, we found that gaps and resumptives can be equally acceptable. German seems to behave like Greek (and unlike English) in this respect (this observation has not been made in the theoretical literature on German, as far as we are aware). Another crosslinguistic difference concerns unembedded resumptives: in German, they are as unacceptable as strong island violations. This is like English, but differs from Greek, where unembedded resumptives were more acceptable than strong island violations.

Taking Experiments 1–3 together, the most interesting overall finding is that the acceptability patterns are basically the same across all three languages, which indicates that the principles underlying these phenomena are crosslinguistically constant. This result demonstrates the importance of employing an experimental methodology for identifying crosslinguistic universals and the locus of crosslinguistic variation. For example, in English,

the effect of resumption in that- and whether-clauses is of the same nature as in Greek and German, but manifests itself in a smaller reduction in unacceptability (compared to the unembedded control). It is unlikely that this fact would have been revealed by the standard informal collection of judgments, given that the acceptability of resumptives remains worse than the acceptability of gaps.

6. Experiment 4: case violation and object extraction in Greek. Experiments 1–3 demonstrated that resumption reverses the effect of embedding and weak islands, but it cannot restore the affected structures to full acceptability. We also found a crosslinguistic effect of embedding; in all three languages, doubly and singly embedded sentences are less acceptable than unembedded ones. This effect was modulated by the type of the embedded clause (non-island, weak island, strong island).

The present experiment has three aims. First, we wanted to further study the embedding effect by including a triple embedding condition and identifying whether it induces a further reduction in acceptability, compared with the double embedding condition studied in Experiments 1–3. Second, we wanted to establish whether embedding can lead to a reduction in acceptability outside the context of wh-extraction. We therefore tested the acceptability of multiply embedded sentences in declaratives.

Third, this experiment investigates possible interactions between embedding and a core morpho-syntactic violation like case mismatch. Experiments 1–3 established an interaction between resumption and embedding in that-clauses, raising the hypothesis that resumption interacts with a processing constraint on embedding. The present experiment studies the interaction between embedding, resumption, and case mismatch with the aim of testing the possibility that case mismatch will be less noticed in more embedded positions.

The materials in the present experiment are an extension of the stimuli for Experiment 2. Examples for the zero, single, and double embedding conditions are given in 16–18. 23 and 24 provide examples for triple embedding in the non-island and the weak island condition, respectively. Note that no strong island condition was included; it was considered redundant, as no variation was found in this condition in the previous experiments.

(23) Pion nomizi o Petros oti pistevi o Kostas oti
who.acc think.3sg the.nom Petros.nom that believe.3sg the.nom Kostas.nom that
θeori i Eleni oti θa ∅/ton apolisume?
speculate.3sg the.nom Eleni that will ∅/him fire.1pl
'Who does Petros think that Kostas believes that Eleni speculates that we will fire?'

(24) Pion nomizi o Petros oti pistevi i Sofia oti anarotiete
who.acc think.3sg the.nom Petros.nom that believe.3sg the.nom Sofia that wonder.3sg
i Maria an θa ∅/ton apolisume?
the.nom Maria whether will ∅/him fire.1pl
'Who does Petros think that Sofia believes that Maria wonders whether we will fire?'

For the declarative control condition, we used declarative versions of examples like 16–18, 23, and 24, as illustrated in 25–27.

(25) Declarative (that-clause)

a. Θa apolisume to Γiani. (zero embedding)
will fire.1pl the.acc Γiannis.acc
'We will fire Γiannis.'

b. I Anna θeori oti θa apolisume ton Petro. (single)
the.nom Anna speculate.3sg that will fire.1pl the.acc Petros.acc
'Anna speculates that we will fire Petros.'

c. O Petros nomizi oti i Eleni θeori oti θa
the.nom Petros.nom think.3sg that the.nom Eleni speculate.3sg that will
apolisume ti Maria. (double)
fire.1pl the.acc Maria
'Petros thinks that Anna speculates that we will fire Maria.'

d. O Takis nomizi oti i Anna pistevi oti i Eleni
the.nom Takis.nom think.3sg that the.nom Anna believe.3sg that the.nom Eleni
θeori oti θa apolisume ti Maria. (triple)
speculate.3sg that will fire.1pl the Maria
'Takis thinks that Anna believes that Eleni speculates that we will fire Maria.'

(26) Declarative (whether-clause)

a. I Maria anarotiete an θa apolisume ton Taki. (single)
the.nom Maria speculate.3sg whether will fire.1pl the.acc Takis.acc
'Maria speculates whether we will fire Takis.'

b. O Petros nomizi oti i Maria anarotiete an θa
the.nom Petros.nom think.3sg that the.nom Maria speculate.3sg whether will
apolisume to Γiorγo. (double)
fire.1pl the.acc Γiorγos.acc
'Petros thinks that Maria speculates whether we will fire Γiorγos.'

c. O Petros nomizi oti i Sofia pistevi oti i Maria
the.nom Petros.nom think.3sg that the.nom Sofia believe.3sg that the.nom Maria
anarotiete an θa apolisume to Γiorγo. (triple)
speculate.3sg whether will fire.1pl the.acc Γiorγos.acc
'Petros thinks that Sofia believes that Maria speculates whether we will fire Γiorγos.'

(27) Declarative (relative clause)

a. i Eleni sinadai tus tipus. (zero)
the.nom Eleni meet.3sg the.acc guys.acc
'Eleni meets the guys.'

b. i Eleni sinadai tus tipus pu θa apolisun to Γiorγo. (single)
the.nom Eleni meet.3sg the.acc guys.acc that will fire.3pl the.acc Γiorγos.acc
'Eleni meets the guys who will fire Γiorγos.'

c. O Petros nomizi oti i Eleni sinadai tus tipus pu
the.nom Petros.nom think.3sg that the.nom Eleni meet.3sg the.acc guys.acc that
θa apolisun ton Taki. (double)
will fire.3pl the Takis.acc
'Petros thinks that Eleni meets the guys who will fire Takis.'

d. O Petros nomizi oti i Sofia pistevi oti i Eleni
the.nom Petros.nom think.3sg that the.nom Sofia believe.3sg that the.nom Eleni

sinadai tus tipus pu θa apolisun ton Taki. (triple)
meet.3sg the.acc guys.acc that will fire.3pl the.acc Takis.acc
'Petros thinks that Sofia believes that Eleni meets the guys who will fire Takis.'

Furthermore, this experiment included materials for the case mismatch condition, where the case marking of the wh-phrase fails to match the accusative case required by its subcategorizing verb. Two conditions were included: one in which the wh-phrase bears genitive (instead of accusative) case marking, and one in which it consists of a prepositional phrase. The following examples illustrate this (we list only the single embedding condition):

(28) Non-island condition (that-clause, case mismatch)

a. Pianu θeori i Anna oti θa ∅/ton apolisume?
who.gen speculate.3sg the.nom Anna that will ∅/him fire.1pl
'Who does Anna speculate that we will fire?'

b. Se pion θeori i Anna oti θa ∅/ton apolisume?
to who.acc speculate.3sg the.nom Anna that will ∅/him fire.1pl
'To who does Anna speculate that we will fire?'

(29) Weak island condition (whether-clause, case mismatch)

a. Pianu anarotiete i Maria an θa ∅/ton apolisume?
who.gen wonder.3sg the.nom Maria whether will ∅/him fire.1pl
'Who does Maria wonder whether we will fire?'

b. Se pion anarotiete i Maria an θa ∅/ton apolisume?
to who.acc wonder.3sg the.nom Maria whether will ∅/him fire.1pl
'To who does Maria wonder whether we will fire?'

6.1. Method. Thirty-three subjects from the same population as in Experiment 2 participated in the experiment. None of the subjects had taken part in the earlier study.

The experiment included three subexperiments. The first one replicated Experiment 2, but with up to three levels of embedding, and with declarative controls. The design crossed the factors Embedding (single, double, triple), Island (that-clause, whether-clause), Resumption (gap, resumptive, declarative). This resulted in

Embedding $\times$ Island $\times$ Resumption $= 3 \times 2 \times 3 = 18$ cells. In addition, unembedded control conditions were included, which added three cells (gap, resumptive, declarative).

The second subexperiment augmented the first subexperiment by including declarative versions of the relative clause stimuli from Experiment 2. These additional controls allow us to test if declarative relative clauses behave differently from declarative that- and whether-clauses. This subexperiment had four cells (zero, singly, doubly, triply embedded relative clauses).

The third subexperiment tested the effect of a case mismatch between the gap or resumptive and the wh-phrase antecedent. The correct case for the antecedent is accusative; we tested two mismatch conditions: genitive antecedent and prepositional phrase antecedent (see 28 and 29 for examples). The design crossed the factors Embedding (single, double), Island (that-clause, whether-clause), Resumption (gap, resumptive), and Case (genitive, prepositional), resulting in
Embedding $\times$ Island $\times$ Resumption $\times$ Antecedent $= 2 \times 2 \times 2 \times 2 = 16$ cells. Again, unembedded controls were included, which added Resumption $\times$ Antecedent $= 2 \times 2 = 4$ cells to the design.

Taken together, the three subdesigns had a total of 45 cells. Nine lexicalizations were used for each cell, yielding a total of 405 stimuli. The stimulus set was divided into nine subsets of 45 stimuli by placing the items in a Latin square. A set of 45 fillers was used, covering the whole acceptability range.

The same procedure as in Experiments 1–3 was used. Instructions were presented in Greek.

6.2. Results. The data were normalized and log-transformed as in Experiments 1–3. For details of the statistical analyses, see Appendix A.

Subexperiment 1. Figure 4 graphs the mean judgments for this subexperiment. An Anova yielded significant main effects of Island (by subjects only), Embedding, and Resumption. The interaction of Island and Resumption was significant, indicating that the acceptability of resumption is sensitive to island violations. The interaction Embedding/Resumption was also

significant, but all other interactions failed to reach significance.

A post-hoc Tukey test was conducted to further investigate the interaction of Island and Resumption. Its results show that gaps are more acceptable than resumptives, and that declaratives are more acceptable than gaps and resumptives. These results hold for both that- and whether-clauses. We also carried out a series of Dunnett tests to compare the single, double, and triple embedding conditions with unembedded controls; see Table 4.

[Figure 4 about here.]

[Table 4 about here.]

Subexperiment 2. Figure 5 graphs the mean judgments for this subexperiment. We conducted an Anova that included the declarative stimuli from the first subexperiment, so as to be able to determine if different types of declaratives show a differential effect of embedding. The Anova therefore crossed the factors Embedding (single, double, triple) and Type (that-clause, whether-clause, relative clauses). A significant main effect of Embedding was found, but there was no main effect of Type, and no interaction between the two factors.

A post-hoc Tukey test on the main effect of Embedding demonstrated that single embedding was significantly more acceptable than double embedding and triple embedding, while the difference between double and triple embedding did not reach significance.

[Figure 5 about here.]

Subexperiment 3. Figure 6 graphs the mean judgments for this subexperiment. An Anova revealed a significant main effect of Antecedent: genitive antecedents were more acceptable than prepositional antecedents. The main effects of Island, Embedding, and Resumption were not significant. No significant interactions were found either.

Again, a Dunnett test was conducted to compare the single and double embedded case violation conditions to an unembedded control; see Table 5 for the results.

[Figure 6 about here.]

[Table 5 about here.]

6.3. Discussion. Subexperiment 1 replicated the effects found in Experiment 2 for Greek. Again, we found that resumptives improve with embedding compared to an unembedded control condition; this was found for doubly embedded resumptives in that-clauses, and for singly embedded resumptives in whether-clauses; see Figure 4. We failed to find a corresponding effect for triple embedding; it seems that additional levels of embedding do not improve resumptives further (in fact there seems to be a tendency in the opposite direction). The present experiment also confirmed the finding that resumptives are never more acceptable than gaps; we found that they were less acceptable than gaps in all conditions.

Declarative controls were included in this experiment to test if embedding per se (even in the absence of extraction) leads to a reduction in acceptability. The results of the first subexperiment confirm this; we found that doubly and triply embedded declarative that- and whether-clauses were less acceptable than unembedded ones. Single embedding seems to have a much weaker effect on acceptability; singly embedded that-clauses were less acceptable than unembedded ones; however, this effect was only significant by subjects and did not extend to whether-clauses.

We also found that extraction incurs a further reduction of acceptability, on top of the one incurred by embedding. The post-hoc analysis of the Island/Resumption interaction showed that declaratives were more acceptable than gaps, which in turn were more acceptable than resumptives.

Subexperiment 2 investigated declarative clauses further and found that there is no interaction between clause type and embedding; that-, whether-, and relative clauses all behave in the same way when it comes to embedding (see Figure 5). Furthermore, we found that single embedding was more acceptable than double embedding, while double embedding was not significantly different from triple embedding.

Subexperiment 3 included stimuli in which the case of the gap or resumptive (accusative) mismatched the case of the wh-phrase (genitive or prepositional). The aim was to test the hypothesis that this case mismatch is less noticeable in more deeply embedded sentences, due to the distance between the gap/resumptive and its antecedent. These results provide only a weak confirmation for this hypothesis: we found a significant improvement in acceptability for

doubly embedded resumptives with a genitive antecedent. However, this effect was small (see Figure 6), and did not extend to prepositional antecedents, as would be expected if this was purely an effect of embedding on case mismatch in resumptives. Note also that the case mismatch always triggers serious unacceptability, even for doubly embedded resumptives, compared to the non-mismatch condition (see Figure 4). This indicates that the case mismatch does not go undetected in the case of double embedding; there is only a small improvement over the unembedded case.

7. Summary of results. The experimental results establish some robust crosslinguistic patterns but also some interesting crosslinguistic differences. In what follows, we summarize the main results of our experiments on island constraints and resumption in object extraction for wh-questions.

7.1. Crosslinguistic similarities.

Pronominals and gaps. The first important crosslinguistic result is that resumptives are at most as acceptable as gaps (but not more acceptable), which means that resumptives cannot 'save' island violations in questions (in the sense of making them fully acceptable) contrary to what has been suggested in the literature.[14] Furthermore, our experimental results showed that a resumptive can reverse the effect of embedding in extraction from non-islands and weak islands: doubly or singly embedded structures with pronominals are more acceptable than unembedded ones. While this effect has been noticed in the theoretical literature (Erteschik-Shir 1992, Tsimpli 1999), the degradation of embedded gaps has gone unnoticed.

Embedding. We consistently found that a structure with single, double, or triple embedding is less acceptable than an unembedded control structure. This finding holds for bare clauses and that-clauses (no islands) as well as for whether-clauses (weak islands). An effect of embedding was detected even in declarative clauses, though it was weaker than for questions. Embedding does not seem to have an effect on extraction from relative clauses (strong islands), which are already highly unacceptable.

Non-islands/weak islands vs. strong islands. Non-islands and weak islands pattern together and contrast with strong islands. Extraction out of non-islands and weak islands gives rise to a mild reduction in acceptability; embedding reduces acceptability, but resumption can compensate for the effect of embedding. By contrast, extraction out of relative clauses always induces strong unacceptability and is immune to embedding and resumption.

The non-island condition is broadly similar to the weak island condition, i.e. both conditions show the same interaction with embedding and resumption. There are some quantitative differences (in English and Greek, extraction out of whether-clauses is less acceptable than extraction out of that-clauses), but the real contrast is with extraction out of strong islands.

7.2. Crosslinguistic variation. The experimental results demonstrate a consistency across all three languages in the overall pattern of interaction between the tested factors. Crosslinguistic variation appears confined to quantitative differences associated with violation of principles that are crosslinguistically identical. We summarize the main crosslinguistic differences:

(i) Resumption in questions is more acceptable in Greek than in German and English. While pronominals in Greek simple questions induce strong unacceptability, they are still significantly more acceptable than strong island violations. By contrast, in German and English, pronominals induce severe unacceptability equal to strong island violations.

(ii) Extraction out of that-clauses is worse in German than in English and Greek; it is almost as unacceptable as extraction out of whether-clauses.

(iii) In all three languages the acceptability of pronominals improves with embedding (for both that- and whether-clauses). However, gaps in English remain significantly better than pronominals in all conditions. This contrasts with German and Greek, where intrusive pronouns can be as acceptable as gaps.

In section 8.3, we attribute these differences to variation in the structural properties of the languages in question.

8. Analysis.

8.1. Locality conditions on movement.

Non-islands/weak islands vs. strong islands. Let us begin with one of the central experimental results, viz. the contrast between extraction from that- and whether-clauses on the one hand, and extraction from relative clauses on the other. This contrast is not surprising; proposals like Rizzi's (1990) treat strong island violations (such as extraction from relative clauses) as Empty Category Principle (ecp) violations, while weak islands violations (such as extraction from whether-clauses) are subject to Relativized Minimality. What is less expected is the parallel between the non-island condition (that-clauses) and the weak island condition (whether-clauses) in object questions.[15]

Let us consider the syntax of the structures in question. We assume that, in all three languages, the element introducing indirect questions is a complementizer occupying C.[16] Further, in all three languages whether-clauses are CP complements, on a par with that-clauses.[17] The contrast between that- and whether-clauses and relative clauses is the familiar contrast between selected and unselected phrases, or complements and adjuncts. We take this as a primitive and assume that the relevant operations that license extraction (e.g. agree) do not have access to adjunct phrases.[18] The contrast then between that- and whether-clauses and strong islands is that the latter, but not the former, violate this condition. We now turn briefly to recent theoretical proposals for weak islands.

Relativized Minimality and weak islands. The basic intuition of Rizzi's (1990) Relativized Minimality underlies most recent formulations of locality conditions in relation to weak islands. According to Relativized Minimality, certain intervenors may interrupt otherwise well-formed chains. Thus, in an example like 30, who in the Spec of intermediate C acts as an intervenor blocking the (government) chain between why and its trace.[19]

(30) a. *Why$_i$ do you know [who$_j$ they fired t_j t_i]?

b. ??What$_i$ do you know [who$_j$ t_j read t_i]?

An immediate problem for this analysis arises with structures like 31, where the specifier position of intermediate C is not occupied by any operator.[20] Rizzi's response was to assume a

covert operator at the specifier of if, which induces the same intervening effect as the overt specifier in 30. However, as has been noted in the literature, Rizzi's solution is stipulative.

(31) a. Why$_i$ do you wonder [if they fired him t$_i$]?

b. Who$_i$ do you wonder [if they fired t$_i$]?

Szabolcsi and Zwarts (1993) and Manzini (1998) develop analyses of weak islands which, among other consequences, eliminate the need for assuming a covert operator in indirect questions introduced by a complementizer.[21] Both analyses abandon Rizzi's (1990) assumption that minimality effects are associated with A-bar specifiers and associate minimality directly with the scope domains of interacting operators, the matrix and intermediate C in examples like 31.

Manzini (1998) presents a syntactic account based on the following definition of minimality (which assumes a minimalist framework):

(32) Minimality

Given an attractor feature F and an attractee feature A$_F$, F attracts A$_F$ only down to the next attractor F′ for A$_F$.

In an example like 33, both matrix and embedded C are specified for the feature Q (Question) and act as attractors of an element bearing such a feature, the wh-phrase. According to 32, the scope of matrix C is closed off by the occurrence of the intermediate C. The reduced acceptability of examples like 33 reflects a violation of Minimality. By contrast, standard declarative C in 34 is not specified for a Q-feature and therefore does not interfere with the scope of the matrix Q-operator.

(33) ??Who do you wonder whether we will fire?

(34) Who do you think we will fire?

Szabolcsi and Zwarts (1993) present a semantic account of weak island effects. In particular, they assume that unlike that-complements of 'volunteer stance' verbs like claim, think or say, complements of a 'non-stance' verb like wonder introduce a Scope Element, the question operator, that interacts with the scope of the wh-element.

The main feature of these two proposals then is that C of indirect questions (whether) is 'richer' than declarative C (that) in that it realizes an operator that takes scope over a domain;

crucially, weak islands arise due to the interaction of two scope domains, that of matrix C and that of the embedded/'intervening' C. The analysis we present in the next section builds on this view.

8.2. Processing cost, cyclicity and resumption.

Embedding and wh-questions. A main finding was that questions extracted from that-clauses are less acceptable than unembedded ones. The observed effect has to be due to processing, since there is no obvious grammatical principle explaining it and the acceptability of such questions is relatively high. Since whether-clauses pattern with that-clauses in a number of ways, the effects in whether-clauses should also be accommodated by whatever processing explanation is invoked for that-clauses. Note also that the effect of embedding is mainly pronounced in questions; though present in declaratives, embedded declaratives were more acceptable than embedded questions.

Related studies in the psycholinguistic literature have established increased processing difficulty in wh-questions extracted from clausal complements, an effect attributed to the involvement of a filler-gap dependency. Thus, Frazier and Clifton (1989) report that measurements of reading times in a self-paced reading task indicated that wh-questions involving embedding were more difficult to process than corresponding yes/no questions, where no filler-gap dependency is involved. The contrast between wh-questions and declaratives found in our data parallels the contrast between wh-questions and yes/no-questions, in that, no filler-gap dependency is involved in declaratives. Contrasts between wh-questions and yes/no questions are also reported by Kluender (1998) and Kluender and Kutas (1993). Interestingly, the relevant contrasts were obtained through a grammaticality judgment task but also in Event Related Potential (erp) experiments using the same materials. These studies therefore establish a parallel between acceptability judgments and processing measures such as erps. Given that processing effects detected by online paradigms manifest themselves also in acceptability judgments, it seems legitimate to assume that the same factors underly the drop in acceptability for wh-questions in our experiments, the increased reading times in Frazier and Clifton's (1989) study, and Kluender's (1998) and

Kluender and Kutas's (1993) acceptability judgment and erp data.

Note that our results do not unambiguously indicate an effect of embedding distinct from string length. However, Frazier and Clifton (1989) have addressed this problem by obtaining reading times for pairs like 35; the two sentences are matched for length but only 35b involves embedding. Longer reading times for 35b indicated an effect of embedding distinct from string length.

(35) a. What did Katie and Tom mail to New York?

b. What did Sue think Tom mailed to New York?

Frazier and Clifton (1989) assume that the processing difficulty of 35b relates to 'carrying a filler across a clause boundary', in technical terms, to the intermediate trace in a C (A-bar) position.[22]

The question of the processing difficulty associated with wh-questions involving embedding is also investigated by Dickey (1996) and Kluender (1998) who present proposals that attempt to provide a unified explanation for center embedding data and the effect of embedding in wh-questions.

Dickey (1996)—building on Kimball (1973)—proposes that the human parser can retain information associated with at most two domains in immediate memory, where a domain is roughly a clause (IP/CP). Thus, when faced with a third clause, the human parser has to shunt one of the previous two from immediate memory; in 36, one of the first two CPs has to be shunted away. In doing so, the parser chooses the most complete CP, where completeness amounts to saturation of subcategorization frames and resolution of fillers with gaps. In 36, the matrix CP is complete because the filler is 'resolved' since it is associated with an intermediate trace at the specifier of C of the embedded clause. Moreover, the matrix filler is specified for ϕ-features (agreement features). This is not the case with the intermediate trace at Spec,CP of the second clause, which is assumed to lack ϕ-features, a fact that makes the second CP less complete than the matrix one; the parser then shunts the matrix CP. By the time then the gap is encountered in the third clause, lack of ϕ-features leads to processing difficulty.

(36) Who do you think Mary claims Bill invited to the party?

Dickey's (1996) hypothesis associates the processing difficulty of three-clause questions with

the number of clauses involved and the lack of ϕ-features for traces. The main weakness of his analysis is that he predicts increased processing cost only for three-clause questions, but not for two. As indicated by our results, a clear effect is detected already in two-clause questions (though the effect is more dramatic in three-clause questions). In two-clause questions, no domain is shunted and so there is no obvious source of processing difficulty. Moreover, it is not obvious how this analysis can be extended to whether-clauses, where the effect is present already with one embedded clause. Finally, shunting should not affect declaratives, since no important information is lost through shunting. However, an effect of embedding was detected in declaratives at the second and third level.

Kluender (1998) focuses on weak islands like 37.

(37) *What did you ask which man was reading?

Following Just and Carpenter (1992) he assumes memory limitations for the human parser and that activation levels of fillers that remain unresolved over longer periods keep decreasing. Embedding, thus, induces a straightforward strain on memory resources. Computational resources may be further exhausted by a number of factors other than embedding, such as the number and type of fillers. For instance, 37 involves two fillers, the first of which has to be carried across a clause boundary. Further, a less referential filler as in 38b may improve the acceptability of examples like 37, as demonstrated by the fact that 38a is more acceptable than 38b (Kluender 1998 uses relative acceptability judgments, where '>' means 'is more acceptable than').[23]

(38) a. What did you wonder who read? >

b. What did you wonder which man read?

In the acceptability study reported by Kluender (1998) he compares the following types of questions and finds 39a to be better than 39b which in turn is better than 39c.

(39) a. What did he think that we should consider? >

b. What did he wonder if we should consider? >

c. What did he wonder who should consider?

His analysis focuses on examples like 39c. He attributes the effect obtained for sentences like 39a to the processing cost associated with carrying a filler across a clause boundary but

offers no explanation for the contrast between 39a and 39b, which parallel the contrast between our that and whether-clauses.

Hawkins (2005) offers an account of processing complexity which, building on Hawkins (1994, 1999), makes reference to the size of the processing domain of a structure and general efficiency principles for the processing of a given domain (e.g. minimize domains, maximize online processing). His theory offers an account of the interaction between resumption and processing complexity in filler gap dependencies (fgd). The main aspect of the account relates to the Proximity Hypothesis: the more relations of combination or dependency are involved between the filler and the predicate in a fgd, the higher the complexity of the structure. For Hawkins (2005) the crucial difference between gaps and pronominals is that the latter only involve a relation of co-indexation between the locally realized pronominal argument and the antecedent. By contrast, gaps, in addition to the co-indexation relation between the gap and the filler, also involve lexical co-occurrence (of the filler and the gap) within the lexical domain of the predicate. The main consequence of this difference is that while co-indexation is marked only once on a chain, information relating to lexical co-occurrence involves every single node intervening between the filler and the gap, thus, increasing the processing load of embedded structures, by increasing the size of the processing domain. Hawkins' main prediction is that in more embedded positions pronouns are preferred over gaps, since the size of the processing domain of gaps keeps increasing with embedding. By contrast, non-embedded gaps are slightly preferred over pronominals. His predictions are borne out for Hebrew relative clauses, on the basis of data Hawkins borrows from Ariel (1990). However, our data do not conform to his predictions. Pronominals in unembedded positions are much worse than gaps, while embedded pronominals are (at best) as acceptable as gaps and never better than gaps, as Hawkins (2005) would predict. It is worth noting that, though Hawkins' analysis cannot be made to work for our data, the analysis we offer in section 8 does incorporate two key aspects of his account: (i) that the relation between the filler and the pronominal is anaphoric in nature and thus immune to the locality restrictions of the human parser and (ii) that the relation between the gap and the filler is a syntactic one, 'registered' through the chain and, thus, subject to locality.

Syntactic Prediction Locality Theory. In the previous section, we have seen that none of the existing theories of processing wh-dependencies offers a satisfactory account of the results of Experiments 1–4. In this section, we present an account of our results based on Gibson's (1998) theory of linguistic complexity, the Syntactic Prediction Locality Theory (splt). splt is a model explaining certain aspects of the language comprehension mechanism in terms of available computational resources. Linguistic complexity is associated with the quantity of computational resources consumed by two distinct components: (i) a memory cost component involved in the storage of parts of the input that may be used in parsing later parts of an input, and (ii) integration cost associated with integrating new input into the structures already built at a given stage in the computation. These costs are defined as in 40 and 42.

(40) Syntactic Prediction Memory Cost

a. The prediction of the matrix predicate, V_0, is associated with no memory cost.

b. For each required syntactic head C_i other than V_0, associate a memory cost M(n) memory units where M(n) is a monotone increasing function and n is the number of new discourse referents that have been processed since C_i was initially predicted. (Gibson 1998:15)

The first clause 40a of the memory cost definition ensures that the matrix predicate with its immediate arguments are not costly. The main part of the definition 40b basically says that once a syntactic prediction is made, the more discourse referents that intervene between the point at which the prediction is made and the point at which it is satisfied, the higher the memory costs associated with this prediction. Thus local resolutions are always less costly than non-local ones. Consider 41:

(41) a. The reporter who the senator attacked admitted the error.

b. The reporter who attacked the senator admitted the error.

In both examples, who introduces a prediction for a gap. Assuming that this prediction is satisfied when the verb (with its associated gap) is encountered, the memory cost associated with satisfying this prediction in 41a is M(1), since one discourse referent, the senator, intervenes between who and attacked. The memory cost of satisfying the same prediction in 41b is M(0) since no discourse referent intervenes between who and attacked. This contrast

explains the finding that subject relative clauses are easier to process that object relative clauses.

(42) Linguistic Integration Cost

> The integration cost associated with integrating a new input head h_2 with a head h_1 that is part of the current structure for the input consists of two parts: (1) a cost dependent on the complexity of the integration (e.g. constructing a new discourse referent); plus (2) a distance-based cost: a monotone increasing function I(n) energy units (EUs) of the number of new discourse referents that have been processed since h_1 was last highly activated. For simplicity, it is assumed that I(n) = n EUs. (Gibson 1998:12–13)

The linguistic integration cost is dependent on two factors. First, the type of element to be integrated matters: new discourse referents (e.g. indefinite NPs) are assumed to involve a higher integration cost than old/established discourse referents, identified by pronominals.[24] Second, like the memory cost, the integration cost is sensitive to the distance between the head being integrated and the head it attaches to, where distance is again calculated in terms of intervening discourse referents. Note that arguments have both an integration and memory cost whereas adjuncts have integration but no memory costs, since they are not predicted.

splt and successive cyclicity. Let us now turn to our data and consider how the splt can provide an account. The main facts to be accounted for are as follows:

(i) Embedded wh-questions are less acceptable than wh-questions without embedding.

(ii) A similar effect of embedding is also present in declarative clauses.

(iii) Wh-questions with embedding are less acceptable than declarative sentences with embedding.

(vi) Wh-questions extracted from that-clauses are more acceptable than wh-questions extracted from whether-clauses.

(v) There is no such effect for declarative clauses.

We adopt here the main assumptions of splt in our account of these facts. However, we argue that locality is sensitive to syntactic heads and hierarchical structure rather than discourse

referents.

We begin with the contrast between singly and doubly embedded wh-questions, exemplified by 35, repeated below as 43. As shown by Frazier and Clifton (1989), 43b is associated with higher reading times and is therefore more difficult to process. Following Frazier and Clifton (1989), we interpret this as an indication that embedding rather than mere string length accounts for the increased processing complexity of 43b. However, this is not a necessary assumption from the point of view of splt. According to splt, what is at issue in both examples is the number of discourse referents intervening between the filler what and the predicate mail with which an appropriate gap is associated. Note that Gibson assumes that verbs can also count as discourse referents, since they introduce discourse events. Given this assumption, the distance between the filler and the gap in 43a is 2, while in 43b it is 3. Since memory and integration costs associated with these examples are calculated on the basis of these distance numbers, it follows that 43a has lower memory and integration costs than 43b. Hence 43a is predicted to involve less processing difficulty than 43b, exactly as found by Frazier and Clifton (1989).

(43) a. What did Katie and Tom mail to New York?

b. What did Sue think Tom mailed to New York?

However, the fact that splt can account for the contrast in 43 by associating distance with discourse referents does not prove that discourse referents are the relevant unit for locality. Since one of the three discourse referents in 43b is a verb (think), it is impossible to know whether what makes the difference between 43a and 43b is the number of discourse referents, irrespective of category (V, N), or the presence of a clausal complement in 43b.

Recent data by Gibson and Warren (2004) shed light on this issue by investigating reading times for sentences such as 44.

(44) a. The manager$_i$ who$_i$ the consultant claimed t$_i$ that the new proposal had pleased t$_i$ will hire five workers tomorrow.

b. The manager$_i$ who$_i$ the consultant's claim about the new proposal had pleased t$_i$ will hire five workers tomorrow.

The two structures in 44 are matched for length, and they are also identical in the number of discourse referents intervening between the filler and the gap. However, they differ in the type of intervening syntactic structure: in 44a, the intervening structure contains a functional head C, while in 44b, a complex NP including a PP modifier intervenes. The experimental results show that 44a is easier to process than 44b, which indicates that the number of discourse referents as such is not sufficient to explain the processing cost for embedded structures. Instead, Gibson and Warren (2004) make the additional assumption that splt distance is calculated relative to intermediate traces that can intervene between the filler and the gap. Such a trace is present in 44a, assuming successive cyclic movement, resulting in a shorter distance between the filler and the gap, but not in 44b, where a complex NP, but no intermediate trace intervenes, resulting in increased processing cost for this structure.

Our results go one step further by showing that the type of intervening functional head also plays a role. In Experiment 1 (see Figures 1(b) and 1(c)), we established a contrast between extraction from that-clauses and from whether-clauses, using examples like 45:

(45) a. Who does Mary claim that we will fire?

b. Who does Mary wonder whether we will fire?

Both examples in 45 involve the same number of discourse referents intervening between the filler and the gap, hence splt predicts that they have identical integration and memory costs, contrary to fact. The acceptability difference cannot be explained by the presence of an intermediate trace either (as suggested by Gibson and Warren 2004), as both structures in 45 involve such a trace. Evidently, the difference between the two types of clauses has to be associated with the complementizer. But note that, according to splt, complexity is associated with discourse referents rather than syntactic heads. It is clear that the definition of 'discourse referent' cannot be extended to include functional elements such as C. Such examples therefore necessitate the assumption that syntactic heads/phrasal nodes are included in the calculation of integration/processing costs. We adopt this assumption, which allows us to explain facts (i)–(iv) as listed above.

We start with point (i), the contrast between embedded and unembedded wh-questions, present in all four experiments. If syntactic heads and not discourse referents are important for

calculating integration and memory costs, then 43b involves at least one extra head (C) compared to 43a.[25] Further, we assume an extra source of processing complexity for such examples, viz. the cost of carrying the filler across a clause boundary. This cost is related to a specific prediction associated with C for a gap later in the clause. In technical terms, this prediction amounts to an intermediate trace associated with C, a standard syntactic assumption capturing the cyclic nature of movement. We should note here that, strictly speaking, this assumption is not necessary for accounting for the contrast in question. Since an independent effect of embedding was detected in declaratives and since questions with fillers are harder to process than yes/no questions, it could well be that the processing cost is a cumulative cost of these two independent effects. That is, there might not be an independent cost of carrying a filler across a clause boundary. However, we favor this assumption because it is in line with the syntactic literature and because, as seen below, such an assumption is necessary for understanding the role of resumption in these structures. In examples like 43b, the integration of intermediate C, that, involves not only the integration of a new syntactic head, but also that of the intermediate trace. Intermediate C then involves a higher integration cost than other lexical heads, since its integration involves the integration of a head and the associated gap/trace (i.e. a head and a specifier).[26] This intermediate trace 'resolves' the matrix filler, and is thus associated with a higher distance-sensitive integration cost. At the same time, it carries the prediction of the gap to the next clause, and is therefore associated with a specific memory cost. The processing difficulty associated with embedding then is due to the higher integration and memory costs incurred by intermediate C, i.e. C licensing an intermediate trace. Thus we predict that examples like 43b involve higher complexity than 43a.

The examples in 44a and 44b present an interesting case since it is 44a, i.e. the example involving intermediate traces that is easier to process. For Gibson and Warren (2004) the intermediate trace in 44a reduces the number of intervening elements, and hence integration and memory cost, which are both sensitive to locality. Thus, despite two integrations of intermediate traces that increase complexity locally at C, the overall processing cost of 44a is reduced in comparison with 44b, where there is no compensation for the intervening complex DP.

Let us now turn to point (ii), the effect of embedding in declaratives (see Figure 5). This effect can be explained by assuming that the integration cost of a CP complement is higher than the integration cost of a DP, i.e. C has a higher complexity-related integration cost (see clause 1 of 42).[27] While the additional cost of integrating a CP is present both in questions and in declaratives, the additional cost of integrating an intermediate trace only occurs in questions; no such trace is present in declaratives, which do not include a wh-element. This explains fact (iii), i.e. that questions with embedding are less acceptable than declaratives with embedding (see Figure 4).

Fact (iv) refers to the contrast between that- and whether-clauses in questions. Building on the theoretical literature reviewed in section 8.1, we assume that the presence of a Q-operator in 45b (Manzini 1998) (or a Scope Element (SE) according to Szabolcsi and Zwarts 1993), is associated with a higher complexity related integration cost, as specified in clause 1 of the Linguistic Integration Cost definition in 42. We should note here that the integration of such a Q-feature is costly exactly because elements of the matrix clause bearing similar features are still not integrated, and thus have not been assigned scope. In other words, the integration cost is due to the interaction of two scope domains. Experimental confirmation for this assumption comes from Anderson's (2004) work on quantifier scope. She presents reading time data that shows that sentences with inverse scope are harder to process than sentences with surface scope.

At the same time our Q-operator analysis also explains fact (v), viz. that the difference between that-clauses and whether-clauses is present for questions, but not for declaratives. Declaratives do not include a scope-bearing wh-element, so no scope interaction between the Q-operator and the wh-element can arise, and no increased processing cost is predicted. This is what we found experimentally: As Figure 5 illustrates, there is no difference in the acceptability of that- and whether-declaratives.

Processing complexity and resumption. Let us now consider the role of resumption, in particular with regard to the following three questions:

(i) Why is resumption unacceptable in simple unembedded questions?

(ii) (a) How do intrusive pronominals interact with the integration costs associated with embedding and indirect questions? (b) Why are intrusive/embedded pronominals better than unembedded ones (modulo strong islands)?

(iii) Why do intrusive pronominals fail to fully compensate for the processing costs associated with embedding/indirect questions and restore the acceptability of such structures?

The unacceptability of pronominals in wh-questions can be accounted for by the fairly standard assumption that these structures are specified for movement (agree/move) which yields a phonologically empty element in situ.[28] This is the case for all three languages. However, for Greek, the existence of clld presents the possibility of a further derivation, viz. one of a clld-ed wh-phrase. We assume that standard cases of clld as in 46a involve movement of a 'clitic doubled' phrase as in 46b (on this see Alexopoulou 1999, Alexopoulou and Kolliakou 2002).

(46) a. to Γiani to sinadisame stin aγora
the.acc Γiani.acc him met.1pl at-the market
'We met Γianis at the market.'

b. stin aγora to sinadisame to Γiani
at-the market him met.1pl the.acc Γiani.acc

Why is then clld of a wh-phrase like 47 unacceptable? We attribute the contrast between 46a and 47 to the type of operator involved in each structure; questions involve a quantificational operator while clld, on a par with English topicalization, involves an anaphoric/referential operator (Lasnik and Stowell 1991, Rizzi 1997, Tsimpli 1999).

(47) *pion ton sinadisate stin aγora
who.acc him met.2pl at-the market
'Who did you meet at the market?'

A crucial point here is that the contrast between 47 and 46a does not relate to the properties of the pronominal. Thus, Greek pronominals can be (A-bar) bound by quantificational operators as the one involved in relative clauses (Lasnik and Stowell 1991) as indicated by the examples

below from relative clauses, where the pronominal is obligatory (see Alexopoulou 2006 for a detailed discussion).[29]

(48) a. kaθe kopela pu tis lei ta mistika tu o Γianis kataliγi
each girl that her.gen say.3sg the secrets his.gen the.nom Γianis.nom reach.3sg
sto siberasma oti ine trelos
to-the conclusion that is insane
'Every girl that Γianis tells his secrets to reaches the conclusion he is mad.'

b. kamia kopela pu tis lei ta mistika tu o Γianis den ton
no girl that her.gen say.3sg the secrets his.gen the.nom Γianis.nom not him
perni sta sovara
take.3sg at-the seriously
'No girl that Γianis tells his secrets to takes him seriously.'

Let us now turn to the second question, viz. how resumption appears to revert processing costs associated with embedding. In the analysis assumed here the main culprit is the cost incurred by the intermediate C and its associated trace. We speculate that the presence of a resumptive makes the parser abandon the syntactic/cyclic resolution of the dependency and revert to an anaphoric dependency. That is, the pronominal searches for its antecedent, the wh-phrase, not through the cyclic syntactic route, but in the previous discourse, as in cases of intrasentential anaphora (for similar ideas see Erteschik-Shir 1992 and Dickey 1996, the discussion of long movement by Cinque 1990 and Hawkins 2005). In other words, the integration cost associated with a pronominal is not sensitive to the locality restrictions that are associated with a syntactic resolution of a filler-gap dependency, as expected for discourse anaphora.

That intrusive resumption is anaphoric in nature has been established since Chao and Sells (1983) and Sells (1984)—see also Cinque (1990). For example, Sells (1984) notes the impossibility of a bound interpretation for the pronominal in 49b. Such examples parallel examples of intrasentential anaphora like 49c.

(49) a. I'd like to meet the linguist that Mary couldn't remember if she had seen ∅/him before.

b. I'd like to meet every linguist that Mary couldn't remember if she had seen ∅/*him

before.

c. We met every linguist. *We met him at the conference.

Similarly, the functional answer in 51[30] is an acceptable response to 50a but not to 50b.[31]

(50) a. Which woman does no Englishman believe will make a good wife?

b. Which woman does no Englishman even wonder whether she will make a good wife?

(51) The one his mother likes best.

Similarly Greek 52b does not allow a bound interpretation of the pronominal and, as a result, cannot accept a functional answer like 52b or a pair-list one like 52c. By contrast, functional and pair-list answers are possible for a gap question as in 53a.

(52) a. Pion ipes oti θa ton eksetasi kaθe γiatros
who.acc said.2sg that will him examine.3sg each doctor.nom
'Who did you say each doctor will examine?'

b. *To diefθidi tu.
the.acc manager his
'His manager.'

c. *Ton Petro i Ikonomu, ti Maria o
the.acc.msc Petros.acc the.nom.fem Ikonomu, the.acc.fem Maria the.nom
Aθanasiu . . .
Aθanasiu . . .
'Petros, Ikonomu (will examine), Maria, Aθanasiu (will examine) . . . '

(53) a. Pion ipes oti θa eksetasi kaθe γiatros
who.acc said.2sg that will examine.3sg each doctor.nom
'Who did you say each doctor will examine?'

b. To dieuθidi tu.
the.acc manager his
'His manager.'

c. Ton Petro i Ikonomu, ti Maria o
the.acc.msc Petros the.nom.fem Ikonomu, the.acc.fem Maria the.nom

Aθanasiu …

Aθanasiu …

'Petros, Ikonomu (will examine), Maria, Aθanasiu (will examine) …'

Intrusive pronominals therefore are not variables; rather, they are linked to their antecedent anaphorically.[32] In this respect, they differ from the pronominals in the Greek relative clauses in 48 where the pronominal is bound by a quantifier.[33]

Let us now turn to the third question, viz. why intrusive pronominals cannot restore embedded/weak island violating questions to full acceptability. The question is all the more important since one of the well-known properties of resumptive chains crosslinguistically is exactly that they do not obey island restrictions (Borer 1984, Sells 1984, McCloskey 1990). Our tentative answer is this: wh-questions in English, Greek, and German are specified for movement (e.g. by a specific feature on matrix C) and therefore are processed as such up to the point of encountering the pronominal. Once encountered, the pronominal gives rise to a different interpretation/processing of the whole sentence, by which the syntactic/cyclic resolution of the dependency is abandoned and an anaphoric resolution is pursued, which, by hypothesis, is less costly.[34] But, crucially, the processing costs (memory and integration) incurred thus far cannot be undone. As a result, intrusive resumption cannot restore the offending structures to full acceptability.[35] This situation is therefore different from cases which are not associated with movement (e.g. if C lacks a relevant feature, as proposed for Irish by McCloskey 2002). In such cases, there is no prediction for a gap and the parser does not enter into a cyclic derivation. No island sensitivity is therefore observed and resumption is the consequence of the absence of movement (see Alexopoulou 2006).

Before we close this section, there are two remaining issues: (i) How is the main finding of Experiment 4, viz. that case mismatches were not tolerated with intrusive pronominals, to be accommodated in the current account? (ii) What is the relation between intrusive resumption and clld? We now turn to these issues, starting with (ii).

Intrusive resumption and clld share an number of important properties. First, on a par with intrusive resumption, clld involves an anaphoric relation between the dislocated antecedent and the pronominal reminiscent of intra-sentential anaphora. For instance, clld-ed

indefinites standardly take wide scope[36] over the universal quantifier kaθe ‘every’ in 54a, numerals and intensional predicates like ψaxno ‘look for’ in 54b,c (Philippaki-Warburton 1985, Anagnostopoulou 1994, Alexopoulou and Kolliakou 2002). Thus, examples 54 are parallel to example 52a.

(54) a. Ena arθro tu Chomsky to δiavase kaθe fititis
an/one article the.gen Chomsky it read.3sg each student.nom
‘There is an article of Chomsky every student read.’ (Wide scope for the indefinite)

b. Ena δoro γia to Γiani (*to) ψaxno eδo ki ena mina ke δe
a/one present for the.acc Γianis.acc it look-for.1sg here and one month and not
boro na vro tipota pu na m’aresi
can.1sg to find.1sg nothing that to me-like.3sg
‘A present for Γianis I’ve been looking for for a month, but I cannot find anything I like.’

c. Ena pukamiso γa to Γiani (pu tu to ixa pari stin Arγedini) to
a shirt for the Γianis.acc (that him.gen it had.1sg got in Argentina) it
ψaxno edo ke meres ke de boro na to vro me tipota
look-for.1sg here and days and not can.1sg it to find.1sg with nothing
‘A shirt for the Γiani (that I had got for him in Argentina) I’ve been looking for for some days but I cannot find it.’

Second, as with intrusive resumption, case mismatches are ungrammatical in clld, a property that distinguishes clld from Left Dislocation where case mismatches are possible (for diagnostics distinguishing clld from left dislocation, see Cinque 1990, Tsimpli 1995, Alexopoulou et al. 2003, Alexopoulou 1999).

Third, though the issue has been the matter of some debate, most researchers agree that clld obeys islands in Greek[37] (Anagnostopoulou 1994, Iatridou 1995, Tsimpli 1995, Alexopoulou and Kolliakou 2002).

These shared properties between intrusive resumption and clld give rise to the hypothesis that intrusive resumption in Greek involves nothing else but unacceptable clld, where the antecedent is just of the wrong type, i.e. quantificational rather than referential.

Our answer is the following. First, the contrast between unembedded questions with pronominals and embedded ones indicates that intrusive resumption is independent from clld; the status of the antecedent (wh-phrase/quantificational as opposed to referential) does not change with embedding; no improvement therefore should be expected in the embedded condition. This improvement can only be understood if intrusive resumption is admitted as distinct from clld. But of course, the two are more similar than different. How are their similarities to be understood? We view clld as a grammaticized version of anaphoric dependencies relevant for intra-sentential anaphora. This assumption explains the interpretational properties of clld; the 'grammaticization' of this dependency relates to the requirement for case agreement between the dislocated phrase and the pronominal and its sensitivity to islands.[38] It is possible that there is a continuum of such anaphoric relations, depending on how loosely/closely the antecedent is integrated with the syntactic structure of the following sentence, ranging from intrasentential anaphora (no syntax relevant), to Left Dislocation (no cyclicity, case mismatches acceptable, no island sensitivity, root phenomenon) to clld (cyclicity, no case mismatches, island sensitivity, non-root phenomenon). In this continuum, clld and intrusive resumption are distinct, the latter possibly closer to intra-sentential anaphora.

Let us finally consider the fact that case mismatches were not tolerated with intrusive pronominals in Experiment 4. At first sight, this finding appears inconsistent with the view that intrusive resumption involves a process resembling intra-sentential anaphora, since case is irrelevant in intra-sentential anaphora. However, note that while similar to intrasentential anaphora, cases of intrusive resumption are different in that the antecedent has not been properly syntactically integrated with previous discourse in a sentence. In more syntactic terms, its case features have not been licensed/checked. Ultimately, when the dependency is resolved (whichever way, syntactically or anaphorically), the wh-phrase is integrated with some predicate in the question, not with some other sentence in previous discourse. This then is what accounts for the case requirement. Note crucially that this does not—at least not necessarily—undermine the hypothesis that some anaphoric process is involved, since intrasentential processes may preserve case as indicated by the unacceptability of 55c in the

context of 55a.

(55) a. Pios kerδise to laxio?
who.nom won the lottery
'Who won the lottery?'

b. o Γianis
the.nom Γianis

c. *to Γiani
the.acc Γiani

8.3. Crosslinguistic variation. As mentioned earlier, the overall pattern in the results we obtained is crosslinguistically identical, indicating that the principles underlying these effects are crosslinguistically constant. However, some crosslinguistic variation arises, prima facie due to quantitative differences in the seriousness of the violations under investigation. In this section, we reduce such quantitative variation to the structural properties of the languages in question.

The first striking difference relates to the acceptability of the pronominal in simple questions like 56. While in all three languages such examples lead to strong unacceptability, the Greek example in 56b is more acceptable than its English and German counterparts since it is significantly better than the strong island violation (see Figure 2(c)). By contrast, English and German questions with (unembedded) pronominals as 56a and 56c are as unacceptable as strong island violations (see Figures 1(d) and 3(c)). As has been established in the previous section, Greek differs from English and German in that unembedded resumptive sentences are not necessarily analyzed as movement chains with an illicit pronominal in place of a gap/trace. Rather, they can be analyzed as cases of clld; under such a derivation, the problem is not the presence of a pronominal in a movement chain, but rather the incompatibility between the wh-phrase as a non-referential/quantificational antecedent and the pronominal in an anaphoric chain. This violation lies at the syntax-semantics interface and, as such, can be assumed to lead to a milder reduction in acceptability (Sorace and Keller 2005).

(56) a. Who will we fire him?

b. Pion θa ton apolisume
who.acc will him fire.1pl

c. Wen entlassen wir ihn?
who.acc fire.1pl we him

The second main contrast is the stronger effect of embedding in German dass-clauses, which parallels the weak island violation in whether-clauses (see Figure 3(a)). We relate this difference to the fact that, unlike Greek and English, the complement CP is extraposed to the right in German (Keller 1995). In particular, we assume that the integration of an extraposed CP is associated with higher cost.

Let us finally consider why gaps remain significantly better than pronominals in English, while intrusive pronominals are as acceptable as gaps in Greek and German (compare Figure 1(c) and Figures 2(b) and 3(b)). We interpret this as an epiphenomenon of the absence of clld and CP extraposition in English. The unavailability of clld makes pronominals in English less tolerated than in Greek. In German on the other hand, CP extraposition makes embedded gaps worse than any of the three languages. Absence of clld and CP extraposition leaves English with relatively acceptable embedded gaps and unacceptable pronominals; as a result, gaps are always more acceptable than pronominals.

8.4. Questions vs. relative clauses. As explained in section 2.4, we have here focused on object questions and have ignored relative clauses as well as subject questions. In the following, we briefly review relevant experimental results.

McDaniel and Cowart (1999) present an experimental study that tests the acceptability of gaps and pronouns in subject relative clauses and investigates the interaction between resumption, C-trace effect and subjacency (weak islands). In particular, they compare the acceptability of pronominals and gaps in the following structures.[39] Both 57 and 58 involve a weak island violation. However, 57, on top of the weak island violation, it involves a that-trace violation.

(57) a. *That's the girl that I wonder when met you.

b. That's the girl that I wonder when she met you.

(58) a. ?That's the girl that I wonder when you met.

b. [judgment?] That's the girl that I wonder when you met her.

Their results differ from ours with respect to the acceptability of resumption. First, in the object position in 58, the pronominal is as acceptable as the gap. Second, in the subject position in 57, the pronominal improves the C-trace violation and is better than the gap. Hence 58b is as acceptable as 58a, while 57b is better than 57a. Furthermore, the weak island violation in 58a is more acceptable than the combination of the C-trace and weak island violation in 57a. But, more interestingly 58, involving a weak island but no trace violation, is as acceptable as 57b which involves both a weak island and C-trace violation.

These results contrast with our results from English questions, where gaps were better than resumptives in object extractions from whether-clauses. We believe that this contrast relates to a more general contrast between questions and relative clauses with respect to the acceptability of resumption.[40] As already mentioned in section 2.4, a number of corpus studies indicate that resumption is productive, not only as a 'last resort' device, but also as a general strategy of relative clause formation (Prince 1990, 1997, Cann et al. 2005). If resumption is independently available in relative clauses, then it is not surprising that pronominals and gaps are equally acceptable in the object position, while gap structures like 57a are worse than all other structures; only 57a involves a grammatical violation. The remaining three structures just involve a weak island violation, which, as shown in our results on questions, is a less serious violation. The contrast between our results and the results of McDaniel and Cowart (1999) therefore relates to the unavailability of the resumptive strategy in questions as opposed to relative clauses.[41]

9. Conclusion. This article has presented a systematic experimental investigation of the interaction between locality and resumption in wh-questions. Resumptives have traditionally been claimed to save island violations, i.e. to improve the acceptability of extraction out of weak and strong islands, though such claims have predominantly involved resumption in relative clauses. We tested this claim using a series of Magnitude Estimation experiments that investigated object extraction in various syntactic configurations: complement clauses with and without that (non-islands), complement clauses with whether (weak islands), and relative

clauses (strong islands). We also tested multiple levels of embedding (single, double, and triple) and included two control conditions: unembedded questions and declarative clauses. In order to be able to differentiate language specific effects from crosslinguistically constant ones, we conducted identical experiments in three languages: English, Greek, and German. The Greek data is particularly relevant as Greek (in contrast to English and German) allows resumptive pronouns in certain syntactic configurations, most importantly clld.

Experiments 1–4 established a robust pattern of results that holds across all three languages. The most striking finding was the absence of a saving effect of resumption in questions: we found that a resumptive pronoun is at most as acceptable as a gap in the same construction, but never more acceptable. This means that resumptives do not remedy island violations in questions, and hence cannot be viewed as a 'last resort' strategy. This highly surprising result is at variance with claims in the theoretical literature. However, we also found that resumption in questions can compensate for embedding in certain cases: resumptives with one or two levels of embedding were more acceptable than unembedded resumptives. This effect was limited to extraction in the non-island and weak island conditions, and was strongest in German.

Another surprising finding was a general effect of embedding, even in structures that are considered fully grammatical, such as wh-questions extracted out of bare clauses or that-clauses (which are not islands). We even found that embedding reduces acceptability in declarative sentences, also considered fully grammatical. Based on experimental results in the literature, we argued that this is a genuine effect of embedding, and not one of sentence length (see section 8.2).

The third major finding was that the non-islands and weak islands pattern together: in both cases extraction led to a mild reduction in acceptability and resumption in questions compensated for embedding. This contrasts with the strong island (relative clause) condition, in which extraction led to strong unacceptability, and resumption and embedding had no effect.

The experimental results demonstrate a remarkable consistency across the three languages in the overall pattern of interactions between the factors we studied. Crosslinguistic

variation is confined to quantitative differences associated with universal grammatical principles: We found that (a) resumption in questions is more acceptable in Greek than in German and English, compared to strong island violations; (b) extraction out of that-clauses is worse in German than in English and Greek (in German it is almost as unacceptable as extraction out of whether-clauses); (c) gaps in English are significantly better than pronominals in all conditions; in Greek and German, resumptives can reach the same acceptability as gaps.

Based on our experimental results, we presented an analysis that explains the gradient nature of the acceptability judgments in terms of the interaction of different cognitive modules. We argued that strong islands involve grammatical violations, which, on a par with violations of core syntactic principles like case, give rise to strong unacceptability. Such violations cannot be remedied by resumption (at least in the wh-questions investigated here). By contrast, weak island violations, on a par with extractions from that-clauses, give rise to mild unacceptability. We argued that this is caused by an interaction of the syntactic properties of these structures with the demands they impose on the human sentence processor. In particular, we built on Gibson's (1998) theory of human sentence processing to develop an account of the processing complexity of A-bar dependencies. In this account, locality conditions associated with the Relativized Minimality effects for weak islands can be viewed as a grammaticization of the resource limitations of the human parser, i.e. a response of the grammar to processing pressures. Resumption can compensate for the processing difficulty associated with these structures, as pronominals can seek antecedents anaphorically (as in discourse). This means they do not necessarily have to rely on local, cyclic movement (which is obligatory for the resolution of traces), and therefore are less sensitive to the locality effects associated with movement.

On more general level, we have demonstrated that there is much to be gained from the crosslinguistic experimental study of gradient acceptability. By eliciting English, Greek, and German data in parallel, we were able to show that resumption has essentially the same effect in all three languages. Crucially, this was true for cases where the theoretical literature suggests crosslinguistic variation. As an example, take the claim that Greek (unlike English

and German) allows resumptives in embedded clauses. Our experimental data showed that the empirical basis for this claim is simply that embedded resumptives are more acceptable than unembedded ones, a finding that holds for all three languages. Only the absolute acceptability of resumptives (independent of embedding) differs between Greek and English or German. A result of this type cannot be obtained based on informal, intuitive acceptability judgments; it requires experimental data such as the Magnitude Estimation data presented here.

According to the current analysis, the effect of resumption is understood in terms of lesser costs induced by resolving a long distance dependency anaphorically rather than syntactically. This approach can be extended to accommodate known interactions between d-linking, weak islands and resumption, which await a proper experimental investigation. Moreover, interesting questions arise with respect to whether there is a range of anaphoric relations or factors that can contribute to a structure being less or more anaphoric (e.g. referential/quantificational antecedent, resumption/gap) which accordingly may induce weaker or stronger effects with islands. A systematic investigation and comparison between types of structures (questions, topicalization and clld) and types of antecedents (d-linked/referential) can illuminate such questions.

Finally, none of the languages tested here allows resumption of the type attested in Semitic and Celtic languages, where pronominals are insensitive to islands. A comparison with such languages is important in order to understand relations between intrusive and true resumption. Data reported and discussed by Erteschik-Shir (1992) and Dickey (1996) indicate that, at least, intrusive resumption in Hebrew is not that different from intrusive resumption in English, Greek and German. The question is whether true resumption will prove to be distinct from intrusive resumption in such languages or is better understood as the end point of a continuum.[42] The question is all the more important given that true resumption is attested predominantly in relative clauses which also appear to favor (or at least tolerate) resumption in Greek and English.

References.

Alexiadou, Artemis, and Anagnostopoulou, Elena. 2000. Asymmetries in the distribution of clticis: the case of Greek restrictive relatives. Clitic phenomena in European languages, ed. by Frits Beukema and Marcel den Dikken, 47–70. Amsterdam: John Benjamins.

Alexopoulou, Theodora. 1999. The Syntax of Discourse Functions in Greek: a non-configurational approach. Ph.D. thesis, University of Edinburgh.

Alexopoulou, Theodora. 2006. Resumption in relative clauses. Natural Language and Linguistic Theory To appear.

Alexopoulou, Theodora, Doron, Edit, and Heycock, Caroline. 2003. Broad subjects and clitic left dislocation,. Peripheries: syntactic edges and their effects, ed. by David Adger, Cecile de Cat, and George Tsoulas, 329–358. Dordrecht: Kluwer.

Alexopoulou, Theodora, and Kolliakou, Dimitra. 2002. On linkhood and clitic left dislocation. Journal of Linguistics 38.193–245.

Anagnostopoulou, Elena. 1994. Clitic Dependencies in Modern Greek. Ph.D. thesis, University of Salzburg.

Anderson, Catherine. 2004. The Structure and Real-time Comprehension of Quantifier Scope Ambiguity. Ph.D. thesis, Northwestern University.

Aoun, Joseph, and Choueiri, Lina. 1997. Resumption and last resort. Ms., University of Southern California.

Ariel, Mira. 1990. Accessing Noun-Phrase Antecedents. London and New York: Routledge.

Bard, Ellen Gurman, Robertson, Dan, and Sorace, Antonella. 1996. Magnitude estimation of linguistic acceptability. Language 72.32–68.

Boeckx, Cedric. 2003. Islands and Chains: Resumption as Derivational Residue. Amsterdam: John Benjamins.

Borer, Hagit. 1984. Restrictive Relatives in Modern Hebrew. Natural Language and Linguistic Theory 2.219–260.

Cann, Ronnie, Kaplan, Tami, and Kempson, Ruth. 2005. Data at the grammar-pragmatics interface: The case of resumptive pronouns in English. Lingua 115.1551–1578.

Chao, Wynn, and Sells, Peter. 1983. On the interpretation of resumptive pronouns.

Proceedings of the 13th Conference of the North-Eastern Linguistic Society, ed. by Peter Sells and C. Jones, 47–61. Amherst: GLSA.

Chomsky, Noam. 1975. The Logical Structure of Linguistic Theory. New York: Plenum Press.

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

Chung, Sandra, and McCloskey, James. 1983. On the interpretation of certain island effects in GPSG. Linguistic Inquiry 14.704–713.

Cinque, Guglielmo. 1990. Types of A′-Dependencies. Cambridge, MA: MIT Press.

Corley, Martin, and Scheepers, Christoph. 2002. Syntactic priming in English sentence production: Categorical and latency evidence from an internet-based study. Psychonomic Bulletin and Review 9.126–131.

Cowart, Wayne. 1997. Experimental Syntax: Applying Objective Methods to Sentence Judgments. Thousand Oaks, CA: Sage Publications.

Dickey, Michael Walsh. 1996. Constraints on the sentence processor and the distribution of resumptive pronouns. Linguistics in the Laboratory, ed. by Michael Walsh Dickey and Susanne Tunstall, no. 19 in Occasional Papers in Linguistics, 157–192. Amherst, MA: University of Massachusetts.

Erteschik-Shir, Nomi. 1992. Resumptive pronouns in islands. Island Constraints: Theory, Acquisition and Processing, ed. by Helen Goodluck and Michael Rochemont, 89–108. Dordrecht: Kluwer.

Fanselow, Gisbert, and Mahajan, Anoop. 2000. Towards a minimalist theory of wh-expletives, wh-copying, and successive cyclicity. Wh-Scope Marking, ed. by Uli Lutz, Gereon Müller, and Arnim von Stechow, 195–230. Amsterdam: John Benjamins.

Featherston, Sam. 2005. Magnitude estimation and what it can do for your syntax: Some wh-constraints in German. Lingua 115.1525–1550.

Frazier, Lyn, and Clifton, Charles. 1989. Successive cyclicity in the grammar and the parser. Language and Cognitive Processes 4.93–126.

Giannakidou, Anastasia. 1997. The Landscape of Polarity Items. Ph.D. thesis, University of Groningen.

Gibson, Edward. 1998. Linguistic complexity: locality of syntactic dependencies. Cognition

68.1–76.

Gibson, Edward, and Warren, Tessa. 2004. Reading time evidence for intermediate linguistic structure in long-distance dependencies. Syntax 7.55–78.

Haegeman, Liliane. 1994. Introduction to Government and Binding Theory. Oxford: Blackwell, 2nd edn.

Hawkins, John. 2005. Efficiency and Complexity in Grammars. Oxford: Oxford University Press.

Hawkins, John A. 1994. A Performance Theory of Order and Constituency. Cambridge: Cambridge University Press.

Hawkins, John A. 1999. Processing complexity and filler-gap dependencies across grammars. Language 75.244–285.

Iatridou, Sabine. 1995. Clitics and island effects. Penn Working Papers in Linguistics 2.11–31.

Joseph, Brian D. 1978. Morphology and Universals in Syntactic Change: Evidence from Medieval and Modern Greek. Ph.D. thesis, Harvard University. Published as Joseph (1990).

Joseph, Brian D. 1980a. Linguistic universals and syntactic change. Language 56.345–370.

Joseph, Brian D. 1980b. Recovery of information in relative clauses: evidence from Greek and Hebrew. Journal of Linguistics 16.237–244.

Joseph, Brian D. 1983. Relativization in Modern Greek. Lingua 60.1–24.

Joseph, Brian D. 1990. Morphology and Universals in Syntactic Change: Evidence from Medieval and Modern Greek. New York: Garland.

Just, M., and Carpenter, P. 1992. A capacity theory of comprehension: Individual differences in working memory. Psychological Review 99.122–149.

Kayne, Richard. 1991. Romance clitics, verb movement, and pro. Linguistic Inquiry 22.647–686.

Keller, Frank. 1995. Towards an account of extraposition in HPSG. Proceedings of the 7th Conference of the European Chapter of the Association for Computational Linguistics, 301–306. Dublin: Association for Computational Linguistics. Student Session.

Keller, Frank, and Alexopoulou, Theodora. 2001. Phonology competes with syntax:

Experimental evidence for the interaction of word order and accent placement in the realization of information structure. Cognition 79.301–372.

Keller, Frank, and Asudeh, Ash. 2001. Constraints on linguistic coreference: Structural vs. pragmatic factors. Proceedings of the 23rd Annual Conference of the Cognitive Science Society, ed. by Johanna D. Moore and Keith Stenning, 483–488. Mahwah, NJ: Lawrence Erlbaum Associates.

Keller, Frank, and Sorace, Antonella. 2003. Gradient auxiliary selection and impersonal passivization in German: An experimental investigation. Journal of Linguistics 39.57–108.

Kimball, John. 1973. Seven principles of surface structure parsing in natural language. Cognition 2.15–47.

Kluender, Robert. 1998. On the distinction between strong and weak islands: A processing perspective. The limits of syntax, ed. by Peter W. Culicover and Louise McNally, 241–269. New York: Academic Press.

Kluender, Robert, and Kutas, Marta. 1993. Subjacency as a processing phenomenon. Language and Cognitive Processes 8.573–633.

Kroch, Anthony. 1981. On the role of resumptive pronouns in amnestying island constraint violations. Proceedings of the Meeting of the Chicago Linguistics Society, 125–135. Chicago: Chicago Linguistics Society.

Lasnik, Howard, and Stowell, Tim. 1991. Weakest Crossover. Linguistic Inquiry 22.687–720.

Lodge, M. 1981. Magnitude scaling: quantitative measurement of opinions. Beverley Hills, CA: Sage Publications.

Longobardi, Giuseppe. 1986. L'estrazione dalle "isole" e lo scope dei sintagmi quantificati. Parallela 2: Aspetti della sintassi dell'italiano contemporaneo. Tübingen: Gunter Narr.

Manzini, M. Rita. 1998. A minimalist theory of weak islands. The limits of syntax, ed. by Peter W. Culicover and Louise McNally, 185–120. New York: Academic Press.

McCloskey, James. 1990. Resumptive pronouns, A′-binding and levels of representation in Irish. Syntax of the modern Celtic languages, ed. by Randall Hendrick, 199–248. New York: Academic Press.

McCloskey, James. 2002. Resumption, successive cyclicity, and the locality of operations. Derivation and Explanation in the minimalist program, ed. by Samuel D. Epstein and Daniel T. Seely, 184–226. Oxford: Blackwell.

McDaniel, Dana, and Cowart, Wayne. 1999. Experimental evidence of a minimalist account of English resumptive pronouns. Cognition 70.B15–B24.

Merchant, Jason. 2004. Resumptivity and non-movement. Studies in Greek Linguistics, vol. 24, 471–481. Thessaloniki: University of Thessaloniki.

Meyer, Roland. 2003. Superiority effects in Russian, Polish and Czech: Comparative evidence from studies on linguistic acceptability. Proceedings of the 12th Conference on Formal Approaches to Slavic Linguistics. Ottawa: University of Ottawa.

Parodi, Teresa, and Tsimpli, Ianthi-Maria. 2005. "Real" and apparent optionality in L2 grammars: finiteness and pronouns in Null Operator structures. Second Language Research 21/3.250–85.

Philippaki-Warburton, Irene. 1985. Word order in Modern Greek. Transactions of the Philological Society 83.113–143.

Philippaki-Warburton, Irene, and Veloudis, Ioannis. 1985. The subjunctive in complement clauses. Proceedings of the 5th Annual Meeting of the Deptartment of Linguistics, 149–167. Thessaloniki: University of Thessaloniki.

Prince, Ellen F. 1990. Syntax and discourse: A look at resumptive pronouns. Proceedings of the Sixteenth Annual Meeting of the Berkeley Linguistic Society, ed. by Kira Hall, 482–97. Berkeley: Berkeley Linguistic Society.

Prince, Ellen F. 1997. On kind-sentences, resumptive pronouns, and relative cluases. Towards a Social Science of Language: Papers in Honor of William Labov, ed. by Gregory R. Guy, John G. Baugh, Deborah Schiffrin, and Crawford Feagin, 223–35. Cambridge: Cambridge University Press.

Rizzi, Luigi. 1990. Relativized Minimality. Cambridge, MA: MIT Press.

Rizzi, Luigi. 1997. The Fine Structure of the Left Periphery. Elements of Grammar: Handbook of Generative Syntax, ed. by Liliane Haegeman, 281–337. Dordrecht: Kluwer.

Ross, John R. 1967. Constraints on Variables in Syntax. Ph.D. thesis, MIT. Published as Ross

(1986).

Ross, John R. 1986. Infinite Syntax! Norwood, NJ: Ablex.

Roussou, Anna. 2002. C, t, and the subject: That-t phenomena revisited. Lingua 112.13–52.

Schütze, Carson. 1996. The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology. Chicago: University Chicago Press.

Sells, Peter. 1984. Syntax and Semantics of Resumptive Pronouns. Ph.D. thesis, University of Massachusetts at Amherst.

Sharvit, Yael. 1999. Resumptive pronouns in relative clauses. Natural Language and Linguistic Theory 17.587–612.

Shlonsky, Ur. 1992. Resumptive Pronouns as a Last Resort. Linguistic Inquiry 23.443–448.

Sorace, Antonella, and Keller, Frank. 2005. Gradience in linguistic data. Lingua 115.1497–1524.

Stavrou, Melita. 1984. Clitic pronominals in direct object pou-relatives (in Greek). Studies in Greek Linguistics, vol. 5, 121–136. Thessaloniki: University of Thessaloniki.

Stevens, Stanley Smith. 1975. Psychophysics: Introduction to its Perceptual, Neural, and Social Prospects. New York: John Wiley.

Suñer, Margarita. 1998. Resumptive restrictive relatives: a crosslinguistic perspective. Language 74.335–364.

Szabolcsi, Anna, and Zwarts, Frans. 1993. Weak islands and an algebraic semantics for scope taking. Natural Language Semantics 1.235–284.

Tsimpli, Ianthi Maria. 1995. Focusing in Modern Greek. Discourse Configurational Languages, ed. by Katalin É. Kiss, 176–206. Oxford: Oxford University Press.

Tsimpli, Ianthi Maria. 1999. Null operators, clitics and identification: A comparison between Greek and English. Studies in Greek Syntax, ed. by Artemis Alexiadou, Geoffrey Horrocks, and Melita Stavrou, 241–262. Dordrecht: Kluwer.

Tsiplakou, Stavroula. 1998. Focus in Greek: its structure and interpretation. Ph.D. thesis, University of London, School of Oriental and Africal Studies.

A. Details of statistical analyses.

A.1. Experiment 1. An Anova yielded significant main effects of Embedding ($F_1(1,54) = 14.330$, $p < .0005$; $F_2(1,8) = 15.704$, $p = .004$), Island ($F_1(3,162) = 55.772$, $p < .0005$; $F_2(3,24) = 45.811$, $p < .0005$), and Resumption ($F_1(1,54) = 67.807$, $p < .0005$; $F_2(1,8) = 81.794$, $p < .0005$). The interactions Island/Resumption ($F_1(3,162) = 30.101$, $p < .0005$; $F_2(3,24) = 35.977$, $p < .0005$) and Embedding/Resumption ($F_1(1,54) = 15.381$, $p < .0005$; $F_2(1,8) = 5.976$, $p = .040$) were also significant. The other interactions were only significant by subjects: Island/Embedding ($F_1(3,162) = 4.217$, $p = .007$; $F_2(3,24) = .877$, $p = .467$), and Island/Embedding/Resumption ($F_1(3,162) = 6.878$, $p < .0005$; $F_2(3,24) = 2.395$, $p = .093$).

A further series of tests was carried out to compare the single and double embedding conditions to the control condition (no embedding). The appropriate statistic is Dunnett's test for comparing multiple conditions to a control. We first report the results of comparing the gapped stimuli to the gapped control condition. For both non-island conditions, there was no significant difference between control and the single embedding condition, while the double embedding was significantly less acceptable than the control ($t_{d_1}(55,9) = 5.734$, $p < .01$; $t_{d_2}(8,9) = 5.110$, $p < .01$ and $t_{d_1}(55,9) = 6.886$, $p < .01$; $t_{d_2}(8,9) = 5.536$, $p < .01$). In the weak island condition, we found that both the single and the double embedding condition were less acceptable than the control ($t_{d_1}(55,9) = 5.710$, $p < .01$; $t_{d_2}(8,9) = 3.891$, $p < .05$ and $t_{d_1}(55,9) = 8.8184$, $p < .01$; $t_{d_2}(8,9) = 6.350$, $p < .01$). Also in the relative clause condition, singly and doubly embedded stimuli were less acceptable than the control ($t_{d_1}(55,9) = 10.825$, $p < .01$; $t_{d_2}(8,9) = 6.209$, $p < .01$ and $t_{d_1}(55,9) = 11.799$, $p < .01$; $t_{d_2}(8,9) = 9.382$, $p < .01$).

In a separate test, we compared the resumptive stimuli to the resumptive control condition. In the non-island condition, there was no significant difference between the singly embedded resumptive and the control, while the doubly embedded resumptive was significantly more acceptable than the control, by subjects only ($t_{d_1}(55,9) = 2.752$, $p < .05$; $t_{d_2}(8,9) = 1.377$, $p > .05$). In the that-clause condition, the singly embedded resumptive was

more acceptable than the control (by subjects only, $t_{d_1}(55,9) = 3.034$, $p < .05$; $t_{d_2}(8,9) = 1.165$, $p > .05$), while the doubly embedded resumptive was not different from the control. No significant differences with the control were found for the weak and strong island conditions.

A.2. Experiment 2. An Anova yielded significant main effects of Embedding ($F_1(1,58) = 26.509$, $p < .0005$; $F_2(1,6) = 19.933$, $p = .004$), Island ($F_1(2,116) = 82.828$, $p < .0005$; $F_2(2,12) = 137.211$, $p < .0005$), and Resumption ($F_1(1,58) = 22.875$, $p < .0005$; $F_2(1,6) = 12.006$, $p = .013$). The interaction of Island and Resumption was also significant ($F_1(2,116) = 10.005$, $p < .0005$; $F_2(2,12) = 4.016$, $p = .046$). All the other interactions were only significant by subjects: Island/Embedding ($F_1(1,116) = 15.072$, $p < .0005$; $F_2(2,12) = 3.409$, $p = .067$), Embedding/Resumption ($F_1(1,58) = 7.705$, $p = .007$; $F_2(1,6) = 4.494$, $p = .078$), and Island/Embedding/Resumption ($F_1(2,116) = 5.888$, $p = .004$; $F_2(2,12) = 3.872$, $p = .050$).

Dunnett's test was used to compare the embedded conditions to the control conditions. We first report the results for the gapped stimuli. For the non-island condition, both the single and the double embedding condition were less acceptable than the control ($t_{d_1}(59,7) = 5.641$, $p < .01$; $t_{d_2}(6,7) = 4.472$, $p < .05$ and $t_{d_1}(59,7) = 8.695$, $p < .01$; $t_{d_2}(6,7) = 9.562$, $p < .01$). Also in the weak island condition, both levels of embedding were significantly less acceptable than the control ($t_{d_1}(59,7) = 8.619$, $p < .01$; $t_{d_2}(6,7) = 14.428$, $p < .01$ and $t_{d_1}(59,7) = 7.532$, $p < .01$; $t_{d_2}(6,7) = 6.005$, $p < .01$). The same picture emerged in the strong island conditions, again both levels of embedding were worse than the control ($t_{d_1}(59,7) = 12.323$, $p < .01$; $t_{d_2}(6,7) = 12.017$, $p < .01$ and $t_{d_1}(59,7) = 12.470$, $p < .01$; $t_{d_2}(6,7) = 17.066$, $p < .01$).

A separate test compared the resumptive stimuli to the resumptive controls. In the non-island condition, neither the single nor the double embedding were significantly different from the control. In the weak island condition, the single embedding condition was significantly more acceptable than the control, by subjects only ($t_{d_1}(59,7) = 3.034$, $p < .05$; $t_{d_2}(6,7) = 2.930$, $p > .05$). There was no difference between the double embedding condition and the control. In the strong island condition, both the single and the double embedding

condition were significantly less acceptable than the control ($t_{d_1}(59,7) = 4.955$, $p < .01$; $t_{d_2}(6,7) = 7.058$, $p < .01$ and $t_{d_1}(59,7) = 4.284$, $p < .01$; $t_{d_2}(6,7) = 6.107$, $p < .01$).

A.3. Experiment 3. An Anova yielded a significant main effect of Island ($F_1(2,72) = 34.415$, $p < .0005$; $F_2(2,12) = 51.787$, $p < .0005$), but the main effects of Embedding and Resumption were not significant. The interactions Island/Resumption ($F_1(2,72) = 5.774$, $p = .005$; $F_2(2,12) = 4.614$, $p = .033$) and Island/Embedding were also significant ($F_1(2,72) = 6.766$, $p = .002$; $F_2(2,12) = 3.917$, $p = .049$). All other interactions failed to reach significance.

We also compared the conditions with single and double embedding to the control (no embedding) using Dunnett's test. We first report the results of comparing the gapped stimuli to the gapped control condition. For the that-clause condition, the control was significantly more acceptable than the single embedding condition ($t_{d_1}(36,7) = 6.527$, $p < .01$; $t_{d_2}(6,7) = 3.494$, $p < .05$) and the double embedding condition ($t_{d_1}(36,7) = 6.900$, $p < .01$; $t_{d_2}(6,7) = 6.033$, $p < .01$). The same pattern was obtained in the whether-clause condition, where the control was more acceptable than both the single and the double embedding condition ($t_{d_1}(36,7) = 8.848$, $p < .01$; $t_{d_2}(6,7) = 5.695$, $p < .01$ and $t_{d_1}(36,7) = 8.493$, $p < .01$; $t_{d_2}(6,7) = 5.636$, $p < .01$). Also, in the relative clause condition, the control was more acceptable than both levels of embedding ($t_{d_1}(36,7) = 11.283$, $p < .01$; $t_{d_2}(6,7) = 13.923$, $p < .01$ and $t_{d_1}(36,7) = 10.247, p < .01$; $t_{d_2}(6,7) = 7.907$, $p < .01$).

A separate test was used to compare the resumptive stimuli to the resumptive control condition. In the that-clause condition, we found that the single embedding condition was more acceptable than the control, by subjects only ($t_{d_1}(36,7) = 3.037$, $p < .05$; $t_{d_2}(6,7) = 2.854$, $p > .05$). Also the double embedding condition was more acceptable than the control ($t_{d_1}(36,7) = 3.839$, $p < .01$; $t_{d_2}(6,7) = 3.701$, $p < .05$). In the whether-clause condition, we found that both the single embedding and the double embedding conditions were significantly more acceptable than the control, by subjects only ($t_{d_1}(36,7) = 4.623, p < .01$; $t_{d_2}(6,7) = 3.314$, $p > .05$ and $t_{d_1}(36,7) = 2.715$, $p < .05$; $t_{d_2}(6,7) = 2.293$, $p > .05$). In the relative clause condition, there was no significant difference between the control and the single and double embedding conditions.

A.4. Experiment 4.

Subexperiment 1. An Anova yielded significant main effects of Island (by subjects only, $F_1(1,32) = 8.727$, $p = .006$; $F_2(1,8) = 5.289$, $p = .050$), Embedding ($F_1(2,64) = 19.180$, $p < .0005$; $F_2(2,16) = 17.902$, $p < .0005$), and Resumption ($F_1(2,64) = 24.132$, $p < .0005$; $F_2(2,16) = 70.939$, $p < .0005$). The interactions Island/Resumption ($F_1(2,64) = 5.703$, $p = .005$; $F_2(2,16) = 4.676$, $p = .025$) and Embedding/Resumption ($F_1(4,128) = 5.588$, $p < .0005$; $F_2(4,32) = 2.900$, $p = .037$) were also significant. All other interactions failed to reach significance.

A Dunnett test was conducted to further investigate the effect of embedding. For the gap condition, singly embedded that-clauses were less acceptable than the unembedded control (by subjects only, ($t_{d_1}(32,7) = 3.147$, $p < .05$; $t_{d_2}(8,7) = 1.995$, $p > .05$); also double and triple embedding was less acceptable than the control ($t_{d_1}(32,7) = 5.356$, $p < .01$; $t_{d_2}(8,7) = 6.660$, $p < .01$ and $t_{d_1}(32,7) = 5.300$, $p < .05$; $t_{d_2}(8,7) = 5.781$, $p < .01$). The same pattern was found for gaps in whether-clauses: single, double, and triple embedding was less acceptable than the control ($t_{d_1}(32,7) = 4.833$, $p < .01$; $t_{d_2}(8,7) = 5.240$, $p < .01$ and $t_{d_1}(32,7) = 6.474$, $p < .01$; $t_{d_2}(8,7) = 5.687$, $p < .01$ and $t_{d_1}(32,7) = 7.383$, $p < .01$; $t_{d_2}(8,7) = 5.879$, $p < .01$)

A separate Dunnett test for the resumptive condition showed that for that-clauses, doubly embedded resumptives were more acceptable than the unembedded control (by items only, $t_{d_1}(32,7) = 1.940$, $p > .05$; $t_{d_2}(8,7) = 3.497$, $p < .05$); for whether-clauses, singly embedded resumptives were more acceptable than the control (by subjects only, $t_{d_1}(32,7) = 2.221$, $p < .05$; $t_{d_2}(8,7) = 1.187$, $p > .05$).

A Dunnett test for the declarative condition showed that both for that-clauses and for whether-clauses, double and triple embedding was less acceptable than the control ($t_{d_1}(32,7) = 4.654$, $p < .01$; $t_{d_2}(8,7) = 8.438$, $p < .01$ and $t_{d_1}(32,7) = 5.198$, $p < .01$; $t_{d_2}(8,7) = 4.809$, $p < .01$ and $t_{d_1}(32,7) = 4.837$, $p < .01$; $t_{d_2}(8,7) = 4.389$, $p < .01$ and $t_{d_1}(32,7) = 5.597$, $p < .01$; $t_{d_2}(8,7) = 5.110$, $p < .01$). For that-clauses, the single embedding was also less acceptable than the control (by subjects only, $t_{d_1}(32,7) = 2.787$, $p < .05$;

$t_{d_2}(8,7) = 0.720$, $p > .05$).

Subexperiment 2. For this subexperiment, we conducted an Anova that included the declarative stimuli from the first subexperiment, so as to be able to determine if different types of declaratives show a differential effect of embedding. The Anova therefore crossed the factors Embedding (single, double, triple) and Type (that-clause, whether-clause, relative clauses). A significant main effect of Embedding was found ($F_1(2,64) = 22.216$, $p < .0005$; $F_2(2,16) = 29.78$, $p < .0005$), but there was no main effect of Type, and no interaction between the two factors.

Subexperiment 3. An Anova revealed a significant main effect of Antecedent ($F_1(1,32) = 4.563$, $p = .040$; $F_2(1,8) = 6.362$, $p = .036$): genitive antecedents were more acceptable than prepositional antecedents. The main effects of Island, Embedding, and Resumption were not significant. No significant interaction were found either.

A Dunnett test was conducted to compare the singly and doubly embedded case violation conditions to an unembedded control. The only significant difference that were found was for doubly embedded clauses with genitive antecedents. Here, resumptives were more acceptable than the control both for that-clauses ($t_{d_1}(32,5) = 3.272$, $p < .01$; $t_{d_2}(8,5) = 3.335$, $p < .05$) and for whether-clauses (by subjects only, $t_{d_1}(32,5) = 3.494$, $p < .01$; $t_{d_2}(8,5) = 1.024$, $p > .05$).

B. Experimental materials.

B.1. Practice and filler items. In each experiments, six practice items were used. All of them were wh-question of varying complexity. Half of the items were grammatical, the other half contained grammatical violations of varying seriousness (e.g. agreement, subcategorization, word order violations). None of the items included resumptive pronouns. The filler items were designed in the same way as the practice items.

B.2. Experiment 1. Modulus:

(59) With who do you want to know whether Bill will go out?

Templates for questions:

(60) a. Who will we $1 (him)?
b. Who does $2 claim (that) we will $1 (him)?
c. Who does $3 think that $2 claims (that) we will $1 (him)?

(61) a. Who does $2 wonder whether we will $1 (him)?
b. Who does $3 think that $2 wonders whether we will $1 (him)?

(62) a. Who does $2 meet the people that will $1 (him)?
b. Who does $3 think that $2 meets the people that will $1 (him)?

Lexicalizations:

(63) a. $1: fire, phone, evict, hire, punish, support, elect, invite, arrest
b. $2: Mary, Ann, Elizabeth, Ruth, Lucy, Laura, Rachel, Susan, Emily
c. $3: Jane, Margaret, Sarah, Jean, Helen, Alice, Diana, Clare, Caroline

B.3. Experiment 2. Modulus:

(64) Me pion iθeles na maθis an vɣike i Maria?
with who.acc wanted.2sg to know.2sg if went-out.3sg the.nom Maria

Templates for questions:

(65) a. Pion θa (ton) $1?
who.acc will (him) $1

b. Pion isxirizete $2 oti θa (ton) $1?
who.acc claim.3sg $2 that will (him) $1

c. Pion nomizi $3 oti isxirizete $2 oti θa (ton) $1?
who.acc think.3sg $3 that claim.3sg $2 that will (him) $1?

(66) a. Pion anarotiete $2 an θa (ton) $1?
who.acc wonder.3sg $2 whether will (him) $1

b. Pion nomizi $3 oti anarotiete $2 an θa (ton) $2?
who.acc think.3sg $3 that wonder.3sg $2 whether will (him) $2

(67) a. Pion sinadai $2 tus tipus pu θa (ton) $1?
who.acc meet.3sg $2 the.acc guys.acc that will (him) $1

b. Pion nomizi $3 oti sinadai $2 tus tipus pu θa (ton) $1?
who.acc think.3sg $3 that meet.3sg $2 the.acc guys.acc that will (him) $1

Lexicalizations:

(68) a. $1: apolisume, kalesume, δioksume, proslavume, vravefsume, ipostiriksume,
fire.1pl, invite.1pl, send-away.1pl, hire.1pl, reward.1pl, support.1pl,
ψifisume
vote-for.1pl

b. $2: i Maria, i Natasa, i Anna, i
the.nom.fem Maria, the.nom.fem Natasa, the.nom.fem Anna, the.nom.fem
Aliki, i Dina, i Mirela, i Sofia
Aliki, the.nom.fem Dina, the.nom.fem Mirela, the.nom.fem Sofia

c. $3: o Petros, o Nikos, o Γianis,
the.nom.msc Petros.nom, the.nom.msc Nikos.nom, the.nom.msc Γianis.nom,
o Kostas, o Panos, o Takis,
the.nom.msc Kostas.nom, the.nom.msc Panos.nom, the.nom.msc Takis.nom,

o Γιοργος

the.nom.msc Γιοργος.nom

B.4. Experiment 3. Modulus:

(69) Mit wem willst du wissen ob Peter ausgeht?

with whom want you know if Peter go-out

Templates for questions:

(70) a. Wen $1 wir (ihn)?

who we him

b. Wen behauptet $2, dass wir (ihn) $1?

who claims that we him

c. Wen denkt $3, dass $2 behauptet, dass wir (ihn) $1?

who thinks that claims that we him

(71) a. Wen überlegt $2, ob wir (ihn) $1?

who ponders if we him

b. Wen denkt $3, dass $2 überlegt, ob wir (ihn) $1?

who thinks that ponders if we him

(72) a. Wen trifft $2 die Leute, die (ihn) $1?

who meets the people who him

b. Wen denkt $3, dass $2 die Leute trifft, die (ihn) $1?

who thinks that the people meets who him

Lexicalizations:

(73) a. $1: entlassen, informieren, vertreiben, bestrafen, unterstützen, wählen, verhaften

fire inform chase away punish support elect arrest

b. $2: Petra, Maria, Sabine, Jutta, Heike, Christine, Andrea

c. $3: Barbara, Monika, Ursula, Brigitte, Renate, Helga, Elisabeth

B.5. Experiment 4. Modulus:

(74) Me poion iθeles na maθis an vγike i Maria?
with whom wanted.2sg to know.2sg if entered.3sg the Maria

Templates for questions:

(75) a. Pion θa (ton) $1?
who.acc will (him) $1

b. Pion θeori $2 oti θa ton $1?
who.acc speculate.3sg $2 that will (him) $1

c. Pion nomizi $3 oti θeori $2 oti θa ton $1?
who.acc think.3sg $3 that speculate.3sg $2 that will (him) $1?

d. Pion nomizi $4 oti pistevi $3 oti θeori $3 oti θa ton $1?
who.acc think.3sg $4 that believe $3 that speculate.3sg $3 that will (him) $1?

(76) a. Pianu θa ton $1?
who.gen will him $1

b. Pianu θeori $2 oti θa ton $1?
who.gen speculate.3sg $2 that will (him) $1

c. Pianu nomizi $3 oti θeori $2 oti θa ton $1?
who.acc think.3sg $3 that speculate.3sg $2 that will him $1

(77) a. Se pion θa ton $1?
to who.acc will him $1

b. Se pion θeori $2 oti θa ton $1?
to who.acc speculate.3sg $2 that will him $1

c. Se pion nomizi $3 oti θeori $4 oti θa ton $1?
to who.acc think.3sg $3 that speculate $4 that will him $1

(78) a. Pion anarotiete $2 an θa ton $1?
who.acc ponder.3sg $2 whether will him $1

b. Pion nomizi $3 oti anarotiete $2 an θa ton $1?
who.acc think.3sg $3 that ponder.3sg $2 whether will him $1

c. Pion nomizi $4 oti pistevi $3 oti anarotiete $2 an θa ton $1?
who.acc think.3sg $4 that believe $3 that ponder.3sg $2 whether will him $1

(79) a. Pianu anarotiete $2 an θa ton $1?
who.gen ponder.3sg $2 whether will him $1

b. Pianu nomizi $3 oti anarotiete $2 an θa ton $1?
who.gen think.3sg $3 that ponder.3sg $2 whether will him $1

(80) a. Se pion anarotiete $2 an θa ton $1?
to who.acc ponder.3sg $2 whether will him $1

b. Se pion nomizi $3 oti anarotiete $2 an θa ton $1?
to who.acc think.3sg $3 that ponder.3sg $2 whether will him $1

Templates for declaratives:

(81) a. θa $1 ton $5
will $1 him $5

b. $2 θeori oti θa $1 $5
$2 speculate.3sg that will $1 $5

c. $3 nomizi oti $2 θeori oti θa $1 $5
$3 think.3sg that $2 speculate.3sg that will $1 $5

d. $4 nomizi oti $3 pistevi oti $2 θeori oti θa $1 $5
$4 think.3sg that $3 believe.3sg that $2 speculate.3sg that will $1 $5

(82) a. $2 anarotiete an θa $1 $5
$2 ponder.3sg whether will $1 $5

b. $3 nomizi oti $2 anarotiete an θa $1 $5
$3 think that $2 ponder.3sg whether will $1 $5

c. $4 nomizi oti $3 pistevi oti $2 anarotiete an θa $1 $5
$4 think.3sg that $3 believe.3sg that $2 ponder.3sg whether will $1 $5

(83) a. $2 sinadai tus tipus
$2 meet.3sg the.acc guys.acc

b. $2 sinadai tus tipus pu θa $1 $5
$2 meet.3sg the.acc guys.acc that will $1 $5

c. $3 nomizi oti $2 sinadai tus tipus pu θa $1 $5
$3 think.3sg that $2 meet.3sg the.acc guys.acc that will $1 $5

d. $4 nomizi oti $3 pistevi oti $2 sinadai tus tipus pu θa $1 $5
$4 think.3sg that $3 believe that $2 meet.3sg the.acc guys.acc that will $1 $5

Lexicalizations:

(84) a. $1: apolisume, kalesume, δioxume, proslavume, vravefsume, ipostiriksume,
fire.1pl, invite.1pl, send-away.1pl, hire.1pl, reward.1pl, support.1pl,
ψifisume, δialexume, kratisume
vote-for.1pl, choose.1pl, keep.1pl

b. $2: i Maria, i Natasa, i Anna, i
the.nom.fem Maria, the.nom.fem Natasa, the.nom.fem Anna, the.nom.fem
Aliki, i Dina, i Mirela, i Sofia, i
Aliki, the.nom.fem Dina, the.nom.fem Mirela, the.nom.fem Sofia, the.nom.fem
Marina, i Aleka
Marina, the.nom.fem Aleka

c. $3: o Petros, o Nikos, o Γianis,
the.nom.msc Petros.nom, the.nom.msc Nikos.nom, the.nom.msc Γianis-non,
o Kostas, o Panos, o Takis,
the.nom.msc Kostas.nom, the.nom.msc Panos.nom, the.nom.msc Takis.nom,
o Γiorgos, Vasilis, o Manos
the.nom.msc Γiorgos.nom, the.nom.msc Vasilis.nom, the.nom.msc Manos.nom

d. $4: o Nikos, o Γianis, o Kostas, o Panos, o Takis, o Γiorgos, Vasilis, o Manos, o Petros

e. $5: to Γiani, ton Kosta, ton Pano,
the.acc.msc Γiani.acc, the.acc.msc Kostas.acc, the.acc.msc Panos.acc,
ton Taki, to Γiorgo, ton Vasili,
the.acc.msc Takis.acc, the.acc.msc Γiorgos.acc, the.acc.msc Vasilis.acc,
to Mano, ton Petro, to Niko
the.acc.msc Manos.acc, the.acc.msc Petros.acc, the.acc.msc Nikos.acc

Notes. *We wish to thank Richard Breheny, Ronnie Cann, Liliane Haegeman, Caroline Heycock, John Hawkins, Ruth Kempson, Dimitra Kolliakou, Bob Ladd, Jim McCloskey, Lutz Marten, Philip Miller, Mits Ota, Teresa Parodi, Ian Roberts, Ivan Sag, Peter Sells, Antonella Sorace, Patrick Sturt, Christoph Scheepers and Ianthi Tsimpli for comments and discussions regarding this work; also the audiences of seminars at the University of Edinburgh, the Research Center for English and Applied Linguistics, Cambridge, the School of Oriental and African Studies, London, the Department of Linguistics, Stanford, the Department of Linguistics, University California Santa Cruz, SILEX, Université de Lille III Charles de Gaulle and the LAGB meeting in Sheffield, 2003. We are also indebted to the editors of Language, Brian Joseph, Jim McCloskey, and Shari Speer and two anonymous reviewers for insightful suggestions and criticisms. The first author thanks Ian Roberts for organizing a visiting fellowship at Cambridge during which this research was completed and written up and gratefully acknowledges support by the Leverhulme Trust Grant R35028.

[1]The diacritic ?* is used by Haegeman (1994) for the example in 85 (ibid, p. 492), which, for current purposes we take as equivalent to 1b. (If anything, 1b ought to be even worse than 85, since d-linking is often assumed to improve weak island violations. We ignore this issue here.)

(85)?*Which man do you wonder when John will meet?

[2]The apparent discrepancy in judgments does not relate to the type of complementizer used, whether vs. if. Chung and McCloskey (1983) make the same assumptions for whether- and if-clauses. Moreover, as discussed in section 8.2, Kluender (1998) reports experimental results on extraction from if-clauses that are parallel to those reported here for extractions from whether-clauses.

[3]D-linking has been argued to improve the acceptability of resumptives in questions (Sells 1984, Iatridou 1995, Anagnostopoulou 1994, Giannakidou 1997).

[4]Shlonsky (1992) analyzes true resumption in Semitic relative clauses as an instance of ‘last resort’ resumption.

[5]This number excludes relative clauses involving resumption in island environments.

[6]One might note that d-linking in questions may improve the acceptability of

pronominals; however, while it is relatively straightforward to abstract away from d-linking in questions and initially study non-d-linked questions, it is not obvious how to neutralize the definite/indefinite contrast in relative clauses. Moreover, d-linking has been argued to interact with resumption in relative clauses as well (Stavrou 1984, Sharvit 1999).

[7]Such facts have led to the hypothesis that, in addition to the standard syntax of A-bar dependencies, there is a syntactic relation (agree) between pu and the nominal/case features of the relativized phrase (Alexopoulou 2006; see also Merchant 2004).

[8]As pointed out to us by the editor, complement clauses introduced by na are, according to many analyses, instances of complementizer-less complement clauses in Greek (Philippaki-Warburton and Veloudis 1985). While it would be interesting to compare oti-complement clauses with na-complement clauses with regard to the phenomena addressed here, such an investigation goes beyond the scope of this article.

[9]Under Sells's (1984) definition, true resumption involves an operator bound pronominal; it is not obvious that clld is a case of true resumption under this definition, since, under standard assumptions, no (quantificational) operator is involved in clld (Cinque 1990, Lasnik and Stowell 1991, Rizzi 1997, Tsimpli 1999).

[10]We owe this point to an anonymous reviewer.

[11]Greek has a further resumptive structure involved in Null Operator Structures (Joseph 1978, 1980a, Tsimpli 1999). The examples in 86, from Tsimpli (1999), involve a pronominal, while the corresponding English structures involve a gap (Lasnik and Stowell 1991).

(86) a. i Maria ine omorfi op na *(tin) kitas
the.nom Maria is pretty op to cl.acc look-at.2sg
'Maria is pretty to look at.'

b. i filosofia ine vareti op na *(ti) δiavazis
the philosophy is boring op to cl.acc read.2sg
'Philosophy is boring to read.'

Parodi and Tsimpli (2005) analyze such cases as instances of intrusive resumption due to an opaque/weak-island like domain created by a rich T in the Greek na-clause which blocks the identification of the empty category. Understanding the relation between this type of

resumption and resumption in questions is beyond the scope of this article.

[12]The pronominal is argued to be as acceptable as the gap in examples like 87 from Tsimpli (1999). In our results, singly embedded pronominals remained significantly worse than corresponding gaps. It is possible that this contrast is not necessarily one between informal judgments and experimentally collected data. It could be that the embedding predicate in 87, suspect (ipoptevome) is more 'opaque' than the one used in the experimental stimuli, claim (isxirizome), thus, making 87 behaving more like a weak island, where singly embedded pronominals and gaps are equally acceptable.

(87) Pion ipoptefθike i Maria oti θa ton kalesume?
who.acc suspected.3sg the.nom Maria that will him invite.1pl
'Who did Maria suspect we will invite?'

[13]Following the suggestion of an anonymous reviewer we tested the possibility that speakers of Southern dialects of German would give lower scores to our data due to preference for the so-called wh-expletive construction 88, widely available in Southern dialects (Fanselow and Mahajan 2000).

(88) Was behauptet Petra, wen wir entlassen?
what claims Petra who we sack
'Who does Petra claim that we will sack?'

We thus divided the experimental data into two classes, based on the language region subjects reported in the demographic questionnaire that preceded the experiment. Speakers from Austria, Switzerland, Baden-Württemberg, and Bavaria were classified as speakers of Southern German, while all others were classified as speakers of Northern German. There were ten Southern speakers and 22 Northern speakers; the data of five subjects had to be excluded from the analysis as they stated Germany as their language region. We re-ran the Anova, now with Dialect as an additional, between-subjects factor. As in the previous analysis, we found a significant main effect of Island and significant interactions of Island and Resumption and Island and Embedding. There was no main effect of Dialect and all interactions involving Dialect were also not significant.

Another Anova was conducted for the unembedded control condition, which revealed a

main effect of Resumption, but no main effect of Dialect. However a significant interaction of Resumption and Dialect was found ($F_1(1,30) = 5.172$, $p = .03$; not enough data for F_2). While Northern and Southern speakers gave comparable judgments to unembedded sentences without resumptives (mean $= .5109$ and mean $= .4796$, respectively), they differed in their assessment of unembedded resumptives (mean $= -.2399$ and mean $= -.6671$, respectively).

[14]Admittedly, such claims predominately involve examples from relative clauses; however, on the basis of the theoretical literature, no contrast is expected between relative clauses and questions.

[15]But note that authors like Chung and McCloskey (1983) regard extraction out of whether-clauses as grammatical.

[16]This analysis is an oversimplification of the English facts. Kayne (1991) provides evidence that English whether is a wh-phrase occupying Spec,CP rather than C. For simplicity, we continue to assume that English whether is a head. The analysis presented in the following sections is compatible with either hypothesis. In a system of multiple specifiers, cyclic movement is not blocked by an element at Spec,CP. Note also that extraction out of whether-clause is of the same acceptability as extraction from if-clauses in English, indicating that whether the complementizer is analyzed as a specifier or a head is not crucial for the islandhood of these structures.

[17]We thus depart from Cinque's (1990) assumption that in English whether-clauses are not directly selected.

[18]Recent minimalist analyses lack an (at least explicit) theory for strong islands (but see Boeckx 2003 for a recent discussion of adjunct islands). In earlier versions of the theory, the distinction was captured on the basis of antecedent government (see also Roussou 2002).

[19]It is standardly assumed that examples like 30b are more acceptable than 30a. The contrast was originally thought to be one between arguments and adjuncts (Cinque 1990). However, more recently, it has been acknowledged that d-linked adjuncts like 89 (from Manzini 1998) are more acceptable than 30a.

(89) For which of these reasons do you know [who they fired t_{who} $t_{for\text{-}which\text{-}of\text{-}these\text{-}reasons}$]?

[20]As observed by Manzini (1998), the problem is inherited by recent reformulations of Relativized Minimality, such as Chomsky's (1995) Minimal Link Condition.

[21]Both analyses naturally account for similar problems raised by extraction out of factive complements, which are also viewed as weak islands. We restrict our discussion here to indirect questions.

[22]Frazier and Clifton (1989) did not control for the number of discourse referents intervening between the filler and the gap, an issue addressed by Gibson and Warren (2004) (see section 42).

[23]Kluender (1998) further assumes that identical factors underlie the unacceptability of strong island violations like 90. However, this approach fails to explain why extractions out of relative clauses are generally worse than extractions out of indirect questions:

(90) *What did you meet the man who was reading?

[24]In this respect, Gibson (1998) disagrees with Kluender (1998) who makes the opposite assumption, i.e. that indefinite NPs are easier to process than definite ones.

[25]An issue arises with respect to the analysis of the coordinate in 43a, which could be taken to involve the same number of phrasal heads as 43b. However, since we further assume that the complexity of intermediate C is increased due to its association with an intermediate trace, 43a will always turn out to be less complex than 43b.

[26]An analysis could be envisaged in which the intermediate trace counts as a distinct discourse referent associated with a specific prediction, therefore incurring higher memory and integration costs. However, it would still be difficult to see how such an analysis could account for the contrast between that and whether-clauses, which, presumably, involve the same number and type of discourse referents.

[27]It is not obvious how splt can account for the effect of embedding in declaratives. In fact, Gibson makes some further assumptions which make the effect of embedding in declaratives rather surprising. He assumes a clause-bound closure, i.e. clauses where all dependents are saturated are shunted from immediate memory. He assumes that when this happens, then the complement clause counts as matrix, in which case its predicate with its arguments are not associated with any memory costs (see clause (a) of 40). The tendency for a drop in

acceptability in particular in three-clause English declaratives may be problematic for this assumption.

[28]There is a question as to which criteria suggest an explanation on the basis of grammar as opposed to processing for a given resumptive structure. The contrast between (unembedded) clld and (unembedded) resumptive questions in Greek indicates that the unacceptability of the latter ought to follow from some grammatical principle, since there is no obvious difference in the processing costs involved in the two structures. Similarly, no obvious grammatical factor can explain why the acceptability of embedded resumptive questions improves.

[29]The question arises why pronominals appear to be acceptable in relative clauses in examples like 48 but not in questions since, under standard assumptions, both involve quantificational operators. Indeed, as with questions, pronominals are dispreferred in restrictive relative clauses in Greek in argument positions of pu-relative clauses and generally in restrictive relative clauses introduced by a relative pronoun. It is in oblique pu-relative clauses that the pronominal is obligatory due to failure of identification of the ϕ-features of the relativized phrase (see Alexopoulou 2006 on this; for resumption in Greek relative clauses see also Joseph 1980b, 1983, Stavrou 1984, Tsimpli 1999, Alexiadou and Anagnostopoulou 2000, Tsiplakou 1998, Merchant 2004).

[30]Answers as in 51 involve functions from Englishmen to their mothers to women as opposed to answers involving an individual (e.g. Mary).

[31]There is an issue with the acceptability status of 50b, which seems to be more acceptable than our corresponding examples. The contrast might relate to the fact that the quantifier and present tense favor a generic interpretation of 50b, which crosslinguistically is known to improve the acceptability of pronominals (Tsimpli 1999, Alexopoulou 2006).

[32]There is evidence that, at least in the case of weak islands, an anaphoric dependency is involved, irrespective of whether a pronominal is present. For example, Cinque (1990)—reporting judgments originally due to Longobardi (1986)— notes that, while 91a is ambiguous between a wide scope reading for the wh-phrase and a wide scope reading for the universal quantifier, the latter reading is absent from 91b, where the wh-phrase is resumed by a pronominal. As in 91b, a wide scope reading for the universal is unavailable also in 92, even

though a gap is involved.

(91) a. Quanti pazienti ritieni che debba visitare t ogni medico?
how-many patients do-you-think that should visit t each doctor
'How many patients do you think each doctor should visit?'

b. ?Quanti pazienti ritieni che li debba visitare t ogni medico
how-many patients do-you-think that them should visit t each doctor
'How many patients do you think each doctor should visit?'

(92) Quanti pazienti te ne sei andato prima che ogni medico potesse visitare ?
how-many patients you cl be go before that every doctor could visit
'How many patients did you go away before each doctor could visit?'

[33]Note that pronominals in oblique positions of relative clauses introduced by complementizers are, by and large, considered 'last resort' elements, exactly like intrusive pronominals in our data (Shlonsky 1992, Suñer 1998, Alexopoulou 2006). But the two types of 'last resort' pronominals appear to behave differently with regard to whether they are anaphorically linked to their antecedents.

[34]We here assume that the local integration costs of gaps and pronominals are identical, since we have no direct evidence to the contrary. If they are not, and, for instance pronominals have a higher integration cost, Greek clld should always be less acceptable than a gap structure like wh-question or focus movement, which, according to informal judgments, is not the case.

[35]There is also the possibility, as suggested to us by an anonymous reviewer, that 'reverting' to a different strategy is itself costly. We leave this possibility open for further research.

[36]At least in episodic sentences.

[37]But note that according to the intuitions of the first author, island-violating questions are worse than island-violating clld.

[38]In the current analysis, sensitivity to islands is the hallmark of cyclicity, and, therefore movement. In this respect we deviate from the dominant view in the Greek literature which is that clld does not involve movement, even though it is sensitive to islands. Note that the source of island sensitivity in questions involving intrusive resumption is different from clld, and due

to the fact that resumption cannot undo costs associated with a cyclic structure. Admittedly, until a thorough comparison of island effects in questions and clld is undertaken, the distinction between clld and intrusive resumption remains not at all straightforward.

[39]The asterisks and question marks indicate the informal acceptability judgments they give for these examples before conducting the experiment. By [judgment?] in 58b the authors indicate their uncertainty about the appropriate judgment/diacritic.

[40]Note that the materials used by McDaniel and Cowart (1999) included whether-clauses as well as indirect questions introduced by why, when, how and where. It is therefore unlikely that the contrast relates to the nature of the element introducing the indirect question (head vs. specifier).

[41]A proper corpus study should of course verify this. However, as far as we are aware, there have been no claims regarding the underreported use of resumption in English questions. Moreover, the contrast between questions and relative clauses with respect to the availability of resumption is not specific to English. Crosslinguistically, resumption is more widely available and often obligatory in relative clauses, and its distribution is generally more restricted in questions, often subject to pragmatic conditions such as d-linking.

[42]Comparisons between questions and relative clauses in Hebrew or Lebanese Arabic would illuminate the issue, since resumption in questions is generally dispreferred in the absence of discourse linking in both languages; by contrast, resumption in relative clauses is acceptable irrespective of d-linking in Hebrew and obligatory in Lebanese Arabic (Borer 1984, Shlonsky 1992, Aoun and Choueiri 1997).

	Gap		Resumptive	
	Single	Double	Single	Double
Clause type	Embedding	Embedding	Embedding	Embedding
Bare clause		*		(*)
That-clause		*	(*)	
Whether-clause	*	*		
Relative clause	*	*		

*: significant by subjects and items; (*): significant by subjects only

Table 1: Result of Dunnett tests comparing embedded clauses with the unembedded control in Experiment 1

	Gap		Resumptive	
	Single	Double	Single	Double
Clause type	Embedding	Embedding	Embedding	Embedding
That-clause	*	*		
Whether-clause	*	*	(*)	
Relative clause	*	*	*	*

*: significant by subjects and items; (*): significant by subjects only

Table 2: Result of Dunnett tests comparing embedded clauses with the unembedded control in Experiment 2

	Gap		Resumptive	
	Single	Double	Single	Double
Clause type	Embedding	Embedding	Embedding	Embedding
That-clause	*	*	(*)	*
Whether-clause	*	*	(*)	(*)
Relative clause	*	*		

*: significant by subjects and items; (*): significant by subjects only

Table 3: Result of Dunnett tests comparing embedded clauses with the unembedded control in Experiment 3

	Gap			Resumptive			Declarative		
Clause type	Single Embd.	Double Embd.	Triple Embd.	Single Embd.	Double Embd.	Triple Embd.	Single Embd.	Double Embd.	Triple Embd.
That-clause	(*)	*	*		[*]		(*)	*	*
Whether-clause	*	*	*	(*)	*			*	*

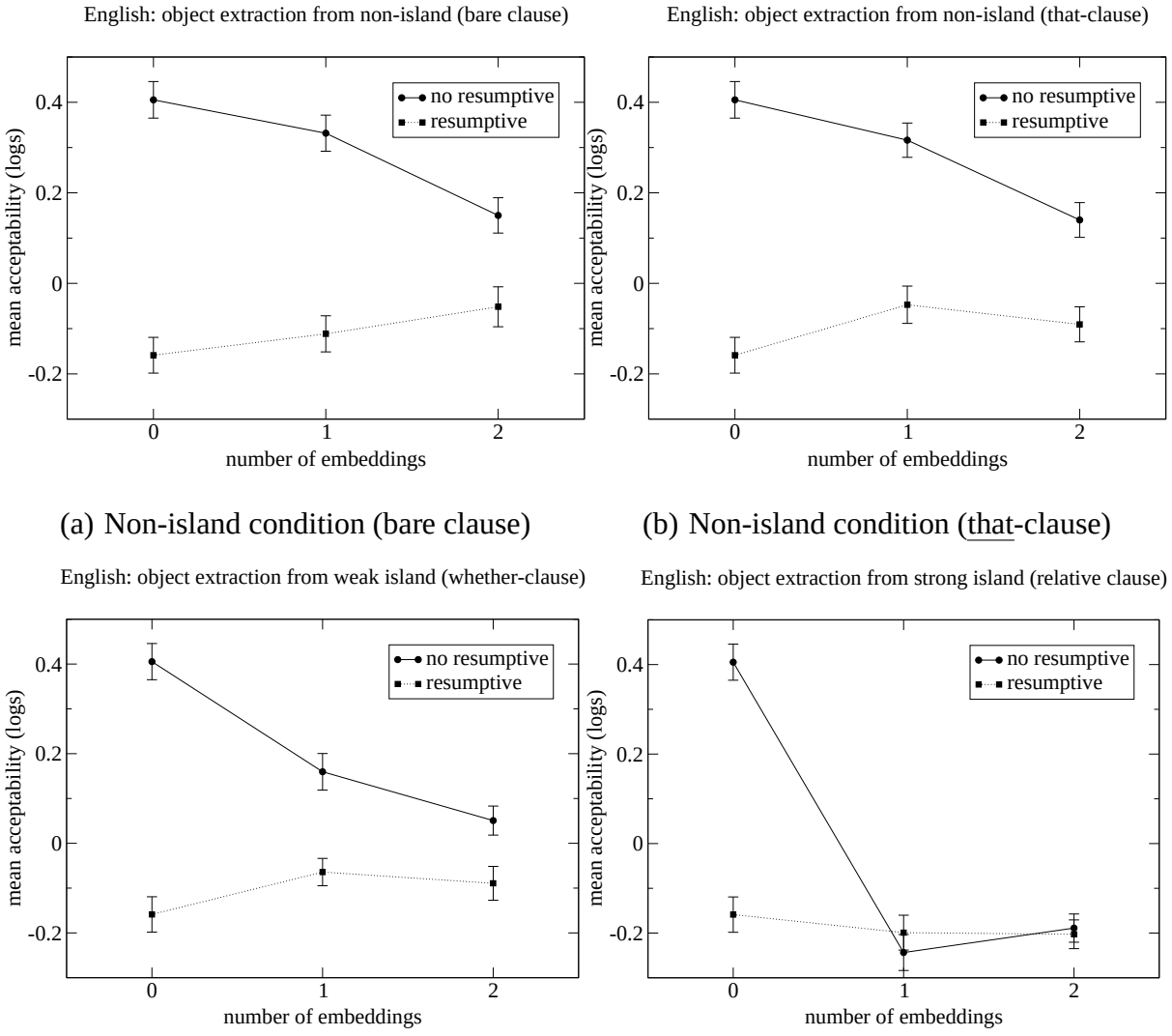
*: significant by subjects and items; (*): significant by subjects only; [*]: significant by items only

Table 4: Result of Dunnett tests comparing embedded clauses with the unembedded control in Experiment 4, Subexperiment 1

	Genitive antecedent				Prepositional antecedent			
	Gap		Resumptive		Gap		Resumptive	
	Single Embd.	Double Embd.	Single Embd.	Double Embd.	Single Embd.	Double Embd.	Single Embd.	Double Embd.
That-clause			*					
Whether-clause			(*)					

*: significant by subjects and items; (*): significant by subjects only

Table 5: Result of Dunnett tests comparing embedded clauses with the unembedded control in Experiment 4, Subexperiment 3



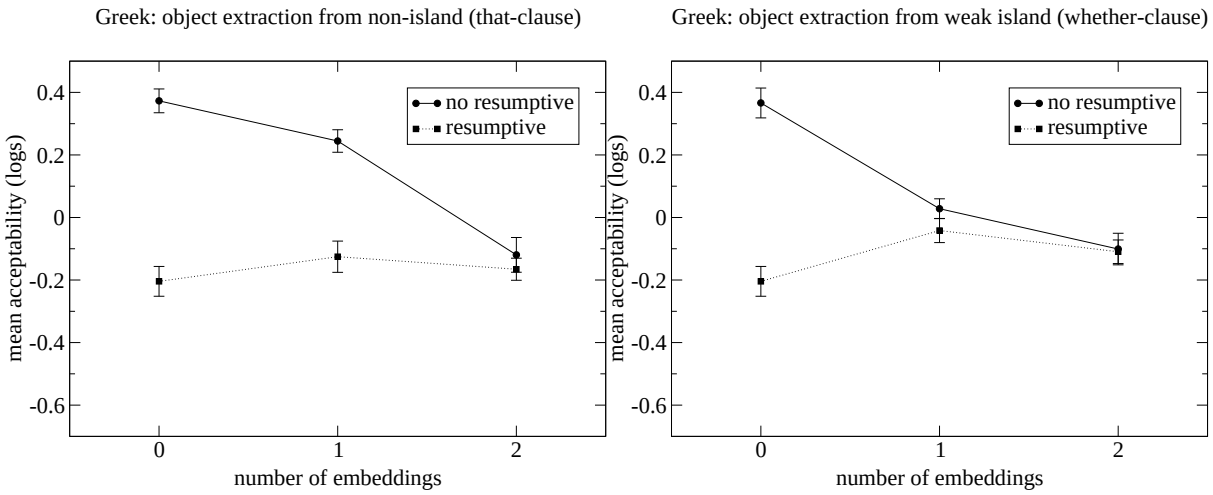
(a) Non-island condition (bare clause)

(b) Non-island condition (that-clause)

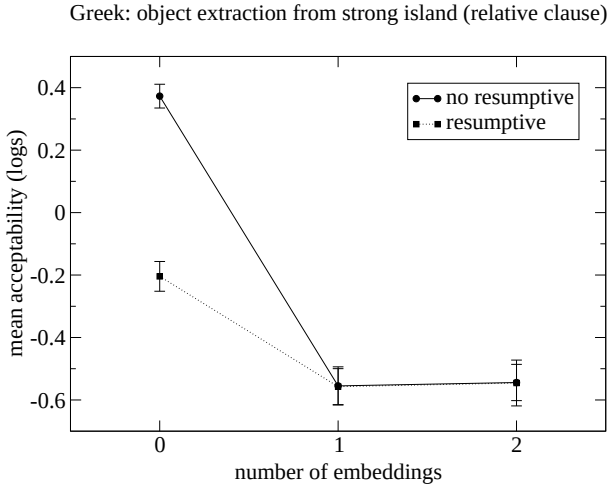
(c) Weak island condition (whether-clause)

(d) Strong island condition (relative clause)

Figure 1: Effect of embedding and resumption on object extraction in English in Experiment 1

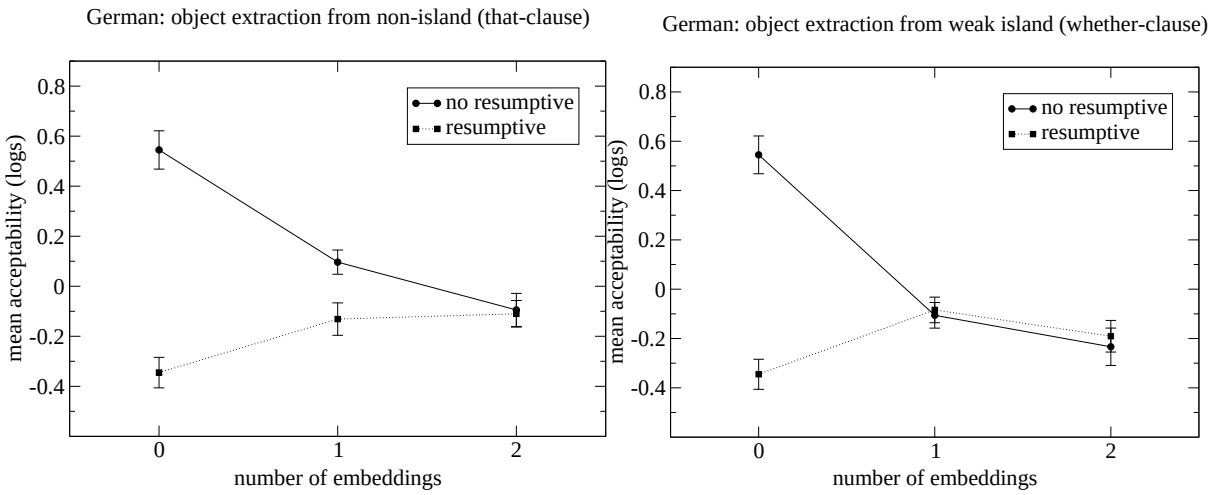


(a) Non-island condition (that-clause) (b) Weak island condition (whether-clause)

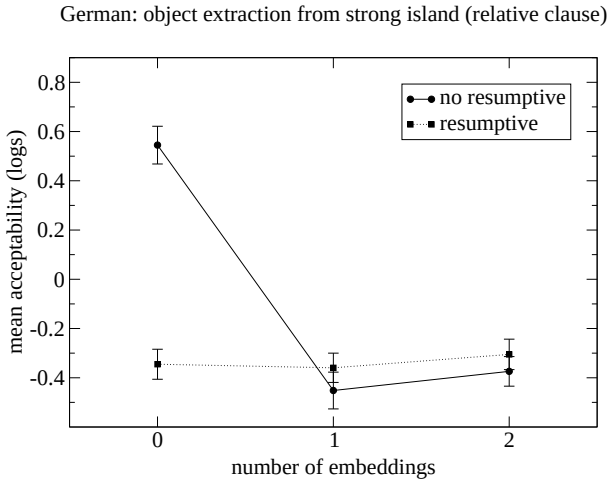


(c) Strong island condition (relative clause)

Figure 2: Effect of embedding and resumption on object extraction in Greek in Experiment 2

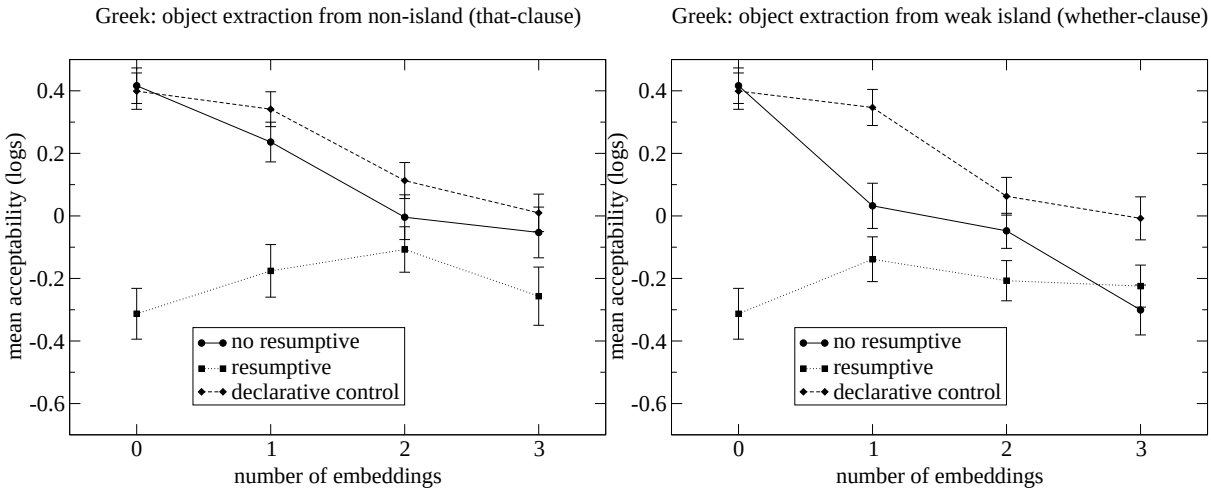


(a) Non-island condition (that-clause) (b) Weak island condition (whether-clause)



(c) Strong island condition (relative clause)

Figure 3: Effect of embedding and resumption on object extraction in German in Experiment 3



(a) Non-island condition (that-clause) (b) Weak island condition (whether-clause)

Figure 4: Effect of embedding and resumption on object extraction in Greek in Experiment 4, Subexperiment 1

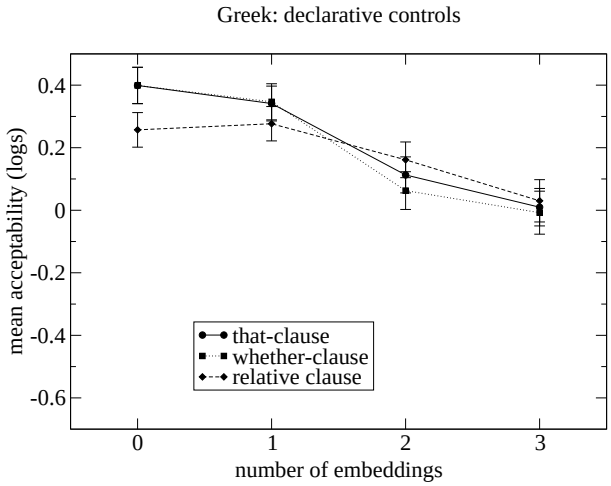
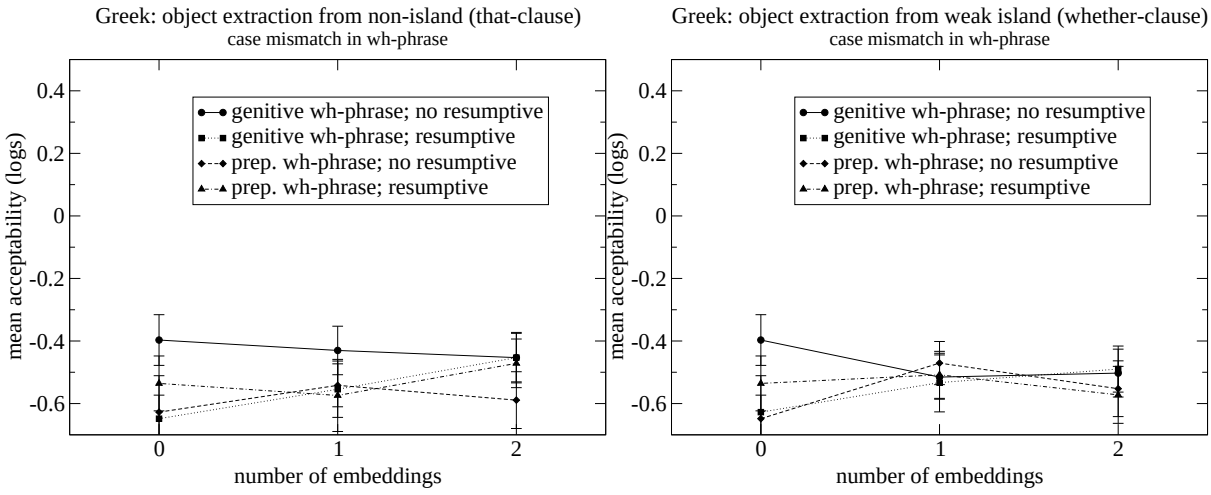


Figure 5: Effect of embedding on declaratives in Greek in Experiment 4, Subexperiment 2



(a) Non-island condition (that-clause) (b) Weak island condition (whether-clause)

Figure 6: Effect of embedding and resumption on object extraction with case mismatch in Greek in Experiment 4, Subexperiment 3