A Crosslinguistic, Experimental Study of Resumptive Pronouns and *that*-Trace Effects

Frank Keller (keller@inf.ed.ac.uk)
School of Informatics, University of Edinburgh
2 Buccleuch Place, Edinburgh EH8 9LW, UK

Theodora Alexopoulou (ta259@cam.ac.uk)
Department of Linguistics, University of Cambridge
Sidgwick Avenue, Cambridge CB4 9DA, UK

Abstract

We present two magnitude estimation experiments on the role of resumptive pronouns in subject extraction in English and German. In the theoretical literature, resumptives have been claimed to remedy effects of island violations, but our results show that this is not the case. Furthermore, our data also confirm the existence of a *that*-trace effect for both languages, contrary to theoretical claims that this effect is not present in German. On a methodological level, our study demonstrates the importance of controlled experiments for obtaining theoretically crucial linguistic judgments, and the potential of the magnitude estimation paradigm for crosslinguistic comparisons.

Keywords: grammaticality judgments; magnitude estimation; resumptive pronouns; island constraints; crosslinguistic experiments.

Introduction

Linguistic Intuitions The data on which linguists base their theories typically consist of grammaticality judgments, i.e., intuitive judgments of the well-formedness of utterances in a given language. When a linguist obtains a grammaticality judgment, he or she performs a small experiment on a native speaker; the resulting data are behavioral data in the same way as other measurements of linguistic performance (e.g., the reaction time data used in psycholinguistics). However, in contrast to experimental psychologists, linguists are generally not concerned with methodological issues, and typically none of the standard experimental controls are imposed in collecting data for linguistic theory. Such methodological negligence can seriously compromise the data obtained as demonstrated by Schütze (1996), who argues for a more reliable mode of data elicitation in linguistics, based on standard methods from experimental psychology.

Bard et al. (1996) and Cowart (1997) demonstrated how the experimental paradigm of magnitude estimation (ME) makes it possible to address problems such as the ones raised by Schütze. ME is an experimental technique standardly used in psychophysics to measure judgments of sensory stimuli (Stevens, 1975). It requires subjects to estimate the magnitude of physical stimuli by assigning numerical values proportional to the stimulus magnitude they perceive. Highly stable judgments can be achieved for a whole range of sensory modalities, such as brightness, loudness, or tactile stimulation. Bard et al. (1996) demonstrated that linguistic judgments can be elicited in the same way as judgments of sensory stimuli, and that ME can yield reliable and fine-grained measurements of linguistic intuitions. ME has since been applied successfully to a range of linguistic phenomena, including binding (Featherston, 2002), superiority (Meyer, 2003), and word order (Keller & Alexopoulou, 2001).

The present paper applies the ME methodology to a longstanding dispute in linguistic theory, viz., the status of resumptive pronouns in extraction. While this is a topic of considerable interest in the literature, progress is hampered by disagreement on both the relevant data and on the theoretical conclusions to be drawn from these data. We address this problem by presenting two ME experiments that shed light on the status of resumptive pronouns in subject extraction in English and German. By running experiments on the same linguistic construction in two languages, we also introduce a methodological innovation: most previous ME studies were only concerned with data for a single language, thus missing out on crosslinguistic generalizations, which are crucial for modern linguistic theory.

Islands and Resumption in Object Extractions Ever since Ross (1967), island constraints have been a major focus of research in syntactic theory. Islands are syntactic environments that do not allow extraction, i.e., an element that occurs within the island can not be dislocated by *wh*-extraction in order to form, e.g., a question or relative clause. (1) gives a number of standard examples: (1-a) is a case of an extraction from a subordinate clause, which is acceptable, as subordinate clauses are not islands for extraction (note the use of *t* to indicate the **gap**, i.e., the site from which the extraction took place). (1-b) is an extraction from an indirect question (*whether*-clause). This extraction is of intermediate acceptability; indirect questions are referred to as **weak islands**. Finally, (1-c) illustrates extraction out of a relative clauses, which is severely unacceptable: relative clauses are **strong islands** for extraction.

(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*?

Island constraints are known to interact with **resumptive pronouns**, i.e., pronouns that function as an overt realization of the gap *t* left behind by extraction. In particular, Ross (1967) and Sells (1984) claim that the presence of a resumptive pronoun can "save" island violations. The expectation then is that the sentences with resumptives in (2) should be more acceptable than the corresponding sentences with gaps in (1).

(2) a. Who did Mary wonder whether they will fire him?
     b. Who did John meet the girl who will marry him?

However, a recent magnitude estimation study of the interaction between resumption and islands in extraction from object questions has only partially confirmed this picture (Alexopoulou & Keller, 2003). Weak islands were indeed found to incur a significant drop in acceptability for examples like (1-b), which were in turn significantly more acceptable than examples like (1-c). But interestingly, **embedding** per se was also found to reduce acceptability (in particular for double embedded gaps like *who did John think Peter claimed Mary would choose?*). Furthermore, contrary to claims in the literature, resumption was not found to "save" island violations: resumptives were at most as acceptable as gaps. However, the acceptability of resumptives did improve when they were embedded in a *that-* or *whether*-clause, indicating that resumption reverses the effect of weak islands and embedding on questions (though it cannot restore them to full acceptability). The present study extends this investigation to subject questions and further explores the nature of the embedding effect by comparing questions with declaratives.

Subject Questions Subject questions differ from object questions in English in a crucial way: they are ungrammatical (see (3-b)), unless the complementizer is absent (see (3-c)). The phenomenon is standardly referred to as the ***that*-trace effect** (Beletti & Rizzi, 1988).

(3) a. Who do you think that Mary will invite *t*?
     b. *Who do you think that *t* will invite Mary?
     c. Who do you think *t* will invite Mary?

Crosslinguistically, authors like Lutz (1996) have claimed that, unlike English, no *that*-trace effects are present in German, as indicated by the grammaticality of (4) (ibid., p. 36).

(4) Wer glaubt   Peter, dass ihn *t* gesehen hat?
     who believes Peter that him *t* seen   has

In this paper, we present two magnitude estimation studies investigating the interaction between islands and subject extractions. We compare the status of *that*-trace violations with that of strong and weak island violations in English and German. We further explore the interaction between islands and resumption which, as in the case of object extractions, has been argued to remedy *that*-trace violations, as in the following examples ((5) is from Sells (1984), and (6) from Mc-Daniel & Cowart (1999)):

(5) Which woman does no Englishman even wonder whether she will make a good wife?

(6) That's the girl that I wonder whether she met you.

Finally, we compare the effect of embedding in questions and in declaratives, in order to establish whether embedding affects long distance dependencies only.

Experiment 1: Resumptives and Subject Extraction in English

This experiment investigated four different types of islands: complement clause without *that* (no island), complement clause with *that* (no 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 complement clause embedding another 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 below.

(7) **No island violation** (bare clause)
     a. Who will *t*/he fire Sue? **(zero embedding)**
     b. Who does Mary claim *t*/he will fire Sue? **(single)**
     c. Who does Jane think Mary claims *t*/he will fire Sue? **(double)**

(8) **No island violation** (*that*-clause)
     a. Who does Mary claim that *t*/he will fire Sue? **(single)**
     b. Who does Jane think that Mary claims that *t*/he will fire Sue? **(double)**

(9) **Weak island violation** (*whether*-clause)
     a. Who does Mary wonder whether *t*/he will fire Sue? **(single)**
     b. Who does Jane think that Mary wonders whether *t*/he will fire Sue? **(double)**

(10) **Strong island violation** (relative clause)
     a. Who does Mary meet the people that *t*/he will fire? **(single)**
     b. Who does Jane think that Mary meets the people that *t*/he will fire? **(double)**

In addition to the *wh*-extraction conditions, a declarative control condition was included: each question item was matched by the corresponding declarative; example (11) gives the declaratives corresponding to the questions in (7).

(11) **Declarative** (bare clause)
     a. Mary will fire Sue. **(zero embedding)**
     b. Jane claims Mary will fire Sue. **(single)**
     c. Rachel thinks Jane claims Mary will fire Sue. **(double)**

Method Forty-four 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.

The design crossed the following factors: EMBEDDING (single or double), ISLAND (complement clause without *that*, complement clause with *that*, complement clause with *whether*, relative clause), and RESUMPTION (gap, resumptive, declarative). This resulted in EMBEDDING × ISLAND × RESUMPTION $= 2 \times 4 \times 3 = 24$ cells. As controls, we included stimuli without embedding (gap, resumptive, declarative), increasing the number of cells to 27. Nine lexicalizations were used for each cell, yielding a total of 243 stimuli. The stimulus set was divided into nine subsets of 27 stimuli by placing the items in a Latin square. A set of 30 fillers was used, covering the whole acceptability range.

The method used was ME as proposed by Stevens (1975) for psychophysics and extended to linguistic stimuli by Bard et al. (1996) and Cowart (1997). Subjects first saw a set of in-

Clause	Gap more acc. than resmpt.	Declarative more acc. than resmpt.	Declarative more acc. than Gap
Bare	**	**	
That		**	* (**)
Whether		**	**
Relative		**	**

*: $p < .05$; **: $p < .01$; (): by subjects only

Table 1: Result of Tukey test for interaction of ISLAND and RESUMPTION in Experiment 1

	Gap		Resumptive		Declarative	
Clause	Single Embd.	Double Embd.	Single Embd.	Double Embd.	Single Embd.	Double Embd.
Bare	(*)	* (**)				(**)
That	**	**				* (**)
Whether	**	**				(**)
Relative	**	**			(*)	* (**)

*: $p < .05$; **: $p < .01$; (): by subjects only

Table 2: Result of Dunnett tests comparing embedded clauses with the unembedded control in Experiment 1

structions that explained the concept of numerical ME using line length. Subjects were instructed to make length estimates relative to the first line they would see, the reference line. They were told to give the reference line an arbitrary number, and then assign a number to each following line so that it represented how long the line was in proportion to the reference line. Several example lines and corresponding numerical estimates were provided to illustrate the concept of proportionality. Then subjects were told that linguistic acceptability could be judged in the same way as line length, i.e., by comparing the acceptability of a target sentence to that of a reference sentence. The task was illustrated by examples. The experiment was conducted over the web using WebExp (Keller et al., 1998), an interactive software package for web-based psycholinguistic experimentation.

Results The data were normalized by dividing each numeric judgment by the value that the subject had assigned to the reference sentence. This operation creates a common scale for all subjects. Then the data were 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). All analyses and figures are based on normalized, log-transformed judgments. Figure 1 graphs the mean judgments for all four island conditions.

An ANOVA yielded significant main effects of EMBEDDING ($F_1(1,43) = 17.482$, $p < .0005$; $F_2(1,8) = 14.525$, $p = .005$), ISLAND ($F_1(3,129) = 25.577$, $p < .0005$; $F_2(3,24) = 9.055$, $p < .0005$), and RESUMPTION ($F_1(2,86) = 57.742$, $p < .0005$; $F_2(2,16) = 213.4149$, $p < .0005$). The interactions ISLAND/RESUMPTION ($F_1(6,258) = 15.552$, $p < .0005$; $F_2(6,48) = 3.908$, $p = .003$) and EMBEDDING/RESUMPTION ($F_1(2,86) = 12.933$, $p < .0005$; $F_2(2,16) = 6.299$, $p = .010$) were significant, while the ISLAND/EMBEDDING interaction was only significant by subjects ($F_1(3,129) = 4.827$, $p = .003$; $F_2(3,24) = 1.377$, $p = .274$), and the ISLAND/EMBEDDING/RESUMPTION failed to reach significance.

A post-hoc Tukey test was conducted to further investigate the interaction of ISLAND and RESUMPTION. Its results are tabulated in Table 1. We also carried out a series of Dunnett tests to compare the single and double embedding conditions with unembedded controls; see Table 2.

Discussion We found a significant interaction of resumption and embedding in this experiment, which demonstrates that the various types of islands tested show differential behavior with respect to subject extraction. A post-hoc test detected the locus of this effect: for bare subordinate clauses, questions with gaps are as acceptable as declaratives, while questions with resumptives are significantly less acceptable. In the other three island conditions (*that*-clause, *whether*-clause, relative clause), the opposite pattern was found: gaps and resumptives were equally unacceptable, and both were significantly less acceptable than declaratives (see Table 1).

This finding provides evidence for the *that*-trace effect, in line with the theoretical literature: subject extraction out of bare complement clauses is acceptable (i.e., as acceptable as declaratives), while subject extraction out of *that*-clauses is unacceptable (i.e., less acceptable than declaratives), see Cowart (1997) for similar results on *that*-trace effects in English. Furthermore, our results show that *whether*-clauses and relative clauses behave like *that*-clauses in this construction.

Regarding the effect of embedding, the following pattern emerged: for gapped clauses, the single and double embedding conditions were significantly less acceptable than the unembedded control condition. This pattern was consistent across all types of islands. In the resumptive condition, however, there was no difference between the single and double embedded condition and the unembedded control; again this pattern was the same across all types of islands (see Table 2).

Taken together, these results clearly demonstrate that there is no effect of resumption for subject extraction: resumptives are never able to compensate for the effect of embedding, i.e., they are equally bad in embedded and unembedded conditions. Moreover, gaps were as bad as resumptives in all island conditions (with the exception of the bare complement clause, where we found the *that*-trace effect). Note in particular that this means that there is no difference between subject extraction out of *that*-, *whether*-, and relative clauses; these constructions all behave like strong islands, in the sense that extraction is as bad as an unembedded resumptive. This contrasts sharply with object extractions, where this pattern was found only for relative clauses (Alexopoulou & Keller, 2003). Moreover, this contrasts also with the results of McDaniel & Cowart (1999) who investigated relative clauses instead of questions and found that resumption can remedy *that*-trace violations for relativization, as illustrated by the following pattern:

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

Finally, the results in the declarative control condition are surprising: we demonstrated that there is a penalty for embedding even for declaratives: double embedding is less acceptable than single embedding for all types of embedding; however, single embedding is not less acceptable than no embedding, with the exception of the relative clause condition.
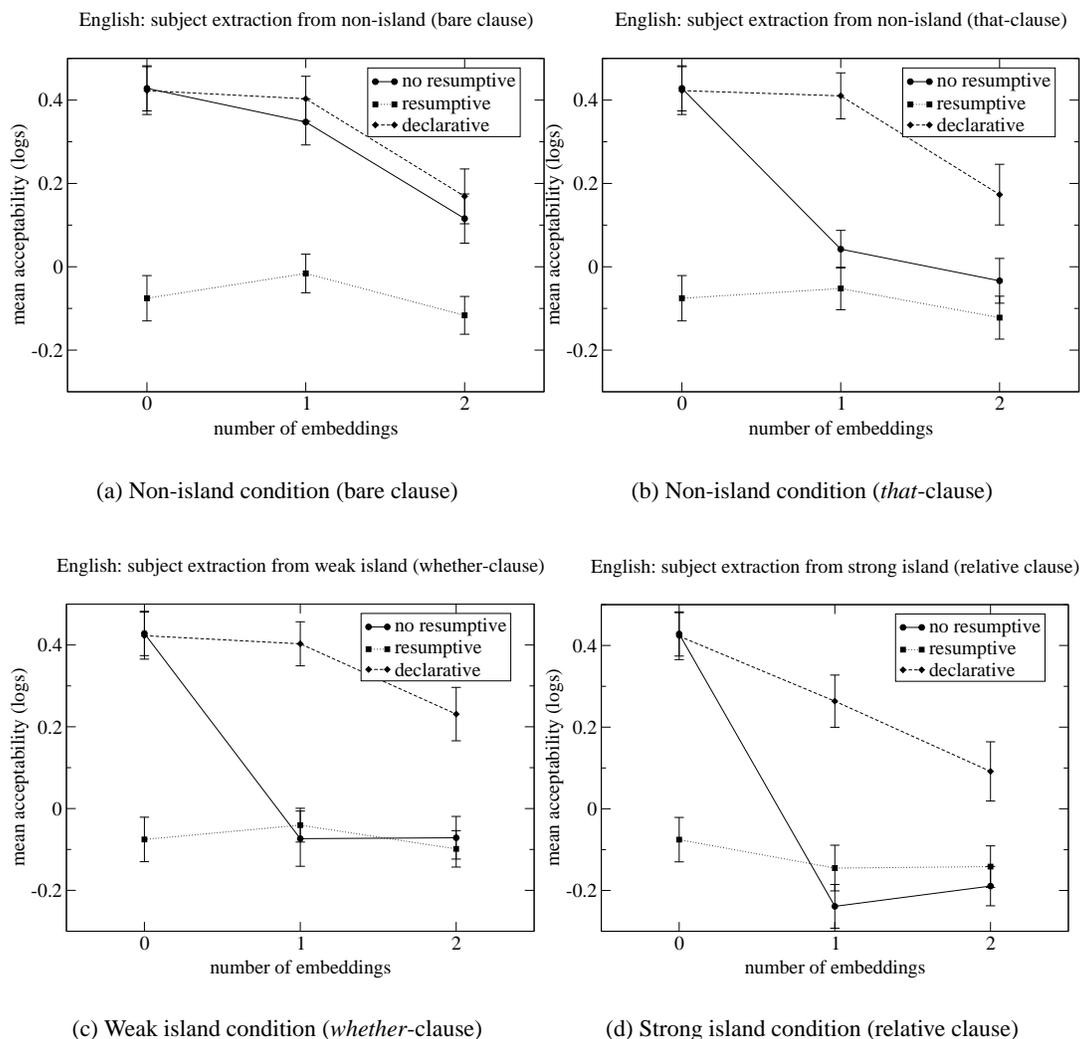
English: subject extraction from non-island (bare clause)

(a) Non-island condition (bare clause)

English: subject extraction from non-island (that-clause)

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

English: subject extraction from weak island (whether-clause)

(c) Weak island condition (*whether*-clause)

English: subject extraction from strong island (relative clause)

(d) Strong island condition (relative clause)

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

Experiment 2: Resumptives and Subject Extraction in German

This experiment replicated the results of Experiment 1 for German. As outlined in the introduction, there are well-documented differences between English and German with respect to the *that*-trace effect; the aim of the present experiment was therefore to assess if these differences also lead to difference in acceptability of resumptives in subject extraction in German.

Unlike English, embedded clauses in German are verb final, and the complementizer is obligatory (see (13-a)). However, complementizer-less embedding is possible for certain matrix verbs; in this case the word order has to be verb second (V2) instead of verb final, and the complementizer has to be omitted (see (13-b)).[1]

---

[1]The status of V2-embedding has been disputed. For example, Reis (1995) argues that this construction is in fact a case of paren- thesis, rather than embedding.

(13) a. Wer behauptet Petra, dass *t*/er Anna entlässt?
    who claim-3sg Petra that *t*/he Anna fire-3sg
 b. Wer behauptet Petra, entlässt *t*/er Anna?
    who claim-3sg Petra fire-3sg *t*/he Anna

The other two types of embedding tested were the same as for Experiment 1: *whether*-clauses and relative clauses (cor- responding to the English examples (9) and (10)). Again, a declarative control condition was included as a baseline to compare with the subject questions.

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

The experimental design was the same as in Experiment 1, but German materials were used, as illustrated by the exam- ples in (13-b). The experimental procedure was identical to that of Experiment 1.
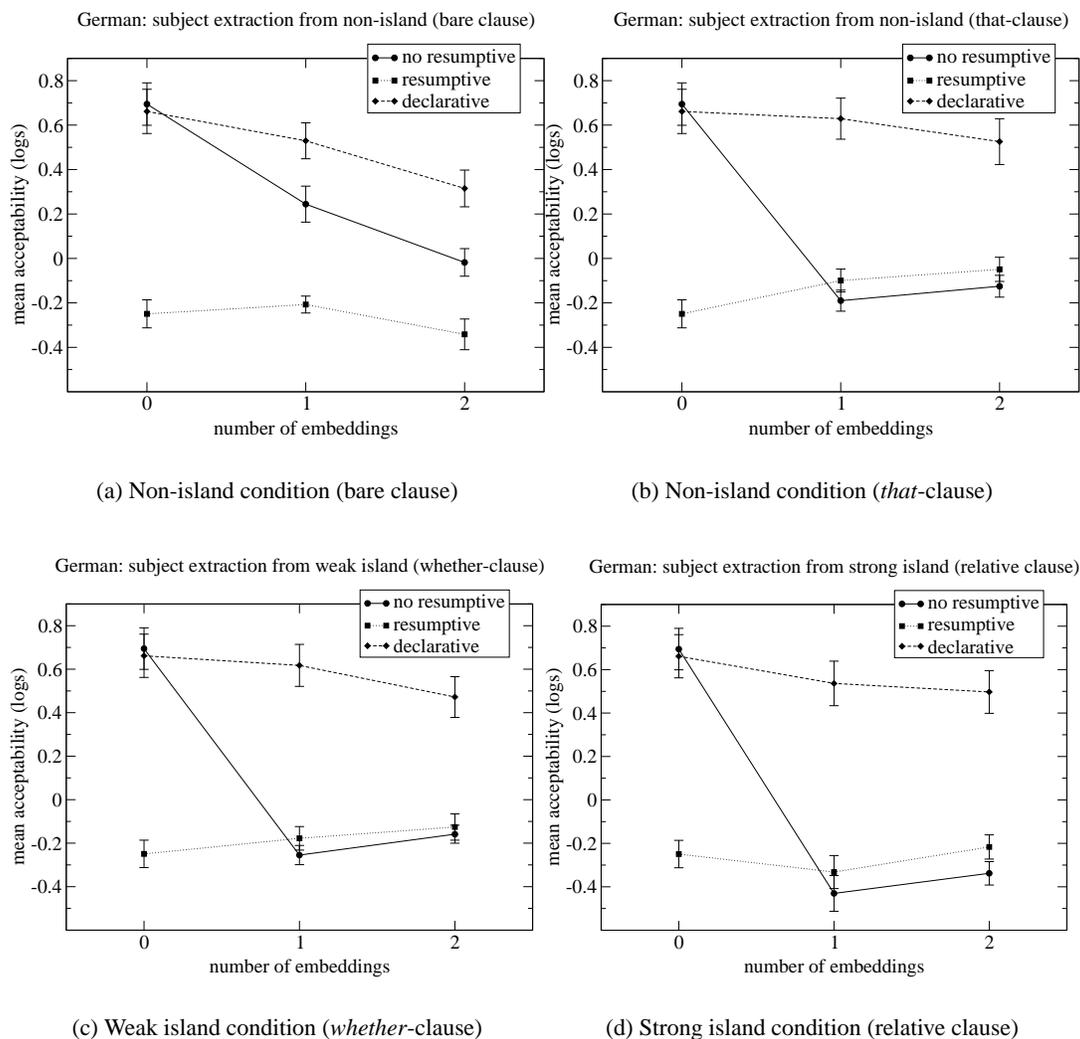
German: subject extraction from non-island (bare clause)

(a) Non-island condition (bare clause)

German: subject extraction from non-island (that-clause)

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

German: subject extraction from weak island (whether-clause)

(c) Weak island condition (*whether*-clause)

German: subject extraction from strong island (relative clause)

(d) Strong island condition (relative clause)

Figure 2: Effect of embedding and resumption on subject extraction in German in Experiment 2

Results The data were normalized and log-transformed as in Experiment 1. Figure 2 graphs the mean judgments for all four island conditions.

An ANOVA yielded main effects of ISLAND ($F_1(3,78) = 9.764$, $p < .0005$; $F_2(3,24) = 3.942$, $p = .020$) and RE-SUMPTION ($F_1(2,52) = 88.483$, $p < .0005$; $F_2(2,16) = 299.130$, $p < .0005$). The main effect of EMBED-DING failed to reach significance. The interactions IS-LAND/RESUMPTION ($F_1(6,156) = 18.055$, $p < .0005$; $F_2(6,48) = 12.888$, $p < .0005$) and ISLAND/EMBEDDING ($F_1(3,78) = 8.945$, $p < .0005$; $F_2(3,24) = 5.802$, $p = .004$) were significant, while the EMBEDDING/RESUMPTION interaction was significant by subjects only ($F_1(2,52) = 6.631$, $p = .007$; $F_2(2,16) = 1.581$, $p = .236$), and the IS-LAND/EMBEDDING/RESUMPTION interaction failed to reach significance.

A post-hoc Tukey test was conducted to further investigate the interaction of ISLAND and RESUMPTION. Its results are tabulated in Table 3. We also carried out a series of Dunnett tests to compare the single and double embedding conditions with unembedded controls; see Table 4.

Discussion The results of this experiment largely confirm the results of Experiment 1 for English. A significant inter-action of resumption and embedding was found, and a post-hoc test demonstrated that gaps were more acceptable than resumptives in bare subordinate clauses (V2 clauses), just like in English (see Table 3). However, we also found that gaps were less acceptable than declaratives in German bare clauses (see Figure 2(a)), while in English, there was no sig-nificant difference (see Figure 1(a)). In the other three island conditions (*that*-, *whether*-, and relative clauses) the findings in both languages were identical, i.e., gaps were as bad as resumptives, while declaratives were significantly better than both gaps and resumptives.

These results demonstrate the presence of *that*-trace effects in German, despite claims to the contrary in the literature (Lutz, 1996). Similar results were obtained by Featherston (2005) (who only deals with gaps, not with resumptives).

Clause	Gap more acc. than resmpt.	Declarative more acc. than resmpt.	Declarative more acc. than Gap
Bare	**	**	* (**)
That		**	**
Whether		**	**
Relative		**	**

*: $p < .05$; **: $p < .01$; (): by subjects only

Table 3: Result of Tukey test for interaction of ISLAND and RESUMPTION in Experiment 2

Clause	Gap Single Embd.	Gap Double Embd.	Resumptive Single Embd.	Resumptive Double Embd.	Declarative Single Embd.	Declarative Double Embd.
Bare	* (**)	**				(**)
That	**	**		(*)		(**)
Whether	**	**				
Relative	**	**			(*)	(*)

*: $p < .05$; **: $p < .01$; (): by subjects only

Table 4: Result of Dunnett tests comparing embedded clauses with the unembedded control in Experiment 2

Regarding the effect of embedding, we found that single and double embedded gapped sentences were less acceptable than an unembedded control, in all four island conditions (see Table 4). This is also what we found for subject extraction in English. For resumptives, we found that double embedded *that*-clauses were more acceptable than the unembedded control. This means that there is evidence for resumption reducing the unacceptability of embedding in this condition, i.e., for the "saving" effect of resumption described in the theoretical literature. However, that these "saved" clauses are still strongly unacceptable. Note that this effect was not present in subject extraction in English.

Finally, we found that embedding has an effect on acceptability in German, even for declaratives. This effect is consistent with the findings for English, though the effect is weaker (it only extends to bare clauses, *that*-clauses, and relative clauses).

Conclusion

In this paper, we presented an experimental study of resumptive pronouns, which in the theoretical linguistics literature have been claimed to remedy the unacceptability caused by extraction from islands. In particular, we investigated the interaction between resumption and the *that*-trace effect, i.e., the observation that subject extraction out of subordinate clauses is acceptable only if the complementizer *that* is absent.

We obtained three main results: Firstly, we found that resumptive pronouns can not remedy island violations, contrary to claims in the theoretical literature. For all types of embedding and for both languages we investigated (English and German), we found that resumptive pronouns are seriously unacceptable, and never better than the corresponding gaps (there was one possible exception in the German data, see Experiment 2). Secondly, we demonstrated the existence of a *that*-trace effect in both English and German. While this effect is well-documented for English, the theoretical literature has so far claimed the absence of this effect for German. Thirdly, we found that embedding leads to a general decrease

in acceptability, even in the declarative control condition (i.e., even when extraction is present). This was particularly pronounced for the second level of embedding and for relative clauses.

All three results contradict claims in the syntactic literature. This demonstrates that it is important to obtain experimental data for theoretically crucial constructions: the relevant data are complex and involve gradient acceptability judgments which cannot be reliably obtained through the introspection of a single native speaker. We expect this to be the case not only for resumption, but for a large number of theoretically relevant judgment data. Furthermore, our study also demonstrates the usefulness of crosslinguistic experimentation, which has rarely been attempted before.

References

Alexopoulou, T., & Keller, F. (2003). Linguistic complexity, locality and resumption. In *Proceedings of the 22nd West Coast Conference on Formal Linguistics*, (pp. 15–28), Somerville, MA. Cascadilla Press.

Bard, E. G., Robertson, D., & Sorace, A. (1996). Magnitude estimation of linguistic acceptability. *Language*, 72, 32–68.

Beletti, A., & Rizzi, L. (1988). Psych-verbs and θ-theory. *Natural Language and Linguistic Theory*, 6, 291–352.

Cowart, W. (1997). *Experimental Syntax: Applying Objective Methods to Sentence Judgments*. Thousand Oaks, CA: Sage Publications.

Featherston, S. (2002). Coreferential objects in German: Experimental evidence on reflexivity. *Linguistische Berichte*, 192, 457–484.

Featherston, S. (2005). Magnitude estimation and what it can do for your syntax: Some wh-constraints in German. *Lingua, to appear*.

Keller, F., & Alexopoulou, T. (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, F., Corley, M., Corley, S., Konieczny, L., & Todirascu, A. (1998). *WebExp: A Java Toolbox for Web-based Psychological Experiments*. Technical Report HCRC/TR-99, Human Communication Research Centre, University of Edinburgh.

Lutz, U. (1996). Some notes on extraction theory. In U. Lutz, & J. Pafel (eds.), *On Extraction and Extraposition in German*, (pp. 1–44). Amsterdam: John Benjamins.

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

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

Reis, M. (1995). Wer glaubst du hat recht? On so-called extractions from verb-second clauses and verb-first parenthetical constructions in German. *Sprache und Pragmatik*, 36, 27–83.

Ross, J. R. (1967). *Constraints on Variables in Syntax*. Ph.D. thesis, MIT.

Schütze, C. T. (1996). *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology*. Chicago: University of Chicago Press.

Sells, P. (1984). *Syntax and Semantics of Resumptive Pronouns*. Ph.D. thesis, University of Massachusetts at Amherst.

Stevens, S. S. (1975). *Psychophysics: Introduction to its Perceptual, Neural, and Social Prospects*. New York: John Wiley.