On the Usefulness of Formal Judgment Tasks in Syntax and in Second-Language Research: The Case of Resumptive Pronouns in English, Turkish, and Mandarin Chinese

Key words: acceptability judgments, resumptive pronouns, experimental syntax

Abstract

Fedorenko and Gibson (2013) have argued against the continued use of informally collected acceptability judgments as the primary methodology in theoretical syntax and semantics research. We provide further support for their position with data from Mandarin and Turkish-language judgment tasks which examined the acceptability of resumptive pronouns (RPs) in relative clauses. Based on previous studies which relied on informal judgments, we expected that RPs should be permitted in certain types of Mandarin relative clauses, but ungrammatical in comparable Turkish relative clauses. The results failed to replicate this contrast: RPs were more acceptable than expected in Turkish, and less acceptable than expected in Mandarin. Furthermore, the Mandarin Chinese experiment showed an unexpected gradient effect. We argue that these results challenge existing theoretical accounts, support the more widespread adoption of experimental tasks in theoretical linguistics and in second-language research, and consistently support the Filler-Gap Domain complexity ranking as proposed by Hawkins (2004). We use the complexity ranking and its supporting evidence as a case study demonstrating that quantitative data, such as the evidence obtained from formal sentence judgment tasks, are indispensable in the defense or criticism of linguistic theories.
1. Introduction

Judgments of the acceptability of any given syntactic structure are the bedrock of arguments in linguistics and in second language acquisition. To take a well-known example, wh-islands constraints have received a great deal of attention in mainstream linguistics and in the second language acquisition literature:

(1) *The boy who Mary described [the way [that Bill attacked t]] is here.

(Hawkins and Chan 1997: 191)

Based on acceptability judgments such as (1), practitioners of mainstream generative grammar have argued that knowledge of the constraints forbidding illicit wh-movements, whether formulated in terms of Shortest Move (Chomsky 1995), Subjacency (Chomsky 1977), or some other principle, must be innately specified in Universal Grammar (henceforth, UG). Second language acquisition theorist Roger Hawkins (2005) accepts the argument that acquisition of an uninterpretable wh-feature (a feature invisible to the semantic grammar component) allows native speakers of English to understand that the wh-movements exemplified in in (1) are illicit, and goes on to argue that, after the critical period (i.e. as teens or adults), native speakers of languages such as Cantonese or Mandarin Chinese cannot obtain the uninterpretable feature that is necessary to develop a sensitivity to movement violations such as (1), because this feature is not selected during their acquisition of Chinese. Without prolonged exposure to English during the critical period, it will be impossible for them to develop an interlanguage grammar allowing wh-movement¹, and consequently, sensitivity to constraints on wh-

¹ According to Roger Hawkins, although native speakers of languages that do not allow wh-movement such as Chinese and Japanese may appear to have acquired wh-movement in
movement in English. For this reason, it is argued, adult L1 Chinese learners of English find
English sentences such as (1) acceptable. Thus, acceptability judgments of sentences such as (1)
(and their Chinese equivalents) are part of the foundation of arguments about the nature of the
language faculty, and about the nature of second language acquisition.

The viability of any linguistic argument based on acceptability judgments can only be as
sound as its foundation—the acceptability judgments themselves. Such judgments can vary
according to many different factors, including declarative knowledge of grammar rules, dialect,
discourse context, frequency of use, and the limitations of working memory (Schütze 1996;
Hofmeister and Norcliffe 2013). For these reasons, some prominent linguists have argued that
acceptability judgments be performed formally or experimentally whenever possible so as to
avoid the formulation of a complex chain of reasoning based on faulty foundations or premises
(Gibson and Federenko 2013; Gibson, Piantadosi, and Federenko 2013; Wasow and Arnold
2005). Traditionally, however, acceptability judgments as reported in the syntax literature have
been informal, meaning that they are not administered to large numbers of participants, and
do not employ a factorial design or tests of statistical significance, as is the case for formal
acceptability judgment tasks. Informal acceptability judgments may consist of one researcher

English questions, these appearances are deceiving, and their L2 linguistic performance may not
directly reflect their underlying L2 linguistic competence: “The present study suggests that,
conversely, there may be cases where apparent target-like performance conceals non-target-
like underlying competence. I have argued that although high proficiency L2 speakers of English
with Japanese as their L1 might look like they have acquired wh-movement driven by a [uwh*:] feature, it turns out that they have a different representation, one consistent with the claim
that uninterpretable features not present in a speaker’s L1 may be inaccessible in later second
using personal introspection or consulting with one native speaker of the language to
determine the acceptability of a given construction (Gibson and Fedorenko 2013), or a linguist
who shares his or her judgments with a group of other linguists at a conference (Sprouse and
Alemeida 2013a). Defenders of informal acceptability judgments counter that the use of formal
judgment tasks is not usually necessary, and that the use of informal judgments rarely leads to
arguments based on mistaken premises (Phillips 2009, Sprouse and Alemeida 2013a, Sprouse,
Schütze, and Almeida, 2013). Sprouse and Almeida (2013b) claim that existing data based on
informal acceptability judgments is in no way flawed: “...the current state of evidence
suggests that questions about the veracity of existing judgment data may have been a
(historically driven) distraction: there appears to be no evidence that the existing data is faulty,
and growing evidence that the traditional methods are appropriate for the majority of
phenomena of interest to syntacticians” (202).

Following Gibson and Fedorenko (2013) and Gibson, Piantadosi, and Fedorenko (2013),
we endorse the use of formal acceptability judgment tasks (henceforth AJTs) or other
quantitative methods to ascertain the nature of the types of grammatical constraints that might
affect a language. The current study presents three previously unpublished experiments from
John Hitz’s 2012 dissertation, providing further support for this position. While Gibson and
Fedorenko’s (2013) experiments focused on implications for mainstream research in syntax, the
experiments reported here have important implications both for mainstream syntax research
and for research in second language acquisition (henceforth SLA). We will show how formal
AJTs can provide data relevant to the accurate formulation of grammatical constraints in native
language grammars, which can in turn inform our understanding of potential L1 transfer effects in the interlanguage grammars of L2 learners.

Hitz intended to elucidate the impact of language transfer in the second language acquisition of English relative clauses (henceforth, RCs) by native speakers of Turkish and Chinese, with respect to the acceptability of resumptive pronouns in RCs. Resumptive pronouns are pronouns which occur in the original (non-displaced) position of a displaced noun phrase, in a position where an empty gap would otherwise go. Although resumptive pronouns (henceforth RPs) commonly occur in RCs and other filler-gap constructions across the world’s languages, languages differ as to whether such pronouns are used productively or not. In English, for example, RPs have a marginal status and are typically judged by native speakers as unacceptable, although they do occur in spontaneous speech (Asudeh 2012: 41-43). In languages such as Hebrew and Arabic, however, RPs are used productively and are fully grammatical (or even required) in certain types of clauses (Asudeh 2012: 27). Furthermore, among languages which allow the productive use of resumptive pronouns, there are differences with regard to which structural positions license RPs and which positions license gaps. Such cross-linguistic differences in the productivity and structural licensing of RPs create challenges for L2 learners. For example, L2 learners of English sometimes produce and/or accept ungrammatical sentences such as in (2a):

(2)  a. *The relatives, who we visited them, last night enjoyed the evening.

    b. The relatives, who we visited __, last night enjoyed the evening.
In (2a), the speaker has inserted a RP (them) in what would normally be the empty gap (or trace) position of the extracted object, as indicated by ‘__’ in (2b). Relying on the informal acceptability judgments put forth in the linguistic literature on English, Chinese, and Turkish RCs, Hitz (2012) assumed that Turkish and English share a common structural feature: both languages license only gaps (not RPs) in simple (monoclausal) relative clauses, as illustrated in (2a-b). This is in contrast to Chinese RCs, in which RPs are said to be acceptable in non-subject positions. Thus, the Chinese equivalent of (2a) should be acceptable. Given this purported distribution of RPs in Turkish and Chinese RCs, Hitz (2012) hypothesized that native speakers of Chinese would accept RPs in English RCs, as in (2a), at a higher rate than native speakers of Turkish would, due to the effects of L1 transfer. He further hypothesized that for reasons related to processing complexity (Hawkins 2004), RPs would be beneficial for L2 learners’ processing of relative clauses, with this benefit showing up most strongly in the most complex clause types. Thus, even the Turkish group might accept RPs at a higher rate than the control group, and both learner groups should accept RPs in prepositional object relatives at a higher rate than in subject relatives. Contrary to the L1 transfer hypothesis, no statistically significant group differences between the L1 Chinese and L1 Turkish speakers were found. Rather, the two learner groups (but not the native speaker group) accepted RPs to a greater extent in direct object and prepositional object RCs as compared with subject RCs (as would be expected on a processing account), but there was no difference between the two learner groups (thus apparently no L1 transfer effect).

To investigate the reason for this lack of group differences, Hitz conducted two follow-up experiments investigating the acceptability of RPs in Turkish RCs and in Mandarin Chinese
RCs according to native speakers of these languages. The results show that a likely reason for a lack of group differences in the English tasks can be traced to faulty assumptions about the distribution of RPs in Turkish and Chinese. Specifically, these follow-up experiments revealed that RPs were more acceptable than expected in Turkish RCs, and less acceptable than expected in Chinese RCs. Thus the two languages were much more similar to each other than the prevailing linguistics literature about the RP distributions in both languages had indicated.

As to the reason for the difference between the control group (which consistently failed to accept RPs) and the two learner groups (which accepted RPs in non-subject positions to some extent), either a transfer account or a processing account is equally plausible given the similar results for the Turkish and Chinese tasks, and in fact both of these factors could be at work at the same time. The data from these experiments do not let us draw any firm conclusions on this matter.

To sum up, the apparently faulty transfer hypotheses formulated by Hitz (2012) in his investigations of the second-language acquisition of English RCs can be attributed to the trust he placed in the informal judgments of the distribution of RPs as reported in linguistic studies of Chinese and Turkish RCs. The results of Hitz’s (2012) experiments have both theoretical and methodological implications. For theories of formal syntax, the results point toward revised formulations of the anti-locality conditions proposed to account for the different RP distributions in Turkish (Kornfilt 2000) and Chinese (Hu and Liu 2007): the restrictions on allowable RP positions apparently need to be more permissive for Turkish but stricter for Mandarin Chinese. From our own functionalist perspective, these data lend support to Hawkins’ (2004) model of performance-grammar correspondence and, more specifically, the
idea that RP distributions across languages are at least partially a reflex of processing complexity as defined in terms of filler-gap domain size. Notably, both the similarities and differences that were found between RP distributions in Turkish and Chinese corresponded to filler-gap domain sizes for these languages as defined by Hawkins (2004). Finally, with respect to methodology in linguistics, it appears that using formal AJTs or other quantitative measures can reduce the possibility that SLA researchers will make mistaken claims for or against language transfer, or that theoretical linguists will use inaccurate linguistic descriptions to build theories of syntax.

In the following pages, we first of all discuss the use and non-use of experimental data as the basis for argumentation in linguistics, and then present one of the reasons why experimental evidence is necessary: to provide a clear pattern of data that can support a particular theory. We then examine the data from three AJTs, as previously summarized.

**1.1 The Use and Non-Use of Formal AJTs in Mainstream Linguistics**

Phillips (2009) claimed that there are no cases in which faulty informal acceptability judgments have become the basis for faulty linguistic theorizing. In response, Gibson and Federenko (2013) cite several previous studies and three additional cases in which apparently faulty informal judgments have formed the basis for theoretical claims in the literature. Of the three cases they discuss, the first involves multiply embedded relative clauses in English, while the second and third involve English interrogatives containing multiple wh-expressions. For all three cases, the authors provide experimental evidence showing that formally-collected acceptability judgments fail to replicate grammaticality contrasts presented in the literature. In
their reply to Gibson and Fedorenko (2013), Sprouse and Almeida (2013a) come to the defense of informal acceptability judgments. They argue that little can be concluded from the three case studies cited by Gibson and Fedorenko (2013) due to the small size and the non-random nature of the sample, and go on to cite a 95% convergence rate between the informal judgments of linguistic structures provided by researchers in randomly selected issues of *Linguistic Inquiry* (for 150 minimal pairs) and judgments of these same structures provided by experimental participants in a series of formal acceptability judgment tasks developed by Sprouse, Schütze and Almeida (2013a) (for 2400 items based on the structures in the original 150 minimal pairs). In short, of the total number of informal acceptability judgments investigated by Sprouse et al., only 5% of them were not in agreement with the results of formal AJTs of the same linguistic phenomena. These results lead Sprouse and Almeida (2013a) to conclude that neither formal nor informal acceptability judgment tasks are necessarily better than one another in any absolute sense, and that selecting the type of judgment task depends on the type of linguistic phenomena that are being investigated. For Sprouse and Almeida (2013a), formal acceptability judgments should be preferred in two areas: firstly, in testing hypotheses about the roles of syntactic constraints versus working memory constraints, and secondly, in exploring the mechanisms that underlie gradient acceptability judgments.

An issue that Sprouse and Almeida (2013a) do not address is how linguists can identify those gradient patterns of judgments that might shed light on the relative contributions of gradient and non-gradient mechanisms in sentence generation. If linguists primarily use informal acceptability judgments, they run the risk of misidentifying gradient patterns as categorical patterns. As Wasow (2007) observes, informal acceptability judgments as used in
research on generative grammar tend to allow only two options: “acceptable” or “unacceptable,” and may therefore potentially misconstrue a gradient phenomenon in need of a more complex explanation as a simpler categorical phenomenon. An examination of the English and Chinese-language AJTs to be discussed in Sections 2 and 3 reveal a gradient pattern of data that informal judgments of the same linguistic phenomena did not show, and therefore demonstrates that the use of formal acceptability judgments may shed light on linguistic phenomena in need of a more sophisticated explanation than can be provided by means of a categorical grammatical rule.

Even aside from the issue of gradience, the current study shows a different relative pattern of judgments than expected for the acceptability of RPs in both Turkish and Chinese. Given the relatively low error rate reported in Sprouse, Schütze and Almeida (2013), why should it be that two experiments on two different languages and populations of speakers yielded unexpected patterns of results? It could be that certain types of linguistic phenomena (in this case, RP distributions) may be more prone to variability across speakers or errors in informal judgments than others. Given that the nature of the phenomenon was what the two experiments had in common, this provides a plausible post-hoc explanation for the results. However, as Gibson et al. (2013: 233) point out in their reply to Sprouse and Almeida (2013a), there is no way of knowing in advance which grammaticality contrasts will be among those that cannot be replicated in a formal experiment. Thus, there is no reliable manner of determining whether to trust informal judgments in any particular case, except perhaps for obvious constraints on constituent order (e.g., *book a), for which informal judgments could almost certainly be trusted.
If informal judgments are not always reliable for showing contrasts in grammaticality, as Gibson et al. (2013) argue, then neither are they up to the task of theory building:

The experiments are necessary in all cases because they provide the only way to objectively measure discrepancies between theory and reality. Theories evaluated only by the intuitions of the investigators involved, are almost necessarily post hoc. This is because, lacking quantitative standards, we have little possibility to be wrong and discover discrepancies between theory and data. Such discrepancies drive scientific progress, and while expert intuitions provide a rich source of hypotheses to investigate, reliable evaluation of such hypotheses requires more sophisticated quantitative methods. (Gibson, Piantadosi and Federenko, 2013: 10)

The AJTs presented here support Gibson et al.’s (2013) position: they cast doubt on previous theoretical proposals for syntactic constraints on RPs in Turkish and Mandarin, while supporting the Filler-Gap Domain complexity ranking for relative clause types as proposed by Hawkins (2004). The problems for existing theoretical proposals will be elaborated within the experiment discussions in Sections 3-4. To highlight the positive contribution of this work in supporting Hawkins’ (2004) theory, we outline the major features of the theory in the following section.
1.2 Hawkins’ (2004) theory of RP use and relative clause complexity

Hawkins (2004) presents a theory of RP use and relative clause complexity which is embedded within his more general model of performance-grammar correspondence. The Performance-Grammar Correspondence Hypothesis (henceforth, PGCH), on which this general model is based, is as follows:

Grammars have conventionalized syntactic structures in proportion to their degree of preference in performance, as evidenced by patterns of selection in corpora and by ease of processing in psycholinguistic experiments. (Hawkins 2004: 3)

The hypothesis states that many syntactic constraints which disallow a particular structural configuration have their origins in constraints on language processing, and that we can find evidence for this in patterns of usage and in psycholinguistic experiments on languages that are not subject to these syntactic constraints but which instead allow more than one syntactic option. Applied to the case of RPs, we can observe that cross-linguistically, RPs tend to be grammatically required more often as RC complexity increases. This observation was first developed by Keenan and Comrie (1977) and formulated in terms of a Noun Phrase Accessibility Hierarchy that measured complexity in terms of the function of the relativized element, where subject RCs are simplest, followed by direct object, indirect object, oblique, and genitive RCs. In Keenan and Comrie’s (1977) sample of 26 languages that have grammatically licensed RPs, all require a gap in subject RCs, the majority require a gap in direct object RCs, and the majority require an RP in indirect object and oblique RCs. Likewise, in performance, we can observe that children acquiring their first language (McKee and McDaniel 2001) and adults
acquiring a second language (Tezel 1998) are more likely to produce or accept an “ungrammatical” RP (i.e., an RP that is used in a position deemed ungrammatical by adult native speakers of the target language) as RC complexity increases. Similarly, for languages that optionally allow either gap or RP in particular positions, such as Hebrew (Ariel 1999) and Cantonese (Francis et al. 2015), corpus and experimental data show that speakers are more likely to produce an RP as RC complexity increases. These three types of evidence (cross-linguistic patterns of syntactic constraints, language acquisition, and language use) support Hawkins’ proposal that RP use (whether grammatically required or not) is linked to RC complexity, which is in turn linked to processing complexity.

As a refinement of Keenan and Comrie’s Noun Phrase Accessiblity Hierarchy, Hawkins (1999, 2004) proposes a revised ranking of RC complexity based on the notion of Filler-Gap Domain, which is defined as follows:

An FGD [Filler-Gap-Domain] consists of the smallest set of terminal and non-terminal nodes dominated by the mother of a filler and on a connected path that must be accessed for gap identification and processing; for subcategorized gaps the path connects the filler to a co-indexed subcategorizor and includes, or is extended to include, any additional arguments of the subcategorizor on which the gap depends for its processing; for non-subcategorized gaps the path connects the filler to the head category that constructs the mother node containing the co-indexed gap; all constituency relations and co-occurrence requirements holding between these nodes belong in the description of the FGD.

(2004: 175)
The terms “filler,” “subcategorizer,” and “gap” are illustrated in the following example:

(3) [NP The girl [who, Sarah helped ____]]

In this example, the filler, girl, must be held in working memory until it can be associated or co-indexed with the appropriate gap after helped. The subcategorizer, helped, is the constituent that determines the RC’s argument structure. In this case, the subcategorizer requires an agent, Sarah, and a theme. The filler-gap domain consists of the constituents and nodes after the.

Following Hawkins’ basic style of tree diagram, Figure 1 illustrates the FGD in sentence (3):

---

2 Especially in studies of English, the wh-relative pronoun who is often assumed to be the filler. However, Hawkins (1999: 251) instead classifies the head noun as the filler in a relative clause, and calls the co-indexed relative pronoun a “filler copy.” The motivation for this is so that FGD can be defined cross-linguistically to include languages such as Mandarin and Turkish that lack an overt wh-pronoun in relative clauses, without having to necessarily assume wh-movement of null relative pronouns in these languages. We acknowledge that Hawkins’ terminology may be considered non-standard, but note that re-defining the filler as the (overt or covert) wh-pronoun would have no effect on the relative rankings of FGD size that Hawkins proposes, and thus would not affect our interpretation of the current results.
Figure 1. *The FGD in an English RC (Adapted from Hawkins, 1999, p. 250)*

The FGD in sentence (3) consists of 8 nodes: \( N_i \), the filler *girl*; \( NP_1 \), the mother of the filler; \( S \); the filler copy (who\(_i\)); \( NP_2 \); N (*Sarah*); VP; and V, the subcategorizer *helped*. The FGD in an English subject RC consists of 6 nodes (since the VP is not included), and the FGD of an RC with a relativized object of a preposition consists of 10 nodes (since PP and P are added).\(^3\) Thus, for English, Hawkins’ FGD complexity ranking aligns with Keenan and Comrie’s Noun Phrase Accessibility Hierarchy: subject < direct object < oblique object.

\(^3\) An anonymous reviewer notes that Hawkins’ metric of complexity does not take the degree to which the RC is embedded in a matrix clause into account, and cites Wasow’s (2002) claim that node-counting is not the only viable measure of complexity. We agree that Hawkins’ complexity metric has these limitations. However, since our experiments manipulated RC complexity only on this basis, and obtained results consistent with the proposed complexity ranking, we believe these other factors (such as depth of embedding of the NP within the matrix clause, which was held constant in the experiments) are not pertinent to understanding the current data.
As detailed in the following sections, the results of all three experiments are consistent with the predictions of Hawkins’ (2004: 178-180) complexity ranking, and as such, provide support for the PGCH. For English, the two non-native speaker groups rated direct object and prepositional object RCs containing an RP slightly (but significantly) higher than subject RCs containing an RP. This can be interpreted as a performance-based complexity effect, although (as we will see) an account based on L1 transfer cannot be ruled out. For Mandarin Chinese, native speakers rated subject and direct object RCs with an RP as unacceptable, while rating indirect object RCs with an RP as marginally acceptable. Although the similarity between subject and object RCs was unexpected based on previous studies that relied on informal judgments, this similarity is predicted by Hawkins’ complexity ranking for Chinese (which, to be explained below, differs from the complexity rankings for English and Turkish in that subject RCs and object RCs have equal FGDs). Finally, for Turkish, native speakers rated direct object and indirect object/oblique RCs with an RP significantly higher than subject RCs with an RP. Again, the results were contrary to studies of Turkish syntax which have claimed that RPs are unacceptable in all three positions, but compatible with Hawkins’ complexity ranking.

In our discussion of each experiment that follows, we will show how Hawkins’ FGD complexity ranking can help explain the pattern of data we have obtained, and we will argue that informal sentence judgments, as used in most of the previous studies of the RP distributions in Mandarin and Turkish, could not provide a convincing basis to support or refute a particular theoretical position.
2. Acceptability judgment task 1: English

In order to illustrate our point regarding the need for quantitative data in linguistic argumentation, and to provide support for the Hawkins’ FGD complexity ranking, we turn to Hitz’s study of language transfer in second language acquisition (SLA). SLA scholars have hotly debated the importance of language transfer, and have not been able to reach a consensus about its importance (Odlin 2003). Because further investigation seemed justified, an English-language acceptability judgment task was designed and administered to native speakers of English, Mandarin Chinese, and Turkish.

2.1 Hypotheses

The majority of published studies appear to agree that RPs are highly acceptable in some Chinese RCs, whereas they are for the most part banned from English and Turkish RCs.\(^4\) That being the case, we postulated that L1 influence would cause L1 Mandarin Chinese speakers to find RPs in English RCs with relativized direct objects and objects of prepositions more acceptable than L1 Turkish speakers at statistically significant levels. This hypothesis was motivated by SLA studies performed by Gass (1979) and Hyltenstam (1984), which claim that L2 learners are likely to accept resumptive pronouns in the L2 if they are grammaticalized in the L1. Furthermore, we postulated that both L2 groups would find RPs in English RCs most

\(^4\) As stated earlier, Keenan and Comrie (1977), Tezel (1998), Kornfilt (2000), and Çağrı (2005) claim that RPs are not acceptable in Turkish RCs with relativized subjects, direct objects, indirect objects, and oblique objects, while scholars of Chinese claim that RPs unacceptable in subject RCs, optional in Chinese direct object RCs, and mandatory in indirect object RCs and in RCs with relativized objects of prepositions (Keenan and Comrie, 1977; Xu and Langendoen, 1985; Hawkins and Chan, 1997; Hsiao, 2003; Hu and Liu, 2007).
acceptable in prepositional object RCs, somewhat less acceptable in direct object RCs, and least acceptable in subject RCs, due to the increasing levels of processing complexity in these three RC types, as operationalized in terms of Hawkins’ (2004) idea of filler-gap domains as discussed previously in Section 1.2. In contrast, native English speakers should find RPs categorically unacceptable across all RC types.\(^5\)

### 2.2 Methodology

#### 2.2.1 Design

An acceptability judgment task was constructed following principles described by Cowart (1997). The experimental items in the task consisted of eight token sets representing six experimental conditions each for a total of 48 experimental items. A sample token set is given in Table 1.

In a token set, the experimental items have the same vocabulary, and differ from one another only in terms of the factors manipulated by the experimenter. As shown in the sample token set in Table 1, two factors were manipulated: Pronoun (presence/absence of a RP), a factor with two levels, and RC Type, a factor with three levels (Subject, Direct Object, Prepositional Object) that differ from one another in their position on the FGD complexity ranking, as discussed in Section 1.2 above.

---

\(^5\) Given that Hofmeister and Norcliffe’s (2013) study of RP processing by native English speakers found no processing advantage for RPs in simple direct object RCs, we believe that the sentence materials in the current study are simple enough that native speakers will show no measurable complexity effects.
Hawkins (1999, 2004) argues that resumptive pronouns aid in the processing of RCs, and tend to increase in acceptability across languages with the size of the filler-gap domain: “The presence of this resumptive or copy pronoun can be argued to make the processing of head and relative clause easier, since an empty category no longer needs to be inferred from its environment but is given formal expression” (257). We therefore expected that the non-native English speakers would be more likely to accept RPs in the direct and prepositional object conditions than in the simpler subject RCs.

Table 1 illustrates that the experimental items had relativized subjects, direct objects, and objects of prepositions, RC types whose FGDs increase in length respectively—6 in a gapped subject RC, 8, in a gapped direct object RC, and 10 in gapped RC with a relativized object of a preposition.

Table 1. Sample Token Set for Acceptability Judgment Task 1: English

<table>
<thead>
<tr>
<th>Subject RC with Resumptive Pronoun</th>
<th>The relatives, [who, they, visited us last night] enjoyed the evening.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject RC with Gap</td>
<td>The relatives, [who, ____ , visited us last night] enjoyed the evening.</td>
</tr>
<tr>
<td>Direct Object RC with Resumptive Pronoun</td>
<td>The relatives, [who, we visited them, last night] enjoyed the evening.</td>
</tr>
<tr>
<td>Direct Object RC with Gap</td>
<td>The relatives, [who, we visited ____ , last night] enjoyed the evening.</td>
</tr>
</tbody>
</table>
Prepositional Object RC with Resumptive Pronoun

The relatives, [who we paid a visit to them,] last night enjoyed the evening.

Prepositional Object RC with Gap

The relatives, [who we paid a visit to ___,] last night enjoyed the evening.

For the third relative clause type shown in Table 1, items with stranded prepositions were used: The relatives who we paid a visit to last night enjoyed the evening. Alternatively, fronted prepositions could have been used, such as The relatives to whom we paid a visit last night enjoyed the evening. RCs with stranded prepositions were chosen in light of Bardovi-Harlig's (1987) finding that L2 English learners acquired stranded prepositions before they acquired pied-piped prepositions, even in EFL contexts, mainly because stranded prepositions occur more frequently in the L2 input with which L2 English learners are surrounded.

Fifty-six filler sentences of varying length and acceptability were combined with the 48 experimental sentences for a grand total of 104 items. All items were randomized according to procedures described in Cowart (1997) and organized in the form of a written questionnaire which was distributed to participants in two versions in which items were ordered differently from one another. Participants were asked to rate items on a 4 point scale in which 1=certainly incorrect, 2=possibly incorrect, 3=possibly correct, and 4=certainly correct.

2.2.2. Participants

Thirty-three native speakers of Turkish, 35 native speakers of Mandarin Chinese, and 34 native speakers of English were recruited from the student body at Purdue University in West Lafayette, Indiana, USA, to participate in the study. They were majoring in a variety of academic
subjects, and none had coursework or training in linguistics. Demographic data concerning the participants appear in Table 2.

Table 2. Participants’ Demographic Data for Acceptability Judgment Task 1: English

<table>
<thead>
<tr>
<th>Number of Participants</th>
<th>Mean Age</th>
<th>Sex</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1 Mandarin Chinese</td>
<td>35</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11 Males, 24 Females</td>
</tr>
<tr>
<td>L1 Turkish</td>
<td>33</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td></td>
<td>17 Males, 16 Females</td>
</tr>
<tr>
<td>Native Speakers of English</td>
<td>34</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td></td>
<td>17 Males, 17 Females</td>
</tr>
</tbody>
</table>

2.2.3. Experimental procedures

At the beginning of each testing session, biodata forms were distributed. Non-native speakers completed the Michigan Placement Test in approximately 30 minutes, followed by a short break of 10-15 minutes. All participants completed a written questionnaire containing the English-language sentence judgment task. The instructions, derived from Cowart (1997: 57), were read aloud, after which participants answered four practice items with the facilitation of the experimenter. The non-native speakers completed the task in approximately 45 minutes.

---

6 The Reading and Grammar sections of the Michigan Placement Test were administered in order to ensure that both groups of L2 English participants had similar levels of English language proficiency, a procedure designed to ensure that L2 language proficiency was not an experimental confound.
to 1 hour, while the native speakers took about 30 minutes to complete the same task. All
participants were paid 10 dollars for each hour of their participation.

2.2.4 Statistical tests used to analyze the data

This study employs a 3-way factorial research design, with Language (L1 Background) as
a between-subjects factor. RC Type and Pronoun (Presence/Absence of RP) were within-
subjects factors.

A mixed-model analysis of variance was used to analyze the patterns of acceptability
judgments. This procedure is a parametric statistical test that can ideally state that the
variation between the different L1 groups' acceptability judgments is caused by an interaction
of three different fixed factors: L1 background, RC type, and the presence/absence of RPs. In
this respect, it is similar to standard ANOVA procedures. The test differs from a standard
ANOVA in that it has one random factor (research participants) in addition to the fixed factors⁷,
and in its output, which does not include an error term because it is already factored into the
statistical test itself (SAS Institute Inc., 1999: 2175). Despite the fact that this test does not
include an error term in its output, the survey results have been adjusted so that the effects of
sample sizes can be accounted for. Least Squares Means tests with Tukey adjustments⁸ were
applied to the data to identify which particular differences between L1 groups' acceptability
judgments caused main effects and interactions.

--------------------
⁷ A random factor is an experimental variable that the researcher has not controlled, whereas
fixed factors have been controlled (Littell, Stroup, and Freund, 2002).

⁸ Tukey adjustments are necessary to control for the possibility of Type 1 error caused by
multiple tests of statistical significance (Leavitt, 2002).
2.3 Results and discussion

The results of the Michigan English Placement test administered to L1 Turkish and Chinese speakers show almost identical results. Out of a total of 80 multiple-choice questions, the L1 Chinese speakers answered 91% correctly, and the L1 Turkish speakers answered 92% correctly. A Mann-Whitney test showed no statistically significant differences between the proficiency test scores ($z=-.542, p=.410$).

None of the transfer-related hypotheses were supported; there were no statistically significant differences between native speakers of English, Turkish, and Chinese in their judgments of RCs with relativized direct objects and objects of prepositions in the resumptive condition. Figure 2 shows that there were few outstanding differences between the native- and non-native speakers on AJT 1.

---

9 A more detailed analysis of the statistics that support this conclusion can be found in Appendix A, Table 2.
The error bars represent the standard error of the mean. Exact values and standard deviations associated with each data point represented in Figure 2 can be seen in Appendix A, Table 1.
The data from AJT 1 demonstrate another reason to perform formal sentence judgment tasks: their ability to depict gradient patterns of data. Although AJT 1 did not show a predicted transfer effect, it did display a gradient pattern of data in that the non-native speakers of English found non-subject RCs in the RP condition to be more acceptable than the subject RCs. Statistical comparisons within L1 groups showed that L1 Turkish and L1 Chinese speakers both found direct object RCs with RPs to be more acceptable than subject RCs with RPs. For L1 Chinese speakers, this difference approached significance (t=3.410, p=.069), while for L1 Turkish speakers this difference was highly significant (t=4.190, p=.004). Similarly, within-groups comparisons showed that the L1 Chinese speakers found RCs with prepositional RCs in the RP condition to be more acceptable than subject RCs with RPs at a statistically significant level (t= 3.640, p=.033). For Turkish speakers, on the other hand, the apparent difference between prepositional RCs with RPs and subject RCs with RPs was not significant (t= 2.74, p=.355). To the extent that RPs were more acceptable in non-subject positions as compared with subject positions, these findings for the two L2 groups are consistent with Hawkins’ FGD complexity ranking and may reflect a slightly greater processing cost for comprehending non-subject RCs.

In contrast to the non-native speakers, within-groups comparisons of native English speakers’ judgments of items with RPs showed no statistically significant differences: the comparison of ratings assigned to subject RCs in the resumptive condition with ratings assigned to direct objects with RPs was not statistically significant (t= -.120, p= 1.000), nor was the comparison between ratings assigned to subject RCs with RPs and RCs with relativized objects of prepositions with RPs (t= .480, p= 1.000). The consistently low ratings given to RCs with RPs
by native adult English speakers are congruent with informal acceptability judgments of these structures, and with experimental evidence obtained by McKee and McDaniel (2001).

As stated earlier, we reasoned that the absence of any transfer effect might have been caused by incorrect assumptions about the distributions of RPs in Chinese and Turkish RCs. In order to explore this possibility, we designed AJTs in these languages.

3. Acceptability judgment task 2: Mandarin Chinese

3.1 Hypotheses

In this study, our hypotheses were based on the "consensus opinion" mentioned in the introduction, which states that Mandarin Chinese does not allow RPs in RCs with relativized subjects, that they are optional and alternate with gaps in RCs with relativized direct objects, and that they are mandatory in RCs with relativized indirect objects and objects of prepositions (Keenan and Comrie 197; Xu and Langendoen 1985; Hawkins and Chan, 1997; Hsiao, 2003; Hu and Liu, 2007). Therefore, we expected sentences with RPs to receive low acceptability ratings in subject RCs, but significantly higher ratings in direct object and indirect object RCs. With respect to gapped clauses, we expected subject RCs and direct object RCs to receive high acceptability ratings, with significantly lower ratings given to indirect object RCs.

3.2 Methodology

3.2.1 Design

As illustrated in (4), Chinese has an unusual word order typology in which SVO clause order is combined with a head-final NP structure. The relative clause (Paul zhaogu de) precedes its head noun (nuren), and thus the gap (in this case, an object gap) precedes its filler.
Due to this unusual word order, Hawkins (2004: 180) proposes that subject RCs and object RCs do not differ in terms of FGD in a transitive clause. Thus, the RC types represented in our Chinese stimuli showed a slightly different pattern of FGDs than those in the English stimuli. Because the direct object intervenes between the subcategorizing verb and the head noun, the FGD for a subject RC in a transitive sentence must include the direct object (unlike in English, where the direct object does not intervene between the head noun and the verb). Thus, for a transitive clause structure, both subject and object RCs have a FGD of 8.\textsuperscript{11} Indirect object RCs (which are comparable to the prepositional object RCs in our English stimuli) must include both the subject and the direct object within the FGD, increasing the FGD to 10 in our stimuli. A tree diagram in the style of Hawkins (2004: 178) illustrates the RC from sentence (4) in Figure 3.

\textsuperscript{11} An anonymous reviewer suggests that the node counts on which Hawkins’ complexity metric is based may not be useful for comparing languages with head-initial vs. head-final NPs. However, Hawkins’ metric was designed precisely to be able to compare across different language types and to explain generalizations (such as cross-linguistic patterns of RP distribution) that occur in languages with different word order typologies. As far as we are aware, Hawkins’ theory is the only one that states that Mandarin Chinese subject and direct object RCs should be equal to one another in RC complexity due to the mixed word-order typology of Chinese, a claim consistent with the results we have obtained.
The experimental items in AJT 2 consisted of 24 sentences containing Mandarin Chinese RCs with relativized subjects, direct objects, and indirect objects. The design was almost identical to that of the English AJT, except that there was only one participant group: native speakers of Mandarin. As in the English AJT, two within-group factors were manipulated: RC Type (subject, direct object, and indirect object), and Pronoun (presence/absence of RP).

Experimental items consisted of four token sets, which were developed with the assistance of a native speaker of Mandarin Chinese and an L2 speaker of Mandarin Chinese, both of whom had training in linguistics. A sample token set illustrating all six conditions can be seen in Table 3.  

---

12 An anonymous reviewer states that 24 item sets should have been created, and that these items should have been distributed in 6 lists under a Latin Square design. We agree that the limited number of token sets limits the generalizability of the results. It will be shown, however, that the results harmonize with recent experimental research performed by Francis et al. (2015) and with data obtained by Ning, Christianson, and Lin (2014).
Table 3. *Sample Token Set for Acceptability Judgment Task 2: Mandarin Chinese*

<table>
<thead>
<tr>
<th>RC Type</th>
<th>Token Set</th>
<th>English Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subject RC with Resumptive Pronoun</strong></td>
<td>( [\text{st}a] \text{zhaogu Paul de nage nuren_i likai le} )</td>
<td>&quot;The woman that she looked after Paul has left.&quot;</td>
</tr>
<tr>
<td><strong>Subject RC with Gap</strong></td>
<td>( [_i \text{VP zhaogu Paul de nage nuren_i likai le} )</td>
<td>&quot;The woman that looked after Paul has left.&quot;</td>
</tr>
<tr>
<td><strong>Direct Object RC with Resumptive Pronoun</strong></td>
<td>( [s \text{Paul [VP zhaogu ta_i] de nage nuren_i likai le} )</td>
<td>&quot;The woman that Paul looked after her has left.&quot;</td>
</tr>
<tr>
<td><strong>Direct Object RC with Gap</strong></td>
<td>( [s \text{Paul [VP zhaogu _i] de nage nuren_i likai le} )</td>
<td>&quot;The woman that Paul looked after has left.&quot;</td>
</tr>
<tr>
<td><strong>Indirect Object RC with Resumptive Pronoun</strong></td>
<td>( [s \text{Paul [VP gei ta_i zhaogu] de nage nuren_i likai le} )</td>
<td>&quot;The woman that Paul showed concern to her has left.&quot;</td>
</tr>
<tr>
<td><strong>Indirect Object RC with Gap</strong></td>
<td>( [s \text{Paul [VP gei _i zhaogu] de nage nuren_i likai le} )</td>
<td>&quot;The woman that Paul showed concern to has left.&quot;</td>
</tr>
</tbody>
</table>
The experimental tokens were supplemented by 24 grammatical and ungrammatical fillers, which were translated from English by a native speaker of Mandarin Chinese. The entire instrument had 48 items altogether, which were randomized according to procedures discussed in Cowart (1997) and then distributed to participants in two separate versions of the questionnaire with different sequences of items. As in AJT 1, participants rated the items on a 4 point scale (1=certainly incorrect, 2=possibly incorrect, 3=possibly correct, and 4=certainly correct).

After the results of the acceptability judgment task were reviewed and analyzed, it was found that one of the token sets had flawed items in the indirect object condition. As a result, all six items from the flawed token set were excluded from the pool of analyzable data. Thus, three token sets that contained a total of 18 experimental items were included in the pool of analyzable data. 13

3.2.2 Participants

All 32 participants were native speakers of Mandarin Chinese and were studying at Purdue University in a variety of academic majors. Of the 32 participants, 16 were males, 16 were females, and 7 had previously participated as non-native speakers in the English AJT. The majority of participants (30) were citizens of the People's Republic of China; the remaining participants (2) were citizens of Taiwan. 14 The participants' average age was 22.2 years old.

13 For a more detailed explanation of the rationale for excluding these items, see Hitz (2012), 177-179.
14 Although L2 speakers of Mandarin are very common in the PRC, such participants were excluded based on the recruitment materials and biodata forms.
3.2.3 Procedures

Firstly, the experimenter distributed the participant biodata form. After these forms were collected, paper copies of the judgment task were distributed, and participants completed the AJT on paper. The directions for completing the questionnaire were written in English and Mandarin Chinese. The experimenter read the directions in English because he did not know Mandarin, and asked if there were any questions about them. He explained that the same directions were written in Mandarin on the questionnaire. Then, he asked the participants to read and evaluate 5 practice sentences. Most participants completed the questionnaire in approximately 20 minutes, and each volunteer received 10 dollars for participating.

3.2.4 Tests of statistical significance

This study used a 2-way factorial research design, with subjects repeated on the RC Type and Pronoun variables. As in the English-language AJT, the statistical procedure employed a mixed model, this time with two fixed factors, RC Type and Pronoun, and one random factor, participants. The RC Type variable has three levels: subject, direct object, and indirect object, and the Pronoun variable has two levels: presence/absence of RP. After the mixed-model procedure determined the variables that qualified as main effects and interactions, Tukey-Adjusted Least Squares Means Tests (hereafter, TALMS tests) were used to determine the differences between item types that caused the main effects and interactions.\(^\text{15}\)

\(^{15}\) Main effects and interactions of various factors are presented in Appendix B, Table 2.
3.3 Results and discussion

There was a highly significant main effect of Pronoun (F (1,31), F= 332.370, p < .001) due to the overall greater acceptability of gapped clauses as compared with resumptives. The lack of any main effect of RC Type (F (2,62), F= 2.160, p=.124) results from the highly significant interaction between Pronoun and RC Type (F (2,62), F= 54.830, p < .001), whereby acceptability ratings for gapped clauses vs. resumptives went in opposite directions across the three types of RCs. This interaction is shown in Figure 4.

![Figure 4. Mean Ratings of Mandarin Chinese RCs](image)

In figure 4, the error bars indicate the standard error of the mean. Exact values for each data point, as well as the standard deviations, can be seen in Appendix B, Table 1.
The lower line in Figure 4 shows that RPs gain in acceptability as the level of embedding increases across RC types. In contrast, the acceptability of gapped RCs decreases in tandem with the level of embedding, as indicated by the upper line. TALMS tests verify that these trends were statistically significant for some comparisons. RPs were more acceptable in indirect object RCs than in subject RCs at a statistically significant level (t= 5.980, p < .001), and were also more acceptable in indirect object RCs than in direct object RCs (t= -4.460, p < .001). However, gapped RCs decreased in acceptability as their syntactic complexity increased. Gapped indirect object RCs were less acceptable than gapped direct object RCs at a statistically significant level (t= 7.380, p < .001), and were also less acceptable than gapped subject RCs (t= -7.640, p < .001).

These results run contrary to our original hypotheses based on the consensus opinion in Chinese linguistics. We had predicted that RPs would be highly acceptable in direct and indirect object RCs, and unacceptable only in subject RCs. Although our predictions with respect to subject RCs were borne out, our results showed that RPs were in fact unacceptable in direct object RCs (average rating: 1.635) and were only of borderline acceptability in indirect object RCs (average rating: 2.364). With respect to gapped clauses, our predictions were correct for subject RCs and direct object RCs: both were highly acceptable, with ratings of 3.948 and 3.906, respectively. These results therefore cast doubt on theoretical proposals based on the consensus opinion in Chinese linguistics. For example, Hu and Liu (2007: 269) propose that RPs are disallowed in subject position but are optional in direct object position, with the latter optionality explained in terms of an alternation between null and overt RPs. However, if (overt) RPs are judged as equally unacceptable in subject and direct object position, as the current
study has shown, the most straightforward syntactic analysis would be to treat direct object relatives as similar to subject relatives but different from indirect object relatives (which do permit overt RPs).

Also contrary to our original predictions, indirect object RCs with gaps (average rating: 2.697), were *more* acceptable than indirect object RCs with RPs (average rating: 2.365). Although these did not differ statistically (t= -2.040, p=.334), we had predicted a difference in the opposite direction. The judgments of the indirect object RCs are gradient in that they were neither completely acceptable, nor completely unacceptable to the experimental participants. This finding contrasts with the majority of published and presumably informal sentence judgments of Chinese indirect object RCs, which state categorically that gapped indirect objects are unacceptable, and that indirect object RCs with resumptives are acceptable (Keenan and Comrie 1977: 93; Hawkins and Chan 1997: 192; Hu and Liu 2007: 269). Thus, our data illustrate that gradient effects cannot as easily be uncovered by informal sentence judgments or be predicted in advance, and illustrate the need for quantitative data that might depict or at least control for any possible gradient effects.

Although our results do not verify the consensus view on which our hypotheses were based, they are consistent with Hawkins’ (2004) prediction that RPs and gaps in any language should be distributed in accordance with hierarchies of RC complexity which predict a greater probability of grammaticalized RPs in more complex RCs, and a greater probability of grammaticalized gaps in simpler RCs (2004: 186). This prediction is borne out in the differences found between subject and direct object RCs (which preferred gaps) and indirect object RCs (for
which there was no clear preference). Furthermore, Hawkins (2004: 180) makes the specific observation that in languages like Chinese, with a SVO clause order and head-final NP, subject RCs and direct object RCs do not differ terms of FGD size. As in other language types, the subject NP is included within the FGD for a direct object relative. Unlike other language types (and as noted above), the direct object in a subject relative intervenes between the head noun and the subcategorizing verb, thus requiring the direct object NP to be included within the FGD of a subject relative and making the FGD of a subject relative just as long as for an object relative (whereas for other language types, subject relatives would have a shorter FGD than object relatives). Because there are other factors besides FGD size which can affect a grammar, Hawkins’ theory does not make specific predictions regarding which structures will be found acceptable in a given language. However, our finding that there was a strong and equal preference for gaps over RPs in both subject and direct object RCs is consistent with Hawkins’ observations regarding the equivalent FGD size of subject and object relatives in languages like Chinese.

Our results, as it turned out, were also largely consistent with the experimental results of Yuan and Zhao (2005), Su (2004), and Ning (2008), Ning et al. (2014), and Francis et al. (2015).¹⁷ Yuan and Zhao (2005) designed an acceptability judgment task in Mandarin Chinese in order to determine the distribution of gaps and RPs in Mandarin Chinese, as a control condition in a study on L2 acquisition of Mandarin. The results for their native speaker group

¹⁷ Some of these studies may be rather obscure to researchers of Chinese linguistics because they are primarily focused on language acquisition (Yuan and Zhao 2005; Su 2004) or sentence processing (Ning 2008) rather than grammar or syntax. In addition, Ning’s (2008) study is an unpublished master’s thesis, and Su’s (2004) study was from conference proceedings.
show that RPs were rated as significantly less acceptable than gaps for both subject RCs and direct object RCs, while subject and direct object RCs did not differ from each other. Further, their results showed that that RPs and gaps were equally acceptable in RCs with relativized indirect objects, although participants generally rated indirect object RCs lower than gapped subject and object RCs. Similarly, in an acceptability judgment task, Ning (2008) found that her participants consistently rated subject RCs and object RCs alike: gapped clauses were consistently preferred over clauses with RPs to the same degree for both subject and object RCs. Likewise, Su’s (2004) production data are in line with that of Yuan and Zhao (2005) and Ning (2008); in an elicitation task administered to 31 adult native speakers of Mandarin (as a control condition in a study of L1 acquisition of Mandarin), RCs with resumptives were never produced in subject and direct object contexts. (Indirect object RCs were not tested.)

More recent studies of Mandarin RCs by Ning et al. (2014) and of Cantonese RCs by Francis et al. (2015) support the claim that RPs aid in sentence processing at greater levels of complexity, but not the claim that subject and direct object RCs are equal to one another in complexity. Ning et al. (2014) performed an eye-tracking study on 16 participants who read sentences containing subject, direct object, and indirect object RCs of varying lengths in RP and gapped conditions. The researchers found that RPs did not facilitate language processing in subject RCs, or in short direct object RCs, but that they did facilitate language processing in direct object RCs which were lengthened by adding an adverbial before the head noun. For indirect object RCs, there was no difference between the RP and gapped conditions. Their overall conclusion is similar to ours: that RPs in gap-filler constructions can facilitate processing of more complex dependencies. Francis et al. (2015) asked 22 native speakers of Cantonese to
complete AJTs and production tasks, and found that RPs were more acceptable and were
produced more often as the structural complexity of the RCs in these tasks increased, a finding
Contrary to the results of AJT 2, however, they found that RPs were more acceptable and were
produced more often in direct object RCs as compared with subject RCs.

The data from Francis et al. (2015) and Ning et al. (2014), while compelling, do not
disprove Hawkins’ claim that Mandarin subject and direct object RCs are equal to one another
in one measure of complexity. Ning et al. (2014) did not measure acceptability or production,
but instead used eye-tracking to assess subtle effects in reading comprehension. It is likely that
their eye-tracking measures were sensitive to some aspect of online processing that
acceptability judgements (Ning 2008; Yuan and Zhao 2005), reading time (Ning 2008), and
production measures (Su 2004) were not. Furthermore, the subject-object asymmetry that they
found was only present in one condition-- the long-RC condition. The results for Cantonese
from Francis et al. (2015) present a more serious challenge to Hawkins’ proposal. Cantonese is
closely related to Mandarin and shares the same word order typology. Therefore, Hawkins
(2004) predicts no subject-object asymmetry in RP use. However, Hawkins himself
acknowledges that other factors besides FGD size may affect processing complexity, and that
even the same complexity factors may be grammaticalized differently in different languages.
Francis et al. (2015: 76) note that the subject-object asymmetry that they found is consistent
with a frequency-based effect, since subject RCs are more common than object RCs in corpora
of Mandarin (Hsiao and MacDonald 2013). We can therefore speculate that different
dimensions of RC complexity may affect RP distributions in Mandarin vs. Cantonese.
In summary, the experimental findings of three studies of Mandarin RPs (Ning 2008; Su 2004; Yuan and Zhao 2005), which deviated from the scholarly consensus in Chinese linguistics and which did not figure in Hitz’s (2012) original hypotheses, turn out to align closely with the findings reported above in experiment 2. Notably, all of these studies relied on experimental data, whereas the scholarly consensus in Chinese linguistics is mostly based on studies that used informally collected judgments. The current results, in combination with those of other experimental studies, cast doubt on previous syntactic analyses (e.g. Hu and Liu 2007) while supporting the predictions of Hawkins’ (2004) FGD complexity ranking.

4. Acceptability judgment task 3: Turkish

4.1 Hypotheses

AJT 3 was designed in order to determine the distribution of RPs in Turkish RCs. The large body of literature on Turkish RCs has generally argued that they do not allow RPs in RCs with relativized subjects, direct objects, indirect objects, and oblique objects (Keenan and Comrie, 1977, Tezel 1998, Lewis, 2000, Kornfilt, 2000, Çağrı, 2005). Moreover, Kornfilt (2000) has argued that a syntactic constraint, the A’-Disjointness requirement,\(^{18}\) prevents the occurrence of RPs in Turkish RCs. This view was the basis for the experimental hypotheses in the current study: we predicted that gapped clauses would be highly acceptable across all

---

\(^{18}\) For more information on this constraint, see McCloskey (1990), Ouhalla (1993), and Kornfilt (2000).
conditions, while clauses with RPs would be unacceptable across all conditions, similar to the predictions for English in AJT 1.

The consensus view described above has been challenged. Meral (2004) argues that resumptive pronouns may alternate "freely" with gaps in Turkish RCs with relativized subjects, direct objects, indirect objects, and oblique objects, and are mandatory in some RCs with relativized objects of postpositions. Meral (2004) states that resumptive pronouns are acceptable in all Turkish RC types. Differing slightly from Meral, Göksel and Kerslake (2004) state that RPs are acceptable in all non-subject RCs. The results from Experiment 3 support Göksel and Kerslake (2004).

4.2 Methodology

4.2.1 Design

Similar to Chinese, Turkish NPs are head-final and therefore relative clauses come before the nouns they modify and fillers come before gaps. Unlike Chinese, however, the basic clause order is SOV. Turkish relative clauses take two forms. In the first type, the subject is relativized and an –en suffix is added to the verb; in the second, the direct, indirect, or oblique object is relativized and the -dik suffix is added to the verb. These RC types are illustrated in (5) and (6) below. Turkish is a pro-drop language, so overt subjects appear in RCs in which the object is relativized only for pragmatic reasons (Cağri, 2005).

(5) ben-i gör-en kişi-ler

I-ACC see-AN person-PL

"The people who saw me"
The structure for Turkish closely resembles that for Chinese, as shown in the Hawkins-style tree diagram below (Figure 5). However, because Turkish is consistently head-final in both the NP and the VP (with the subcategorizing verb adjacent to the head noun), FGDs for subject RCs are shorter than for object RCs (Hawkins 2004: 180). In this respect, Turkish resembles English but differs from Chinese. In a simple transitive clause, subject RCs have an FGD of 6 while direct object RCs have an FGD of 8. Indirect object RCs in a ditransitive clause have an FGD of 10, and so do oblique object RCs. The direct object RC from sentence (6) is illustrated in the following figure.
In experiment 3, one factor, RC Type, had three levels: subject RC, direct object RC, and RC with a relativized indirect object/relativized oblique object, while the other factor, Pronoun (Presence/Absence of RP), had two levels. A sample token set is given in Table 4.

---

19 Comrie (1989), one of the developers of the NPAH, puts indirect object RCs and RCs with relativized oblique objects at the same level of his NPAH taxonomy. Similar to Comrie, Hawkins (1999, p. 253, footnote 8) states that the filler-gap domains in indirect object RCs and oblique object RCs are identical to one another. Following the precedents set by Comrie (1989) and Hawkins (2004), we placed indirect object RCs and oblique object RCs in one category for analysis. In the Turkish AJT, the oblique object RCs contained relativized nouns in ablative case, indicated by the morpheme –den.
Table 4. Sample Token Set for Acceptability Judgment Task 3: Turkish

<table>
<thead>
<tr>
<th>Structure</th>
<th>Turkish Sentence</th>
<th>English Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subject RC with Resumptive Pronoun</strong></td>
<td>[5 Kendisi-ni, beyi, öv-en] adam, çok sevilen bir kişi-dir</td>
<td><strong>The man</strong> who <strong>he</strong> praises us is a much-loved individual.**</td>
</tr>
<tr>
<td></td>
<td>he-NOM, 1.PL-ACC, praise-SR²⁰ man very loved one person-PRED</td>
<td></td>
</tr>
<tr>
<td><strong>Subject RC with Gap</strong></td>
<td>[5 ___i, beyi, öv-en] adam, çok sevilen bir kişi-dir</td>
<td><strong>The man</strong> who praises us is a much-loved individual.**</td>
</tr>
<tr>
<td></td>
<td>1.PL-ACC, praise-NSR man very loved one person-PRED</td>
<td></td>
</tr>
<tr>
<td><strong>Direct Object RC with Resumptive Pronoun</strong></td>
<td>[5 pro Kendisi-ni, öv- düğ- ümüz] adam, çok sevilen bir kişi-dir</td>
<td><strong>The man</strong> who we praised <strong>him</strong> is a much-loved individual.**</td>
</tr>
<tr>
<td></td>
<td>pro he-ACC, praise-NSR²¹-1.PL man very loved one person-PRED</td>
<td></td>
</tr>
<tr>
<td><strong>Direct Object RC with Gap</strong></td>
<td>[5 pro ___i, öv- düğ- ümüz] adam, çok sevilen bir kişi-dir</td>
<td><strong>The man</strong> who we praised is a much-loved individual.**</td>
</tr>
<tr>
<td></td>
<td>pro praise-NSR-1.PL man very loved one person-PRED</td>
<td></td>
</tr>
</tbody>
</table>

²⁰ Turkish subject RCs use a special bound morpheme, -an, to indicate the presence of a subject RC. Because Turkish employs vowel harmony, the vowel in the morpheme changes to harmonize with vowels in preceding syllables. This morpheme is glossed as SR (subject relative clause).

²¹ The morpheme -dák is used to signal the presence of a non-subject RC. Due to vowel harmony, the vowel in the morpheme may change. This morpheme is glossed as NSR (non-subject relative clause).
Indirect Object RC/Oblique Object RC with Resumptive Pronoun

\[ S \text{ pro } \text{ Kendisi-nei } \text{ iltifat } \text{ et-tiğ-imiz } \text{ adamı, çok sevilen bir kişi-dir} \]

pro he-DAT compliment do-NSR-1.PL man very loved one person-PRED

“The man who we gave compliments to him is a much-loved individual.”

Indirect Object RC/Oblique Object RC with Gap

\[ S \text{ pro } ____ \text{ iltifat } \text{ et-tiğ-imiz } \text{ adamı, çok sevilen bir kişi-dir} \]

pro ____ compliment do-NSR-1.PL man very loved one person-PRED

“The man who we gave compliments to is a much-loved individual.”

The instrument contained 4 token sets consisting of 24 experimental items in addition to 24 fillers for a total of 48 items altogether. The items were randomized according to procedures described by Cowart (1997) and distributed to participants in two different questionnaire forms in which items were ordered differently from one another.

4.2.2 Participants

AJT 3 was administered to a total of 16 native speakers of Turkish who were recruited from the student body at Purdue University in 2010. As the majority were graduate students, the average age was 32 years old; 8 were males, and 8 were females. None had training in linguistics. Nine participants in this study had previously participated in the English-language AJT.

4.2.3 Experimental procedures and statistical tests used to analyze the data

These were identical to the ones listed for AJT 2.
4.3 Results and discussion

Results from the mixed model statistical test show highly significant main effects for RC Type, Pronoun, and for their interaction.22

Pairwise TALMS test comparisons showed that one condition was primarily responsible for all statistically significant differences between individual types of items. Subject RCs with RPs were rated significantly lower than all other conditions, including direct object RCs with RPs (t= 4.780, p=.001) and indirect/oblique object RCs with RPs (t=5.650, p< .001). All other pairwise TALMS test comparisons that did not involve subject RCs in the RP condition were not statistically significant.

This difference between subject relatives with RPs and all other sentence types can be seen clearly in Figure 6.23

22 Table 2 in Appendix C lists the exact statistics elicited by the Mixed Model Procedure.
23 It should be noted that for the indirect/oblique object condition in Table 7, two token sets contained items with relativized indirect objects, and two token sets contained items with relativized oblique objects. Participants gave an average rating of 3.875 to gapped items with relativized indirect objects, and an average rating of 3.53 to items with RPs in relativized indirect object RCs. They gave an average rating of 3.46 to gapped items with relativized oblique objects, and an average rating of 3.13 to items with RPs in RCs with relativized oblique objects.
These results differed from our original predictions. Although gapped RCs were rated highly in all conditions as expected, RPs were also found to be acceptable in both direct object RCs (average rating 3.109) and in RCs with relativized indirect objects and in oblique object RCs (3.328), contrary to the claims made by Keenan and Comrie (1977), Tezel (1998), Kornfilt (2000), and Çağrı (2005) that RPs are unacceptable in normal Turkish RCs. It should be recalled that in the non-subject conditions, RCs with RPs did not differ from gapped RCs at statistically significant levels, despite the apparent slight differences shown in Figure 6. These results

24 The error bars indicate the standard error of the mean. Exact values of each data point and the standard deviations can be seen in Appendix C, Table 1.
therefore cast doubt on formal syntactic proposals such as Kornfilt’s (2000) formulation of the 
A’-Disjointness requirement—a constraint on the binding of pronouns by antecedents in A’ 
positions. Kornfilt (2000) formulates this constraint for Turkish in such a way as to prevent RPs 
in all positions within the highest clause (i.e., subject, direct object, indirect object, oblique 
object). The current results suggest that RPs are more likely banned from highest subject 
position only, similar to Irish (McCloskey 1990) and Hebrew (Borer 1984).

Although failing to support Kornfilt’s (2000) analysis, the current data are consistent 
with Hawkins’ (2004) FGD complexity ranking, which predicts a greater likelihood of acceptable 
RPs in non-subject as compared with subject positions. Furthermore, Turkish RCs with 
relativized direct, indirect, and oblique objects are more complex from a morphological 
perspective than Turkish subject RCs, in addition to having larger FGDs as per Hawkins’ 
complexity metric. For example, non-subject Turkish RCs must have a morpheme that indicates 
agreement in number with the RC subject, and that is appended to the morpheme used to 
mark a non-subject RC, -dik. In contrast, Turkish subject RCs contain no such agreement 
morpheme. In other words, L1 Turkish speakers might find Turkish non-subject RCs more 
difficult to process than subject RCs for two different reasons: a) the larger filler-gap domain in 
non-subject RCs as compared with subject RCs, and b) the comparatively elaborate morphology 
in non-subject RCs as compared with subject RCs. For more detailed information on these 
issues, see Özge, Marinis, and Zeyrek (2010).

The results for the resumptive conditions were similar to those for Mandarin Chinese, 
except that in Mandarin, direct object RCs with RPs patterned more similarly to subject RCs,
and the complexity effect for RPs (i.e. enhanced acceptability in more deeply embedded positions) was only shown in the indirect object condition. Thus, the Turkish data, which showed increased acceptability for non-subject RCs with RPs as compared with subject RCs with RPs, patterned in much the way that we had originally hypothesized for Mandarin. The Mandarin data itself patterned slightly differently than expected in that direct object RCs with RPs were no more acceptable than subject RCs with RPs. Importantly, though, both Mandarin and Turkish showed increased acceptability for RPs in some non-subject conditions, in contrast to English, for which RPs were judged by the native speaker control group in AJT 1 as equally unacceptable across all RC types. Of these three languages, only the English native speaker results were as expected. We believe that these results from AJT 2 and AJT 3 might help explain the otherwise puzzling results of AJT 1. The lack of any differences between L1 Turkish and L1 Chinese groups as well as the slight but significant difference between subject and non-subject conditions for both learner groups might be attributed to the fact that RPs are apparently licensed in some non-subject positions in both Turkish and Mandarin. In other words, the results of AJT 2 and AJT 3 tell us that L1 transfer cannot be ruled out as an explanation for the results of AJT 1.

Overall, the results obtained from AJT 3 contrast with most published information about the distribution of RPs and gaps in Turkish RCs, all of which appears to be based on informal sentence judgments. More research needs to be done on the distribution of RPs in Turkish RCs, and we believe that informal sentence judgments are not capable of clarifying this issue in sufficient detail. Furthermore, support for the Hawkins’ FGD complexity ranking is illustrated
here by the increasing acceptability of RPs in subject, direct object, and indirect object/oblique contexts.

5. Conclusion

In this section, we will review each experiment in light of the article’s goal of illustrating the need for quantitative research in theory building, discuss the limitations of each experiment, and describe avenues for further research.

5.1 English-language AJT 1: The cross-linguistic transfer of RPs

AJT 1 showed that formal sentence judgment tasks are needed to portray any possible gradient effects. The gradient effect shown in the non-native speakers’ judgments of RCs in the RP condition could not be elucidated by informal sentence judgments.

More generally, AJT 1 showed that SLA researchers interested in researching language transfer need to base their L1-related hypotheses on experimental data. Currently, many SLA researchers still base their L1-related hypotheses on informal sentence judgments (Jarvis and Pavlenko 2008: 49), and this situation needs to change.

The motivation for the gradient data is consistent with Hawkins’ FGD complexity ranking. Non-native speakers may have found RPs to be more acceptable in non-subject RCs as compared with subject RCs due to the greater processing complexity of non-subject RCs. However, the results of AJT 2 and 3 show that an L1 transfer account, or an account that combines processing complexity and L1 transfer, cannot be ruled out.

We encourage future SLA researchers interested in language transfer of RPs to base their transfer-related hypotheses on experimental data for the L1 (instead of only testing L2 performance), to use a larger variety of token sets in a Latin Square design, and to include more
complex RCs with deeper levels of embedding. If we had included more items like these, the very slight gradient effect likely would have become more apparent. Finally, we have only data obtained from receptive tasks. These need to be supplemented with productive tasks as well.

5.2 AJT 2: The distribution of gaps and RPs in Mandarin Chinese

AJT 2 showed that the actual distribution of gaps and RPs in Mandarin Chinese RCs as revealed by experimental data is quite different from the patterns shown by informal sentence judgment tasks. Thus, our own results, in combination with the experimental results of other researchers (Ning 2008; Yuan and Zhao 2005), cast doubt on previous formal syntactic analyses which assumed that RPs are banned from subject RCs but optional in direct object RCs (e.g. Hu and Liu 2007). On the positive side, our results concur with the FGD complexity ranking of Hawkins (2004), and in particular the prediction that RP distributions should be the same in subject and direct object RCs for languages like Mandarin. The precise formulation of constraints on RP distributions in Chinese RCs is quite controversial as shown in the discussion of AJT 2, and more experimental research is clearly needed.

AJT 2 shows another motivation for formal sentence judgments: a gradient data pattern, which informal sentence judgments cannot adequately reveal, but which syntactic analyses of RP distributions ought to take into account. As with the English RCs in AJT 1, RPs gain in acceptability as the filler-gap domain expands, a finding that supports the FGD complexity ranking of Hawkins (2004).

We are indebted to two anonymous reviewers for pointing out these shortcomings.
The limitations of AJT 2 are similar to AJT 1. The gradient pattern could have been more pronounced if the experiment had included additional conditions with more complexity and embedding. Also, the small number of token sets and the lack of a Latin Square Design do limit our ability to generalize these findings, but, as we have stated earlier, the validity of the results may be strengthened by their consistency with those of other experiments. More production data on the distribution of RPs in Mandarin across a wider range of RC types of varying levels of complexity is needed.

5.3 AJT 3: The distribution of gaps and RPs in Turkish

AJT 3 showed that our experimental data were at odds with all informal judgments regarding the distribution of gaps and RPs in Turkish RCs except for those of Göksel and Kerslake (2004), and therefore call into question formal syntactic analyses proposed by Kornfilt (2000). The results are, however, again consistent with Hawkins’ FGD complexity metric, as RPs in non-subject positions are more acceptable than in subject positions. As far as we are aware, ours is the first and only experimental study to address these issues. As such, it remains unclear as to what extent RPs are grammaticalized in Turkish and under what conditions they are used. In addition, it is not known which type of RC complexity (FGD size, morphological complexity, or some combination) might be affecting participants’ judgments. Further experimental research is needed to answer these questions.

As with AJTs 1 and 2, AJT 3 could have benefitted from a larger number of token sets. Also, it is possible that subject RCs in the RP condition could have become more acceptable if more postpositional or adverbial phrases intervened between the RP and the head noun:
compared to the other subject RCs in the RP condition in AJTs 1 and 2, participants rated them fairly high (1.906), and some participants found them to be consistently acceptable.

It can be seen that the experimental results of AJTs 1, 2, and 3 have important limitations and raise more questions than they resolve in regards to only one small area of syntactic research: resumptive pronouns. However, at least three things are clear: (1) syntactic analyses of RPs in Mandarin (e.g. Hu and Liu 2007) and Turkish (e.g., Kornfilt 2000) may need to be revised such that the relevant syntactic constraints are more restrictive for Mandarin (treating direct object RCs on par with subject RCs) but less restrictive for Turkish (allowing RPs in non-subject positions); (2) acceptability judgments correlated with RC complexity for L2 English speakers and for L1 Mandarin and Turkish speakers, consistent with Hawkins’ FGD complexity ranking, thus providing some support for his PGCH; and (3) informally-collected sentence judgments provide little hope of establishing reliable data on RP distributions or of revealing subtle, gradient distinctions. Thus, following Gibson and Fedorenko (2013), we endorse the continued use of formal sentence judgment tasks or other quantitative measures in both theoretical syntax research, which relies on particular contrasts in acceptability to form complex chains of reasoning, and in second-language research, which relies on particular contrasts in acceptability to form hypotheses about language transfer.
Acknowledgments

We would like to acknowledge the extremely helpful contributions of Dr. April Ginther, the Co-Chair of John Hitz’s Ph.D. dissertation committee, who provided valuable guidance regarding logical argumentation, organization of ideas, research design, and statistical analyses. We would like to thank another member of the dissertation committee, Dr. John Sundquist, for giving many detailed and helpful comments throughout, and respectfully recognize the contributions of Dr. Linda Bergmann, a dissertation committee member whose feedback enabled us to explain our arguments in the clearest possible way.

We are very grateful to the Purdue Statistics Department, and especially to Faye Zheng, our research consultant, who helped us choose the appropriate statistical model.

We offer our sincere thanks to Dr. Charles Lam and Dr. Lixia Cheng, who helped enormously with the design of the Mandarin AJT, and to Dr. Xin Li, who helped us to evaluate and critique the individual items.

We express our sincere gratitude to Evrim Eser Hitz, Nevin Eser, and Ayşe Eser for their assistance with the word choices and experimental items in the Turkish AJT.

We would like to thank the anonymous reviewers who provided some incisive, detailed, and useful comments about an earlier draft of this article that was submitted to another journal.

And, finally, we are very grateful to the anonymous reviewers and editorial staff of *Linguistics*. We really appreciate their detailed comments on the first submission of this article, and their patience with our work.
References

Aoun, Joseph and Li, Yen-Hui. 1989. *Three cases of logical relations: Relative scope, bound pronouns, and anaphoric relations*. University of Southern California, M.S.


*Language*, 75(2), pp. 244-285.


Hitz, John. 2012. *A study of the constraints affecting resumption in Turkish and Mandarin Chinese relative clauses, and the transfer of these constraints to English as a second language*. Ph.D. Dissertation. Purdue University.


Labelle, Marie. 1990. Predication, wh-movement, and the development of relative 

Leavitt, Frederick. 2002. *Evaluating scientific research: Separating fact from fiction.* Long 


Berkeley: University of California Press.

Institute, Inc., Cary, NC.

Meral, Hasan. 2004. *Resumptive pronouns in Turkish.* MA dissertation, Bogazici University, 
Istanbul, Turkey.

McCloskey, James. 1990. Resumptive pronouns, A’-Binding and levels of representation in Irish. 
(in) *Syntax and Semantics*, 23, (Syntax of Modern Celtic Languages), 199-248. Academic 
Press Inc.

*Language Acquisition*, 9(2), 113-156.

clauses.* MA dissertation, National Cheng-Chi University, Taiwan.

in Mandarin relative clauses: An eye-tracking study.* Poster session presented at the 
27th Annual CUNY Conference on Human Sentence Processing, Columbus, Ohio, USA.


Sprouse, Jon & Almeida, Diogo. 2013b. The role of experimental syntax in an integrated

Sprouse, Jon, Schütze, Carson, & Almeida, Diogo. 2013. A comparison of informal and formal
*Lingua* (134), 219-248.


Tezel, Zubeyde. 1998. *The acquisition of restrictive relative clause configurations in English as a

61, 1–27.


(11): 1481-1496.

Yuan, Boping, and Zhao, Yang. 2005. Resumptive pronouns in English-Chinese and
Arabic-Chinese interlanguages. *International Review of Applied Linguistics in
Appendix A

The means of the ratings assigned to the items in the various experimental conditions in AJT 1, as well as the standard deviations between them, are illustrated in Table 1.

Table 1. Mean Ratings of Experimental Items

<table>
<thead>
<tr>
<th></th>
<th>Mandarin Gaps</th>
<th>Turkish Gaps</th>
<th>English Gaps</th>
<th>Mandarin RP</th>
<th>Turkish RP</th>
<th>English RP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td>3.58 SD=.35</td>
<td>3.76 SD=.306</td>
<td>3.74 SD=.41</td>
<td>1.27 SD=.44</td>
<td>1.19 SD=.37</td>
<td>1.13 SD=.23</td>
</tr>
<tr>
<td>Direct</td>
<td>3.22 SD=.59</td>
<td>3.28 SD=.875</td>
<td>3.53 SD=.61</td>
<td>1.58 SD=.74</td>
<td>1.59 SD=.73</td>
<td>1.11 SD=.24</td>
</tr>
<tr>
<td>Object</td>
<td>3.51 SD=.52</td>
<td>3.02 SD=.978</td>
<td>3.53 SD=.75</td>
<td>1.60 SD=.697</td>
<td>1.52 SD=.68</td>
<td>1.17 SD=.35</td>
</tr>
</tbody>
</table>

The main effects and interactions of various factors that might have affected participants’ performance in AJT 1 are depicted in Table 2.

Table 1. Influence of Various Factors on AJT 1

<table>
<thead>
<tr>
<th></th>
<th>Numerator</th>
<th>Denominator</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>language</td>
<td>2</td>
<td>198</td>
<td>0.68</td>
<td>0.5053</td>
</tr>
<tr>
<td>Type</td>
<td>2</td>
<td>393</td>
<td>1.69</td>
<td>0.1852</td>
</tr>
<tr>
<td>Factor</td>
<td>Numerator DF</td>
<td>Denominator DF</td>
<td>F Value</td>
<td>Pr &gt; F</td>
</tr>
<tr>
<td>-----------------------</td>
<td>--------------</td>
<td>----------------</td>
<td>---------</td>
<td>--------</td>
</tr>
<tr>
<td>language*type</td>
<td>4</td>
<td>393</td>
<td>5.28</td>
<td>0.0004</td>
</tr>
<tr>
<td>pronoun</td>
<td>1</td>
<td>198</td>
<td>923.63</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>type*pronoun</td>
<td>2</td>
<td>393</td>
<td>37.04</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>language*pronoun</td>
<td>2</td>
<td>198</td>
<td>6.20</td>
<td>0.0024</td>
</tr>
<tr>
<td>language<em>type</em>pronoun</td>
<td>4</td>
<td>393</td>
<td>6.03</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

The mixed-model statistical procedure showed a significant main effect for one of the fixed factors, Pronoun (F(1, 198), F = 923.63, p < .001). The significant main effect of Pronoun indicates that, as expected, gapped RCs were rated much higher than resumptive RCs across all language groups and all RC types.

The lack of any main effect for Language (F(2, 198) =.680, p=.505) and RC Type (F(2, 393) =1.68, p=.185) is the result of variation across participant groups and clause types among the different experimental conditions. However, this variation was found to be systematic, as indicated by the significant interactions between Language and RC Type (F(4, 393) =5.280, p=.0004), Language and Pronoun (F(2, 98)=6.202, p = .002), RC Type and Pronoun (F(4, 393) =37.040, p < .001), and between Language, RC Type, and Pronoun (F(4, 393) =6.030, p < .001).
The most interesting of these is the significant three-way interaction between Language (L1 background), RC Type, and Pronoun (Presence/Absence of RP).

We conducted several pairwise comparisons using Tukey-Adjusted Least Mean Squares tests (hereafter, TALMS tests) to check for significant differences in the ratings between L1 groups. Contrary to our expectation for L1 transfer effects, there were no statistically significant differences between the two groups of non-native speakers in regards to their judgments of direct object RCs with RPs (t= .030, p=1.000), and in regards to their judgments of RCs with relativized objects of prepositions and RPs (t= 1.070, p=.999). Likewise, the L1 Chinese speakers did not differ at statistically significant levels from the native English speakers in terms of their judgments of similar items on the questionnaire, and the L1 Turkish speakers differed at statistically significant levels from the L1 English speakers on only one condition: gapped RCs with stranded prepositions (t=3.520, p= .049). These data appear to show that the non-native English speakers were so advanced in their L2 RC acquisition that their acceptability judgments for sentences with RPs closely resembled those of native speakers.

---

1 For a detailed explanation of this finding, see Hitz (2012: 152-165). Because Chinese has prepositions and SVO word order, just as English does, L1 Chinese speakers may have an advantage over L1 Turkish speakers in the acquisition of English RCs with stranded prepositions. In contrast to English and Chinese, Turkish allows postpositions only and is characterized by a canonical SOV word order.
Appendix B

The means of the ratings assigned to the items in the various experimental conditions in AJT 1, as well as the standard deviations between them, are illustrated in Table 1.

Table 1. *Mean Ratings of Experimental Items*

<table>
<thead>
<tr>
<th>Condition</th>
<th>Subject</th>
<th>Direct</th>
<th>Indirect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gapped</td>
<td>3.948</td>
<td>3.906</td>
<td>2.697</td>
</tr>
<tr>
<td>RP</td>
<td>1.385</td>
<td>1.635</td>
<td>2.36</td>
</tr>
</tbody>
</table>

The main effects and interactions elicited by the Mixed Model procedure can be seen in Table 2.

Table 2. *Main Effects and Interactions: L1 Mandarin Survey*

<table>
<thead>
<tr>
<th>Effect</th>
<th>Numerator</th>
<th>Denominator</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>2</td>
<td>62</td>
<td>2.16</td>
<td>0.1244</td>
</tr>
<tr>
<td>pronoun</td>
<td>1</td>
<td>31</td>
<td>332.37</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>pronoun*type</td>
<td>2</td>
<td>62</td>
<td>54.83</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>
Appendix C

The means of the ratings assigned to the items in the various experimental conditions AJT 3, as well as the standard deviations between them, are illustrated in Table 1.

Table 1. Mean Ratings of Experimental Items

<table>
<thead>
<tr>
<th>Subject</th>
<th>Direct</th>
<th>Indirect / Object</th>
<th>Oblique</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gapped</td>
<td>3.984</td>
<td>3.81</td>
<td>3.67</td>
</tr>
<tr>
<td>Condition</td>
<td>SD=.422</td>
<td>SD=.370</td>
<td>SD=.425</td>
</tr>
<tr>
<td>RP</td>
<td>1.906</td>
<td>3.10</td>
<td>3.32</td>
</tr>
<tr>
<td>Condition</td>
<td>SD=.1.060</td>
<td>SD=.921</td>
<td>SD=.751</td>
</tr>
</tbody>
</table>

The main effects and interactions elicited by the Mixed Model procedure can be seen in Table 2.

Table 2. Main Effects and Interactions: L1 Turkish Survey

<table>
<thead>
<tr>
<th>Effect</th>
<th>Num DF</th>
<th>Den DF</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>2</td>
<td>30</td>
<td>8.76</td>
<td>0.0003</td>
</tr>
<tr>
<td>Pronoun</td>
<td>1</td>
<td>15</td>
<td>44.4</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>---------</td>
<td>---</td>
<td>----</td>
<td>------</td>
<td>--------</td>
</tr>
<tr>
<td>Type*</td>
<td>2</td>
<td>30</td>
<td>9.89</td>
<td>0.0001</td>
</tr>
<tr>
<td>Pronoun</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>