PROCESSING WH-DEPENDENCIES IN L2 ENGLISH: THE ROLE OF L1 AND WORKING MEMORY CAPACITY

Thesis submitted to the

Institute for Graduate Studies in the Social Science

In partial fulfillment of the requirement for the degree of

Doctor of Philosophy

in

Foreign Language Education

by

Filiz Çele

Boğaziçi University

2010

Thesis Abstract

Filiz Çele, "Processing *wh*-dependencies in L2 English: The role of L1 and working memory capacity"

This study investigates online processing of long-distance wh-dependencies in English by a group of Turkish- and Spanish-speakers of second language (L2) English in comparison to a group of native English speakers to explore whether end-state L2 speakers achieve native-like processing in the domain of wh-extractions. The study also examines the role of first language (L1) and working memory capacity (WMC) in on-line processing of wh-dependencies. To this end, speakers of L1 Turkish as well as speakers of L1 Spanish have been included in the study. Turkish is a wh-in-situ language. Nevertheless, it allows overt wh-movement via scrambling. Therefore, it provides an interesting testing case to verify the influence of scrambling in correctly accepting grammatical wh-extractions and rejecting ungrammatical wh-extractions with island violations. In this vein, the study focuses on the question of whether or not L1 Turkish speakers will be as accurate and as fast as L1 speakers of Spanish, a language with overt wh-movement in processing long-distance wh-extractions.

Additional questions investigated in this study are (1) whether there is a subject-object asymmetry in wh-extractions from finite and nonfinite clauses, and (2) whether there is a relationship between the WMC and sentence processing performance in the L2.

An online grammaticality judgment task (OGJT) involved both grammatical wh-extractions from finite and nonfinite clauses and ungrammatical wh-extractions with island violations and this task was presented in two conditions, namely, the full-sentence condition and the self-paced word-by-word reading in the moving window condition. Both response accuracy and response latency (i.e. reading time) were measured in these conditions. To determine the WMC, all participants were tested on two online working memory (WM) tasks in English: 1) automated reading span (ARSAN) task, and automated operation span (AOSAPN) task. Spanish and Turkish participants also took the ARSPAN task in their respective L1.

The accuracy results from the two conditions revealed that Turkish and Spanish speakers were as accurate as native English speakers in correctly accepting grammatical wh-extractions and rejecting ungrammatical wh-extractions with island violations, except for subject extraction from nonfinite clauses, and wh-extractions with *that*-trace violations. L2 learners were also similar to native speakers in reading patterns. Furthermore, there was no difference between Turkish and Spanish groups in comprehension accuracy and reading time for grammatical and ungrammatical wh-dependencies in L2 English. This suggests that L2 speakers whose L1 allows overt wh-movement (i.e. Spanish speakers of English) do not outperform L2 speakers with a wh-in-situ L1 (i.e. Turkish speakers of English). The presence of overt wh-movement in scrambled sentences in the L1 (as in the case of Turkish) might be playing a role in accurate processing of wh-extractions in the L2. The results may also suggest that in the end-state L2, speakers achieve native-like processing irrespective of the syntactic properties of their L1.

With respect to problematic sentence types, all participants had problems in processing subject extraction from nonfinite clauses but no such problem was observed in object extractions. Although the two L2 groups were significantly less accurate than native English speakers in processing subject extraction from nonfinite clauses, we can still say that the subject-object asymmetry in wh-extractions in nonfinite clauses is a characteristic of both native and nonnative sentence processing. This asymmetry can be accounted for complex processing involving multiple changes in the assignment of theta roles and Cases in subject extraction from nonfinite clauses.

Results from word-by-word RTs showed that the locus of the difficulty for all participants in subject extractions from nonfinite clauses is the embedded overt object NP, where they experienced a filled-gap effect. This finding is consistent with the predictions of the both gap-based accounts (e.g. the Active Filler Hypothesis (AFS) of Frazier, 1989) and gapless accounts (e.g. the Direct Association Hypothesis (DAH) of Pickering & Barry, 1993) and also with the Principled-parsing account of Pritchett (1992); Gibson, (1991); Weinberg (1999). Also, this provides evidence that L2 learners apply similar processing strategies to those of the native speakers in processing wh-dependencies in L2.

With respect to *that*-trace violation, the two L2 groups, particularly the Spanish group was less accurate than native speakers. This might be due to the fact that both Spanish and Turkish are pro-drop languages, in which null-subject sentences are grammatically possible. Moreover, in L1 Spanish, the complementizer obligatorily precedes the subject trace in the embedded clauses.

The RT analysis of *that*-trace violation showed that the embedded verb following the complementizer *that* is the locus of the difficulty for all participants, reflected in their longer RTs at embedded verb. However, the Spanish group's RTs were shorter, suggesting that the Spanish speakers were not as surprised as Turkish learners and the native speakers once they met a finite verb after the complementizer, which might be a result of local influence of the L1. Also, word-by-word RTs showed that none of the groups expected an illicit gap site inside the islands, except for subject island violation, which provides suggestive evidence that both native and nonnative speakers are sensitive to island constraints in L2 English.

Results of the WM tasks demonstrated that the ARSPAN and AOSPAN scores were significantly correlated with each other, suggesting that they measure the same construct. The Turkish and Spanish ARSPAN scores were significantly correlated with the L2 English ARSPAN score, which suggests that the WM is a language-independent construct. However, no significant relationship was observed between the WMC and accuracy and RTs for the OGJT, which implies that WMC cannot account for the difficulty in subject extraction from nonfinite clauses or *that*-trace violation, and provide evidence for the Separate Resource Theory of Waters and Caplan (1996), according to which verbal WM has two separate pools of resources: (1) a specialized verbal WM for interpretive process (i.e., automatic, first-pass language processing), and (2) and a more general verbal WM for post-interpretive processes (i.e., controlled conscious processing of the propositional content of the sentence and using it to accomplish tasks, like reasoning, planning actions, storing information in long term memory etc.)

Tez Özeti

Filiz Çele "İkinci dil olarak İngilizce'de *Wh*-soru bağımlılıklarını işleme: Birinci dil ve işler-bellek kapasitesinin rolü"

Bu çalışma, birinci dili (anadili) (D1) Türkçe ve İspanyolca olan ve ikinci dil (yabancı dil) (D2) İngilizce'de nihai dilbilgisi düzeyine erişmiş iki ayrı grubun, İngilizce'deki kompleks *wh*-soru yapılarında İngilizce'yi anadil olarak konuşanlar gibi tümce işleme yapıp yapmadıklarını anlamak amacıyla planlanmıştır. Çalışma bilgisayar ortamında verilen testler aracılığıyla yapılmış ve kişilerin gerçek zaman dilimi içinde soru tümcelerini işleme hızı ve bu tümcelerin dilbilgisel açıdan kabul edilebilirliklerine ilişkin doğru yargıya varıp varamadıkları incelenmiştir. Çalışma, ayrıca D1 ve işler bellek kapasitesinin (İBK) soru tümcelerinin gerçek zaman dilimi içinde işlenmesindeki rolünü araştırmaktadır.

Bu amaçla çalışmaya Türkçe D1 konuşanlarıyla birlikte İspanyolca D1 konuşanları da dahil edilmiştir. Türkçe, İngilizce ve İspanyolca'nın aksine soru yapılarının yüzeysel düzeyde tümce başına taşınmadığı (wh-in situ) bir dildir. Fakat çalkalama (scrambling) yoluyla açık soru formlarının taşımasına olanak verir. Dolayısıyla bu çalışma, D2'de dilbilgisine uygun soru çıkarımlarını kabul etme ve ada ihlalleri nedeniyle dilbilgisine aykırı olan çıkarımları reddetmede anadildeki çalkalamanın etkisini sınama imkanı sağlar. Bu bağlamda, çalışma, Türkçe D1 konuşanların uzun mesafeli soru çıkarımlarını işlemede, açık soru taşımasının olanaklı olduğu İspanyolca'yı ana dili olarak konuşanlar kadar doğru ve hızlı olup olmadığını ortaya çıkarmayı hedeflemektedir.

Bunların dışında çalışmada yanıt aranan diger sorular, çekimli ve çekimsiz tümceciklerden soru formları çıkarımlarında özne-nesne asimetrisinin olup olmadığı ve D2'de İBK ve tümce işleme arasında bir ilişkinin olup olmadığıdır. Kullanılan çevrimiçi dilbilgisel yargı testi (ÇDYT), hem çekimli hem de çekimsiz tümceciklerden dilbilgisine uygun *Wh*-çıkarımlarını, hem de ada ihlalleri içeren *Wh*-çıkarımlarını kapsamaktadır. Test iki ayrı şekilde uygulanmıştır: tam tümce durumu ve kayan ekran durumunda kisişel hıza göre sözcük sözcük okuma. Yanıtların doğruluğu ve yanıt verme süresi (okuma süresi) ölçülmüştür. İBK'yi belirlemek için ise tüm katılımcılara iki İngilizce çevrimiçi işler bellek (İB) testi uygulanmıştır: 1) Otomatik okuma süresi testi (OOST) ve 2) otomatik işlem süresi testi (OİST). İspanyol ve Türk katılımcılar OOST testini ana dillerinde de almıştır.

Bu iki durumda verilen yargı testleri hem doğru yanıtlar hem de yanıt verme süresi göz önüne alınarak değerlendirilmiştir. Bulgular, hem Türk hem İspanyol grubun birçok tümce kategorisinde anadili İngilizce olanlar kadar başarılı olduklarını ortaya koymuştur. Başarısız olunan kategoriler çekimsiz tümceciklerden özne çıkarımı ve ki-tümleyen (*that*-trace) ihlalleridir. Aynı zamanda ikinci dil konuşanlarının okuma örüntüleri anadili İngilizce olanlarla benzerlik göstermektedir. Bununla birlikte Türk ve İspanyol gruplar arasında ikinci dil İngilizce'de dilbilgisine uygun olan ve olmayan *Wh*-bağımlılıklarını anlama başarısı ve okuma süresi bakımından bir fark bulunmamıştır. Bu bulgu, anadili açık soru formu taşımasına olanak veren bir D1 konuşanlarının (İngilizce

konuşan İspanyollar) ve anadili soru formu taşımasına izin vermeyen (wh-in situ) D1 konuşanlarına göre (İngilizce konuşan Türkler) daha başarılı olmadıklarını göstermektedir. Anadil Türkçe'de çalkalama içeren tümcelerde açık soru formu taşımasının bulunmasının ikinci dil İngilizce'de *Wh*-çıkarımlarınn doğru işlenmesinde önemli bir payı olabilir. Elde edilen sonuçlar ileri düzey D2 İngilizce yetisine ulaşmış kişilerin anadillerinden gelebilecek etkileri göstermediklerini ortaya koymuştur.

Sorunlu tümce çeşitleri açısından tüm katılımcıların çekimsiz tümceciklerden özne çıkarımını işlemede sorun yaşarken, nesne çıkarımında benzer bir sorun yaşadıkları gözlenmemiştir. Çekimsiz tümceciklerden özne çıkarımında her iki ikinci dil grubunun da anadili İngilizce olanlara göre daha az doğru yanıt vermiş olmasına rağmen, çekimsiz tümceciklerde *wh*-çıkarımlarında özne-nesne asimetrisinin hem anadilde hem de ikinci dilde tümce işlemenin bir özelliği olduğu söylenebilir. Bu asimetri, çekimsiz tümceciklerden özne çıkarımında theta rolleri ve durum eklerinin atanmasında çoklu değişimler içeren karmaşık işlemeyle açıklanabilir.

Sözcük sözcük okuma süresi sonuçları çekimsiz tümceciklerde özne çıkarımında tüm katılımcıların karşılaştığı güçlüğün odak noktasının, yan tümcecikte "doldurulmuş boşluk" (filled-gap) etkisi yaşadıkları yerde bulunan açık nesne ad öbeği (AÖ) olduğunu göstermiştir. Bu bulgu, hem boşluğa dayalı (ör. Frazier'ın 1989'da ortaya attığı Aktif Doldurucu Stratejisi Kuramı) hem de boşluk içermeyen (ör. Pickering & Barry'nin 1993- Doğrudan İlişki Kuramı) açıklamalarıyla uyumludur. Ayrıca bu bulgu ikinci dil konuşanlarının ikinci dilde *Wh*-bağımlılıklarını işlemede anadili İngilizce olanlarla benzer işleme stratejileri uyguladıklarına dair kanıt sağlamaktadır.

That-izi ihlalleri açısından iki ikinci dil grubu da—özellikle İspanyol grubu—anadili İngilizce olanlara göre konuşanlarına göre daha az başarı göstermiştir. Bu durum İspanyolca ve Türkçe'nin öznesiz tümcelerin dilbilgisel açıdan olanaklı olduğu adıl düşmeli diller olmasına bağlı olabilir. Ayrıca İspanyolca'da yan tümceciklerde tümleyen zorunlu olarak öznenin izinden önce gelmektedir.

That ihlali taşıyan tümcelerin okuma süreleri incelendiğinde, tümleyen *that*'i izleyen yan tümcecik eylemininin tüm katılımcılar için sorunun ana kaynağını oluşturduğu görülmüştür ki bu durum yan tümcecik eyleminde daha uzun okuma süresi olarak kendini göstermektedir. Fakat, İspanyol grubun okuma süresi daha kısa çıkmıştır. Bu da İspanyol katılımcıların, muhtemelen anadillerinin etkisiyle, tümleyenden sonra çekimli bir eylemle karşılaştıklarında Türkler veya anadili İngilizce olanlar kadar şaşırmadıkları anlamına gelmektedir. Bununla birlikte sözcük sözcük okuma süreleri hiçbir grubun ada içinde özne adası ihlali haricinde ada tümceciklerinde uyumsuz bir boşluk beklemediğine işaret etmiştir. Buradan anadili İngilizce olanların da olmayanların da ikinci dilde ada kısıtlamalarına karşı hassas oldukları sonucu çıkarılablir.

İB testlerinin sonuçları OOST ve OİST puanları arasında önemli bir ilişki olduğunu ortaya koymaktadır; bu da aynı olguyu ölçtüklerine işaret etmektedir. Türkçe ve İspanyolca OOST puanları ile D2 İngilizce'de verilen OOST puanları arasında önemli bir ilişki bulunmaktadır. Bu da İB'in dilden bağımsız bir olgu olduğunu göstermektedir. Fakat, İBK ve ÇDYT doğruluk ve okuma süreleri arasında önemli bir ilişki bulunmamıştır. Buna göre İBK çekimsiz tümceciklerde özne çıkarımı veya *that*-izi ihlallerini açıklayamamaktadır. Bu bulgu Waters ve Caplan'ın (1996) sözel İB'in iki ayrı kaynağı—1) yorumlama süreçlerine özel sözel İB (otomatik, ilk geçiş dil işlemesi vb.) 2) yorumlama sonrası süreçler (tümcenin önermesel içeriğinin kontrollü, bilinçli işlemesi ve mantık yürütme, eylem planlama, uzun süreli bellekte bilgi saklama gibi görevler için kullanılması vb.) için daha genel bir sözel İB—olduğunu iddia eden Ayrı Kaynak Teorisi'ni (*Separate Resource Theory*) desteklemektedir.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my advisor Assoc. Prof. Ayşe Gürel, for her constant support and encouragement throughout the whole work with this dissertation. She always provided constructive feedbacks and comments on each chapter of this dissertation. Without her help, my thesis would not have achieved the current quality of content. I have learned a lot about the parameters of true academics from her strict and professional attitude toward research.

I would like to thank Prof. Dr. Deniz Zeyrek for her insights, valuable comments and suggestions on this dissertation. I also thank Assoc. Prof. Belma Haznedar for her meticulous reading, constructive feedbacks and comments on this dissertation. Her critical reading contributed a lot to the dissertation.

I am deeply grateful to Assist. Prof. Balkız Öztürk for her valuable comments and suggestions on the syntactic background of this dissertation. Her feedbacks were excellent with respect to syntactic theories which underlie this study.

I would like to thank Claudia Felser who helped me construct the experimental items for grammaticality judgment task in this study, and understand critical issues in designing experimental tasks in online sentence processing. I also thank Harald Clahsen for his valuable insights and suggestions on the content of this dissertation, and his help in finding a lab to run my experiments at the department of Language and Linguistics at University of Essex.

My gratitude extends to Language and Linguistics members of University of Essex who helped me pilot and run the experiments: Sonja Eisenbeiss, Louise Esher, Kaili Clackson, Belen Gaspar Garcia, Claire White and Ian Cunnings and Nick Bertenshaw

I am also grateful to Assoc. Prof. Yasemin Bayyurt for her valuable support and encouragement during the whole process of this dissertation. She always provided a good environment for me to finish this dissertation.

My special thanks go to Assist. Prof. Gülcan Erçetin and Assist. Prof. Bengü Börkan and Hüseyin Şimşek for their invaluable comments and suggestions on statistics. They always helped me find a way out of problems I had in my statistical analyses.

Many thanks go to my friends who encouraged and supported me during the slow process of this dissertation: Sebiha Kavak, Zeliş Çoban, Filiz Rızaoğlu, Feride Özdemir, Ebru Bağ, Saadet Tıkaç, Emine Merdin and Mehmet Maden. I also thank all the participants in this study for taking five online experiments. It is certain that this thesis would not have been possible without them.

Finally, I would like to thank my family for their constant love, support, and encouragement throughout this study. I would never have been able to finish this work without them.

I would like to acknowledge that the research in this disseration was fully supported by the Boğaziçi University Scientific Fund’s research grant (no: 07D601D) given to Assoc. Prof. Ayşe Gürel. I would also like to thank Assoc. Prof. Ayşe Gürel for making it possible for me to receive financial support for my dissertation within her TÜBA-GEBİP grant program.

CONTENTS

CHAPTER 1: INTRODUCTION 001

CHAPTER 2: SYNTACTIC BACKGROUND 010
2.1. Wh-movement: Introduction 010
2.2. Overt movement 012
2.3 Covert movement 017
2.3.1 Selectional requirement 018
2.3.2 Scope taking properties 020
2.3.3 Weak crossover effects 021
2.3.4. Locality effects 022
2.4. Island constraints and Subjacency in English 026
2.4.1. Complex NP Island (CNPI) 028
2.4.2. Relative Clause Island (RCI) 030
2.4.3. Adjunct Island (AI) 032
2.4.4. Subject Island (SI) 034
2.4.5. *That*–trace effect 035
2.5. Wh-movement in Spanish 037
2.6. Island Constraints and Subjacency in Spanish 040
2.6.1. Complex NP Island (CNPI) 042
2.6.2. Relative Clause Island (RCI) 042
2.6.3. Adjunct Island (AI) 043
2.6.4. Subject Island (SI) 043
2.6.5. *That*–trace effect 044
2.7. Wh-movement in Turkish 046
2.7.1. Covert movement in Turkish 047
2.7.1.1. Selectional requirement 048
2.7.1.2. Scope taking properties 049
2.7.1.3. Weak Crossover effects 051
2.7.1.4. Locality effects 051
2.7.2. Scrambling in Turkish 053
2.7.2.1. Reconstruction effect 055
2.7.2.2. Weak Crossover effects 057
2. 8. Scrambling and Island constraints in Turkish 058
2.8.1. Complex NP Island (CNPI) 059
2.8.2. Relative Clause Island (RCI) 060
2.8.3. Adjunct Island (AI) 061
2.8.4. Subject Island (SI) 062
2.8.5. *That*–trace effect 063

CHAPTER 3: SENTENCE PROCESSING IN THE FIRST AND THE SECOND LANGUAGE 064
3.1. Introduction 064
3.2. Theories of L1 sentence processing 064

3.2.1. Modular (two-stage) theories in L1 sentence processing …………… 066
3.2.2. Interactive (constraint-based/satisfaction) theories in L1 sentence processing ……………………………………………………………… 070
3.3. Processing wh-dependencies in the L1 ……………………………………… 073
3.3.1. Gap-based accounts in processing of wh-dependencies ……………… 075
3.3.2..Gap-free (traceless) account of wh-processing ………………………… 082
3.4. Working memory and L1 sentence processing ……………………………… 089
3.4.1. Introduction … ……………………………………………………………… 089
3.4.2. Working memory constraints in L1 sentence processing …………… 091
3.4.3. The nature of WM resources: Single Resource Theory vs. Separate Language Interpretation Resource Theory …………………………… 094
3.4.4. Measuring Working Memory Capacity: Working Memory Span Tasks ………………………………………………………………………… 105
3.5. Theories of L2 sentence processing …………………………………………… 111
3.5.1. Introduction …………………………………………………………… 111
3.5.2. Studies in L2 sentence processing …………………………………… 112
3.5.3 Processing wh-dependencies in the L2 ………………………………… 115
3.5.4. Working memory and L2 sentence processing ……………………… 130

CHAPTER 4: METHODOLOGY ……………………………………………………… 142
4.1. Introduction ……………………………………………………………………… 142
4.2. Participants ……………………………………………………………………… 151
4.3. Materials ………………………………………………………………………… 153
4.4. Instruments ……………………………………………………………………… 154
4.5. Procedure ……………………………………………………….......................... 161

CHPAPTER 5: RESULTS ……………………………………………………………… 162
5.1. Introduction ……………………………………………………………………… 162
5.2. Results from the full-sentence condition ………………………………………… 165
5.2.1. Accuracy judgments on grammatical *wh*-extraction types …………… 165
5.2.2. Reading times on grammatical *wh*-extraction types …......................... 174
5.2.3. Accuracy judgments on ungrammatical wh-extraction types............. 182
5.2.4. Reading times on the ungrammatical wh-sentences …………………… 190
5.2.5. Summary of the accuracy and RT analyses for the grammatical and ungrammatical wh-extractions ……………. ………………………... 196
5.3. Results from the moving window condition ……………………………………… 198
5.3.1. Accuracy judgments on grammatical wh-extractions types ………… 199
5.3.2. Word-by-word reading times for grammatical wh-extraction types …………………………………………………………………….. 207
5.3.2.1. Word-by-word reading times for object and subject extractions from finite clauses ……………………………… 207
5.3.2.2. Word-by-word reading times for subject and object extraction from nonfinite clauses …………………………… 217
5.3.2.3. Word-by-word reading times for object extraction from finite clauses with complementizer *that* ……………………… 226

5.3.3. Accuracy judgments on ungrammatical wh-extractions 232
5.3.4. Word-by-word reading times for the ungrammatical wh-extractions with island violations 240
5.3.4.1. Word-by-word reading times for the ungrammatical wh-sentences with that- trace violation 240
5.3.4.2. Word-by-word reading times for ungrammatical wh-sentences with subject- island violation 246
5.3.4.3. Word-by-word reading times for ungrammatical wh-sentence with relative clause island violation 252
5.3.4.4. Word-by-word reading times for ungrammatical wh-sentences with adjunct island violation 256
5.3.4.5. Word-by-word reading times for ungrammatical wh-sentences with complex NP island violation 261
5.3.5 Summary of the results from accuracy and reading time analyses in the moving window 266
5.4. Results from working memory tasks 270
5.4. 1. Introduction 270
5.4. 2. Results of the correlations among the WM measures 274
5.4.3. WM measures and accuracy and RTs for wh-extractions in the full-sentence 279
5.4.4 WM measures and RTs for wh-extraction in the self-paced word-by-word reading 293
5.4.5. Summary 297

CHPAPTER 6: DISCUSION AND CONCLUSION 299
6.1. Introduction 299
6.2. Discussion 300
6.3. Conclusion 316

APPENDICES 320
A. OGJT Experimental items 320
B. Sentences used in English ARSPAN 323
C. Sentences used in Spanish ARSPAN 326
D. Sentences used in Turkish ARSPAN 329

REFERENCES 332

TABLES

1. L2 speakers' background information………………………………………………….. 152
2. Mean accuracy scores for five grammatical wh-extraction types ………………... 166
3. Mean RTs for five grammatical wh-extraction types …………………..……….. 175
4. Mean accuracy scores for ungrammatical wh-extraction types …………………. 182
5. Mean RTs for five ungrammatical wh-extraction types …………………………. 190
6. Mean accuracy scores for five grammatical wh-extraction types in the MWC….. 199
7a. Word-by-word RTs for subject and object extraction from finite clauses in milliseconds (ms.) …………………………………………………………………... 210
7b Subject and object extractions from finite clauses ……………………………….. 210
8a. Word-by-word RTs for subject and object extraction from nonfinite clauses …… 219
8b. Regions for subject and object extractions from nonfinite clauses ……………… 220
9a. Mean RTs for word-by-word reading for object extractions from finite clauses with and without complementizer *that* in ms ……………………........................ 227
9b. Object extraction from finite clause with '*that*' ………………………………….. 228
10. Mean accuracy scores for five ungrammatical wh-extraction types ……………... 233
11a Mean RTs for subject extractions from finite clauses with *that-trace* ..………….. 242
11b Regions in ungrammatical wh-extractions with *that*-trace violations ……………. 242
12a. Mean RTs for the ungrammatical *wh*-sentences with subject island violation ….. 248
12b Regions for subject Island ……………………………………………………….. 248
13a Mean RTs for the ungrammatical *wh*-sentences with RCI violation …………….. 253
13b Regions for relative clause island ……………………………………………….. 253
14a. Mean RTs for the ungrammatical *wh*-sentences with AI violation ……………… 258
14b Adjunct island …………………………………………………………………… 258
15a. Mean RTs for the ungrammatical *wh*-sentences with complex NP/DP island violation ……………………………………………………………………….. 262
15b. Complex NP/DP Island ………………………………………………………….... 262
16. Mean scores of L1 ARSPAN, L2 ARSPAN and AOSPAN ……………………...... 274
17. Correlations among the WM measures of English speakers …………………… 277
18. Correlations among the WM measures of Spanish speakers …………………….. 277
19 Correlations among the WM measures of Turkish speakers ……………… 278
20 Correlations among the WM measures and accuracy scores for English native speakers ……………………………………………………………………….... 284
21 Correlations among the WM measures and accuracy scores of Spanish speakers... 286
22 Correlations among the WM measures and accuracy scores of Turkish Speakers………………………………………………………………………… 290

FIGURES

1. Illustration of the automated operation span task ………………………………….. 158
2. Mean accuracy scores for five grammatical wh-extraction types ………………….. 167
3. English native speakers' mean accuracy scores on five grammatical wh-extraction types ……………………………………………………... …… 169
4. Spanish speakers' mean accuracy scores on five grammatical wh-extraction types ……………………………………………………………………... 170
5. Turkish speakers' mean accuracy scores on five grammatical wh-extraction types ……………………………………………………………………… 172
6. Mean RTs for the five grammatical types by three language groups ……….......... 176
7. English native speakers' mean RTs on five grammatical wh-extraction types …………………………………………………………………......... 178
8. Spanish speakers' mean RTs for five grammatical wh-extraction types …..……….. 179
9. Turkish speakers' mean RTs on five grammatical wh-extraction types …………… 180
10. Mean accuracy scores for five ungrammatical types by three groups ……………... 184
11. English native speakers' mean accuracy scores for ungrammatical wh-extraction types ………………………………………………………….. 186
12. Spanish speakers' mean accuracy scores for ungrammatical wh-extraction types ……………………………………………………………………… 187
13. Turkish speakers' mean accuracy scores for ungrammatical wh-extraction types ....……………………………………………………………… 188
14. Mean RTs of language groups across five ungrammatical wh-extraction types …… 191
15. English native speakers' mean RTs for ungrammatical wh-extraction types ……… 193
16. Spanish speakers' mean RTs for five ungrammatical wh-extract types …………... 194
17. Turkish speakers' mean RTs on five ungrammatical wh-extraction types ………... 195
18 Mean accuracy scores for five types of grammatical wh-extractions in the MWC…. 201
19. English native speakers' mean accuracy scores for grammatical wh-extractions in the MWC ……………………………………………………….. 203
20. Spanish learners' mean accuracy scores for grammatical wh-extractions in the MWC ………………………………………………………………….. 204
21. Turkish speakers' mean accuracy scores grammatical wh-extractions in the MWC ……………………………………………………………………… 205
22. Word-by-word RTs for subject extraction from finite clauses ……….…………… 212
23. Word-by-word RTs for object extraction from finite clauses …………………….. 212
24. Word-by-word reading times for subject extraction from nonfinite clauses ……… 222
25. Word-by-word reading times for object extractions from nonfinite clauses . ……… 222
26. Word-by-word reading times for object extractions from finite clauses with complementizer *that* ……………………………………………………………. 229
27. Word-by-word reading times for object extractions from finite clauses with complementizer *that* …………………………………………………………….. 229
28. Mean accuracy scores for five types of ungrammatical wh-extractions …………... 234
29. English native speakers' mean accuracy scores for five ungrammatical wh-extraction types ……………………………………………………………. 236
30 Spanish speakers' mean accuracy scores for five types of ungrammatical wh-extractions ……………………………………………………………………. 237

31. Turkish speakers' mean accuracy scores for five types of wh-extractions 238
32. Word-by-word RTs for the ungrammatical subject extractions from finite clauses with *that*-trace …………………………………………………………….......................... 243
33. Word-by-word reading for the ungrammatical *wh*-sentence with SI violation …… 249
34. Word-by-word RTs for the ungrammatical wh-sentences with RCI violation 254
35. Word-by-word reading times for the ungrammatical wh-sentences with AI violation ……………………………………………………………………………….. 259
36 Word-by-word reading for the ungrammatical wh-sentences with complex NP/DP island violation ………………………………………………………………. 263

ABBREVIATIONS

ACC	accusative
AFS	active filler strategy
AI	adjunct island
ARSPAN	automated reading span
AOSPAN	automated operation span
CED	condition on extraction domain
CNPI	complex noun phrase island
DAH	direct association hypothesis
DAT	dative
ECP	empty category principle
GEN	genitive
L1	first language
L2	second language
LF	logical form
LOC	locative
LTM	long term memory
MCP	minimal chain principle
NOM	nominalizer
OEFF	object extraction from finite clause
OEFFT	object extraction from finite clause with 'that'
OEFNONF	object extraction from nonfinite clause
OGJT	online grammatical judgment task
PF	phonological form
POSS	possessive
PRES	present
PROG	progressive
RCI	relative clause island
RT	reading time
SEFF	subject extraction from finite clause
SEFNONF	subject extraction from nonfinite clause
SI	subject island
SG	singular
SR	single resource
SSIR	separate sentence interpretation resource
STM	short term memory
TT	*that*-trace
WM	working memory

CHAPTER 1

INTRODUCTION

Second language (L2) acquisition of wh-movement has been the focus of a considerable amount of research within generative framework since constraints on wh-movement has been a good testing ground to examine whether or not L2 grammar is constrained by Universal Grammar (UG), which is assumed to guide first language (L1) acquisition. Many L2 studies examined the knowledge of Subjacency Principle (or island constraints) in English in the grammars of adult L2 learners with a wh-in-situ L1 like Chinese, or Japanese on the basis of their judgments of grammatical as well as ungrammatical wh-constructions. The lack of overt movement in the L1 ensures that L2 learners cannot rely on the L1 to construct constraints on wh-movement in their L2. The results of some studies show that L2 learners are not able to reject subjacency violation, suggesting that the UG is not available in adult L2 acquisition (e.g., Schachter, 1989; Schachter & Yip, 1990; Johnson & Newport, 1991) while others show that the L2 learners are more or less as good as native speakers in rejecting subjacency violation (e.g., White & Juffs, 1998, Juffs & Harrington, 1995).

However, ongoing changes in the definition of island constraints on movement within the generative grammar framework (i.e., Transformational Generative Grammar (Chomsky, 1957), Government and Binding (GB) (Chomsky, 1981), Minimalist Program (MP) (Chomsky, 1995), Phases (Chomsky, 2000) have challenged the idea of subjacency condition as a good testing case to examine whether L2 grammar is constrained by the principle of UG. For example, parametric differences in bounding nodes with some languages like Italian and Spanish having Complementizer Phrase (CP) as a bounding node instead of Tense Phrase (TP) (Rizzi, 1982), and

inconsistencies found across different types of subjacency violations, and finally Huang's (1982) attempt to unify to island constraints under the Condition on Extraction Domain (CED) paved the way for the reformulation of bounding nodes, which were originally analyzed as barriers in which movement cannot cross a barrier, where an XP is a barrier iff it is not in a complement position (Chomsky, 1986; 1995; 2005; Cinque, 1990; Manzini, 1992). Within the CED, subjects and adjuncts are universal islands for movement because they are in non-complement position, whereas arguments are not islands because they are in complement position (e.g. Müller, 2007; Uriagereka, 1999).

Furthermore, recent analyses in Turkish, a wh-in-situ language with free word order, have demonstrated that leftward scrambling in Turkish exhibits an argument-adjunct asymmetry similar to those attested in Chinese and Japanese. Arguments are not islands for leftward movement via scrambling, but adjuncts are (i.e., Ikizoglu, 2007). Note, however, that in Turkish, subjects marked with genitive case can move out of islands (i.e., relative clause islands, and sentential subject islands), but non-subject constituents cannot move. Like non-subject constituents, subjects which are not marked for genitive case cannot move out of adjunct islands (e.g., Aygen, 2002; İkizoglu, 2007). Moreover, adjuncts marked for one of the dative, ablative, locative, or instrumental /comitative cases can move out of islands but bear adjuncts cannot (i.e., İkizoglu, 2007). Based on these observations, it has been suggested that island constraints on movement have more to do with the nature (i.e., cases) of the extracted element rather than the position it is extracted from (i.e., scrambling over a constituent with a same case is not allowed (Meral, 2010).

Thus, Turkish provides a good testing ground to verify the role of L1 in the L2 acquisition of wh-constraints in an overt wh-movement language like English. Within this background, this thesis explores whether adult L2 English learners with different L1 backgrounds—one with overt
wh-movement like English (i.e. Spanish) and the other one with wh-movement via scrambling, which is subject to island constraints (i.e. Turkish) can achieve native-like success in processing long distance wh-constructions in English in terms of accuracy and response latency.

L1 sentence processing theories can be classified into two large categories: (1) modular (two-stage) theories, and (2) interactive (constraint-based/satisfaction) theories. Among the modular theories, the Garden-path Theory proposed by Frazier and her colleagues (e.g., Frazier & Fodor, 1978; Rayner, Carlson & Frazier, 1983; Ferreira & Clifton, 1986; Frazier, 1987) has been the most influential one. The basic assumption of this model for ambiguity resolution during initial processing is that the human sentence processor (or parser) is serial, retaining exactly one analysis at the initial stage of parsing and it is modular, as it uses syntactic information before it uses other information in resolving ambiguity. The parser eventually applies non-syntactic information, such as semantic or pragmatic information or discourse context to resolve ambiguity. Frazier suggests two processing strategies (i.e., *Minimal Attachment*, and *Late Closure*, both of which were basically motivated in terms of reducing memory load), that guide the parser when it encounters an ambiguity. The former requires attachment of an incoming material into the phrase being constructed using the fewest nodes consistent with rules of phrase structure, whereas the latter requires that when possible, the parser attaches an incoming material into clause or phrase currently being parsed.

In contrast to modular theories, interactive (constraint-based/ satisfaction) theories assume that all potentially relevant sources of information, such as syntactic and semantic information, discourse context, and lexical frequency, are used immediately and simultaneously during the initial stage of sentence processing (e.g., MacDonald, Pearlmutter, & Seidenberg, 1994; McRae, Spivey-Knowlton, & Tanenhaus, 1998; Trueswell Tanenhaus & Garnsey, 1994; Trueswell, 1996; Garnsey, Pearlmutter, Myers, & Lotocky, 1997).

Most constraint-based/satisfaction theories have a lexicalist approach to sentence processing in that lexical representations contain syntactically relevant information (i.e. verb-argument structure information lexical category information, and morphological information such as tense and number); therefore, many syntactic ambiguities are associated with lexical ambiguities and are resolved in a similar fashion as lexical ambiguities (e.g. MacDonald et al., 1994; Trueswell, 1996).

With respect to processing of wh-sentences, two accounts have been proposed in L1 sentence processing: (1) gap-based accounts, and (2) gapless accounts. The basic assumption of the gap-based accounts is that the parser uses gaps (i.e. traces) left behind by the moved wh-phrase (also known as *filler*) to form filler-gap dependencies. The two-well known filler-driven strategies are the Active Filler Strategy (AFS) of Frazier & Clifton (1989), and the Minimal Chain Principle (MCP) of De Vincenzi (1991). The former suggests that as soon a wh-filler has been identified, the parser ranks the option of assigning it to a gap above all other options. The latter, on the other hand, proposes to postulate required chain members at the earliest point grammatically possible but postulate no potentially unnecessary chain members. The MCP manages to unify the processing of distinct types of empty categories such as ‘pro’ the null pronominal which occurs in case-marked positions, ‘NP-trace’ and wh-trace. This allows the processing of null subject languages like Italian to be unified with the processing of overt subject languages like English (De Vincenzi, 1991).

Gapless accounts (e.g. Pickering & Barry, 1991) propose a direct association hypothesis for wh-dependency formation in which the parser directly associates the filler with its subcategorizer without making use of gaps.

All of the theories in L1 sentence processing have attributed an important role to working memory (WM) resources, and explain how and when differences in WM resources interact with task demands and affect speed and accuracy in comprehension. Currently, WM is defined as a multi-component system responsible for active maintenance of information in the face of ongoing processing and/or distraction. Active maintenance of information is the result of converging processes, most notably, domain-specific storage and rehearsal and a domain general central executive (Baddeley & Hitch, 1974). In language comprehension, this model predicts that structures that exceed some level of complexity or require maintenance of too many unattached elements will be difficult for all language comprehenders. Two accounts have been proposed to explain the role of WM resources in syntactic processing. One of them is the single-resource (SR) account (e.g. Just & Carpenter, 1992), which suggests a single pool of WM resources shared by maintaining information in an active state and executing processes that manipulate information. When a task demand exceeds the available resources, the resources can be used either to maintain information or to execute processes. Individuals may vary with respect to the size of the shared pool of WM resources. Therefore, comprehenders with greater WM resources are expected to process sentences that require high memory demands more rapidly and accurately, compared to those with low WM resources.

Another account of WM is the separate sentence interpretation resource (SSIR) account (e.g., Waters & Caplan, 1996), which assumes, within the verbal WM, a separate pool of WM resources which are dedicated to syntactic processing only. Therefore, syntactic processing is assumed to be insensitive to WM limitations. For this reason, this account does not expect a difference between high- and low-span comprehenders in processing syntactically complex, or difficult structures.

Research in L2 sentence processing has only recently gained momentum with an aim to examine processing mechanisms and strategies used in L2 grammar building. Thus, there is an increasing number of L2 studies that use online behavioral measures (e.g. self-paced reading and eye-tracking, cross-modal priming) and neurophysiological measures (e.g., event-related potential (ERP), functional magnetic resonance imaging (fMRI), electroencephalography (EEG), and positron emission tomography (PET)) to examine L2 learners' processing of syntactically ambiguous structures (i.e., main verb/reduced relative ambiguities (e.g., Juffs, 1998); relative clause and PP attachment ambiguities involving complex *genitive* (NP-of-NP) antecedents (e.g., Frenck-Mestre and Pynte, 1997; Frenck-Mestre, 2002; Dussias, 2003; Felser et al., 2003; Papadopoulou and Clahsen, 2003); and direct object/sentential complement ambiguities (e.g., Juffs and

Harrington, 1996; Juffs, 1998a; 1998b; Felser and Roberts, 2004; and Juffs, 2004; Dussias & Cramer Scaltz, 2008).

Some of the major issues that L2 researchers have investigated involve the architectures, mechanisms and representations that underlie L2 processing, the type of processing strategies and information sources used by L2 learners in online processing of L2 input, and the differences and similarities between native and nonnative sentence processing. However, these studies revealed divergent results as regards processing strategies and information sources that L2 learners use in online L2 ambiguity resolution (see Clahsen and Felser, 2006 for an overview).

With respect to L2 studies examining processing of wh-dependencies, there are some pioneering studies that need to be noted here to set up the background of the present investigation. The first study was conducted by Juffs and Harrington (1995; 1996) to examine response accuracy and latency of Chinese learners of L2 English in processing grammatical and ungrammatical wh-extractions. The researchers attempted to identify whether native and nonnative differences are due to a processing problem or due to a competence deficit in ultimate attainment in L2 acquisition. Juffs (2005), in a follow-up study, tried to establish clearly the role of L1 by looking at L2 English data of learners with L1 Japanese, Korean Chinese and Spanish in online processing of wh-questions. However the results of these studies neither revealed clear L1 influence nor a strong subject-object asymmetry in processing finite and nonfinite clauses. In the same vein, Marinis, Roberts, Felser & Clahsen (2005) Felser & Roberts (2007) reported that unlike native speakers, adult L2 learners did not make use of gaps in the formation of filler-gap dependencies in real-time processing, which suggests that L2 learners do not process wh-dependencies in the same way as native speakers.

The present study aims to contribute to L2 sentence processing literature by examining Turkish and Spanish end-state L2 speakers' online processing of wh-dependencies in English. The study was designed and conducted to explore several issues. One of the main aims of this study is to explore whether end-state adult L2 learners can achieve native norms in terms of accuracy and speed in processing wh-extractions in the L2 English. Secondly, the study aims to examine whether the L1 (i.e. Turkish and Spanish) still plays a role in end-state L2 processing. Furthermore, the study also attempts to find out whether the subject-object asymmetry previously reported in processing wh-extractions from finite and nonfinite clauses is also observed in this investigation. Finally, the present study explores the role of WM resources in L2 sentence processing. Previously, a few researchers looked at the differences between high-and low span L2 learners in L2 processing (Juffs, 2004; 2005; ; Roberts, Marinis, Felser, and Clahsen, 2006; Felser and Roberts, 2007; Havik et al. 2009), However, possibly due to methodological issues, no conclusive finding was obtained to establish clearly the relationship between the WM capacity and efficiency in L2 processing. Therefore, this study attempts to explore the role of the WM capacity in online processing of wh-extractions through more reliable and valid WM tasks in the L2 as well as the L1, measuring the WM with both a reading and an operation span test.

The rest of the thesis is organized as follows: Chapter 2 presents syntactic background of wh-movement in English, Spanish and Turkish. Chapter 3 gives an overview of literature in L1 and L2 sentence processing in general and wh-processing in particular. The role of the WM capacity in L1 and L2 processing will also be addressed separately in this chapter. Chapter 4 presents methodology, including sections on research questions, participants, materials, procedure of the study. Chapter 5 reports accuracy and response/reading time (RT) results from five types of grammatical and ungrammatical wh-extractions presented in the full-sentence and the self-paced word-by-word reading in moving window condition. Also, it includes the results from the two WM tasks (ARSAPN and AOSPAN) in English and the results of L1 ARSPAN in Spanish and Turkish. Chapter 6 includes discussion, conclusion, and limitations of this study with suggestions for further studies.

CHAPTER 2

SYNTACTIC BACKGROUND

2.1. Wh-movement: Introduction

Chomsky's idea of Universal Grammar (UG) is concerned with general properties of language rather than the idiosyncrasies of a particular language. It is a theory of knowledge that deals with the internal structure of the human mind-how the computational system links sounds to meaning (Cook, & Newson, 2007: 11). Starting from 1957, Chomsky's UG theory has been continuously developed and undergone changes under different models (e.g., Transformational Generative Grammar (Chomsky, 1957), Government and Binding (GB) (Chomsky, 1981), Minimalist Program (MP) (Chomsky, 1995), Phases (Chomsky, 2000)). Since the introduction of the GB Theory, it has been claimed that linguistic knowledge consists of principles and parameters (P&P). Principles are believed to apply invariantly to all languages and parameters are believed to have settings that vary among languages. Acquiring language means learning how these principles apply to a particular language and which value is appropriate for each parameter for that language (Cook, & Newson, 2007:11). The basic concepts of the P&P approach remained constant across all syntactic theories developed after 1980.

One type of parametric variation across languages is the wh-parameter, which determines whether or not a wh-phrase can be moved to the front of a main interrogative structure containing it. This parameter appears to be binary in nature, thus, it allows for only two possibilities; a language either does or doesn't allow overt wh-movement. For example, languages like English allow overt wh-movement, in which wh-phrases can be moved to the Spec-CP position, leaving a trace in their base-positions, whereas, languages like Turkish, Japanese and Chinese do not allow wh-phrases to move to the Spec-CP position at overt syntax.

In the following section, I will first discuss how overt and covert wh-movements work in languages by discussing examples from English in comparison to Chinese and Japanese two wh-in-situ languages, which are studied extensively in this context. I will then present some arguments for the LF-movement approach. In section 2.4, I will discuss specific details about the island constraints and subjacency in English. This will be followed by a discussion of the same features in Spanish and finally the characteristics of wh-movement in Turkish will be presented.

2.2. Overt movement

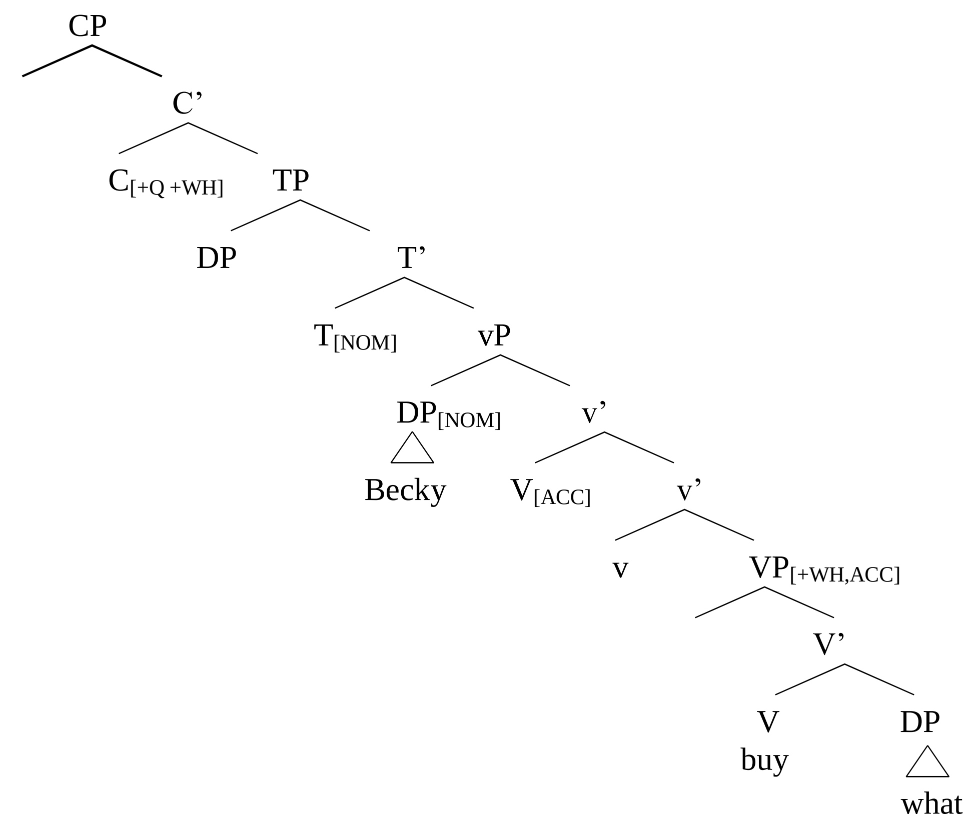
Within the framework of the generative grammar, languages like English and Spanish are considered to exhibit overt wh-movement in the formation of wh-questions. According to the GB Theory (Chomsky, 1981), wh-questions are formed by moving wh-phrases from their base position at the deep structure (d-structure) to the specifier of the Complementizer Phrase [Spec-CP] at the surface structure (s-structure) by a single transformational movement rule, *Move-alpha.* A moved wh-phrase leaves behind (in the position out of which they move) an empty trace of itself. The moved wh-phrase is the antecedent of its empty category (i.e., the trace) and serves to bind the trace. In addition, the wh-phrase and its trace together form a movement chain. The binding relation between a trace and its antecedent is marked by attaching identical subscript letter called indices to them as in (1a-b[1]).

1. a. Becky bought the syntax book.

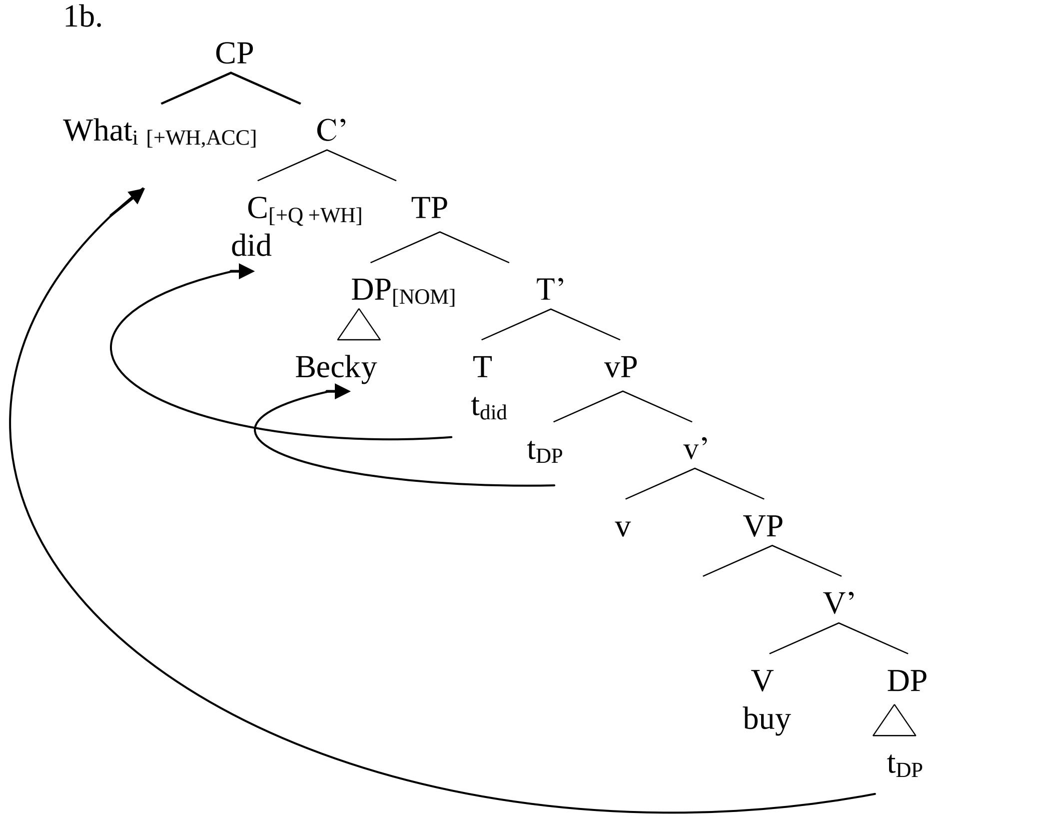
   b. What did Becky buy?

      [$What_i$ did Becky buy t_i]? (Carnie, 2007: p.318)

1a.


CP
C’
C[+Q +WH]
TP
DP
T’
T[NOM]
vP
DP[NOM]
Becky
v’
V[ACC]
v’
v
VP[+WH,ACC]
V’
V
buy
DP
what

1b.


CP
Whati [+WH,ACC]
C’
C[+Q +WH]
did
TP
DP[NOM]
Becky
T’
T
tdid
vP
tDP
v’
v
VP
V’
V
buy
DP
tDP

Becky and *what* both get their theta roles in the d-structure positions. '*What'* also gets its Case in this base position. Three other operations apply: There is DP movement of *Becky* to the specifier of TP to check [NOM] feature. There is insertion of *do* to support the *–ed* and there is also a T-to-C movement to fill the null [+Q] complementizer.

The motivation behind the overt wh-movement in the GB Model is that the head of the interrogative sentence, the head C, has a [+wh] feature. A wh-phrase which has also a [+wh] feature undergoes movement in order to be near the [+wh] feature in the C. In other words, a wh-phrase with a [+wh] feature moves up to the Spec-CP to agree with the [+wh] feature on C. Therefore, wh-movement is obligatory to the front of an interrogative clause (Cook & Newson, 2007).

Chomsky (1995) reformulated grammatical representations and their well-formedness within the Minimalist Program (MP). The MP reduces the set of four levels of representation d-structure, s-structure, LF, and Phonological Form (PF) of standard GB theory to the two *interface levels*: LF and PF. LF interfaces with semantic-conceptual systems of cognition and PF is connected to articulatory–perceptual modules (Marantz, 2006:351). The phonetic, grammatical, and semantic properties of words are described in terms of sets of features. PF representations should contain only phonetic features, and LF representations should contain only semantic features. This requirement is imposed by a UG constraint known as the principle of full interpretation (PFI), which states that a representation for a given expression must contain all and only elements which contribute directly to its interpretation at the relevant level. While a derivation which satisfies PFI converges, the one which does not, will not converge (Radford, 2000:171).

The derivation of a sentence involves the following steps: Lexical items, each of which has sets of phonetic, semantic, and grammatical features are taken from the lexicon by an operation of selection. By the process of merger, constituents are combined together in a pairwise fashion to form a phrase structure tree. After spell-out (where the phrase structure generated by the process of selection and where merger feeds into PF and LF components, the phonetic and semantic features of items are processed separately. The former is being processed by the PF operations which compute PF representations, and the latter being processed by LF operations which compute LF representations (Radford, 2000:171).

Grammatical features include number (singular/plural) features, gender (masculine/feminine/inanimate) features, person features, and features that determine the morphological form of items such as case feature and inflectional features. Some of these features (e.g., the person, gender, number features of pronouns like *she*) have semantic content and so are interpretable (at LF), while others (e.g., the case features of pronouns and inflectional features of nonfinite verbs) are uninterpretable and so must be erased in the course of the derivation in order to ensure they do not appear in LF representations (Radford, 2000).

Also, words carry three types of grammatical features: *head features*, which determine their intrinsic grammatical properties, *specifier features*, which determine the kinds of specifiers they allow, and *complement-features*, which determine the kinds of complements they take. Specifier- and complement-features (and those head-features which are purely formal and hence have no semantic content) are uninterpretable and so must be erased by a process of checking (Radford, 2000).

In the MP, a feature checking mechanism has been introduced to account for movement. Accordingly, derivations involve movement operations. Wh-movement is one of them, in which a wh-phrase moves from their base positions to Spec-CP, leaving a copy in its base position. The assumption is that the question affix, Q occupies the head C position of an interrogative CP and it carries an interrogative specifier-feature. Wh-operators like *who* carry an interrogative head-feature and they move to Spec-CP in order to check the interrogative specifier-head feature carried by Q. Since wh-questions in English generally begin with *wh*, [wh] is used to refer to the interrogative feature. The [wh] specifier-feature of Q is checked by the [wh] head-feature carried by wh-phrases, and thereby erased, since specifier-features are uninterpretable at LF. However, the [*wh*] head-feature carried by *who* is not erased since it has a role at LF in identifying *who* as an interrogative operator (Radford, 2000: p. 273).

Within this framework, overt and covert movements have been explained by the deterministic role of strong and weak features. Some features are strong: they must be checked off before PF, where they are intolerable. Weak features, on the other hand, can survive at PF without causing a crash; therefore, they will be only checked off after Spell-out, in accordance with Procrastinate, which favors the late application of any process. In this view, linguistic parameterization depends on whether a given feature is strong or weak in a language. If the language decides a feature is strong, this will be associated with overt movement of the element bearing the feature, whereas if the language has a weak feature, the language will display covert movement (Cook &, Newson, 2007).

2.3. Covert movement

Within generative grammar theory, wh-phrases are characterized as quantifiers. Like ordinary quantificational NPs, they are non-referential. In standard semantic treatments, they are represented in quantificational schemas suitable for interpretation as in (3). In the GB, wh-phrases are operators binding variables at LF, like other quantificational NPs.

2. Who_i did John see t_i ?

3. (Which x: x is a person) (did John see x)?

Huang (1982) was the first who proposed that in-situ wh-phrases are also quantifiers. As quantifiers undergo movement at LF, wh-phrases having quantificational properties undergo movement to their scope position at LF as in (4-5). The wh-word *shenme* in (4) stays in-situ in at overt syntax, but it obligatorily undergoes a raising process at LF after mapping to PF to produce the LF representation in (5) (Watanabe, 2003: 203).

4. Zhangsan xiang-zhidao [Lisi mai-le shenme] (PF)
   Zhangsan wonder [Lisi bought what]
   "Zhangsan wonder what Lisi bought."

5. [$_{CP}$ Zhangsan xiang-zhidao [$_{CP}$ $shenme_i$ [$_{IP}$ Lisi mai-le t_i]]] (LF)
   Zhangsan wonder what Lisi bought

Evidence for LF (covert movement) comes from parallels in scope, selection, weak crossover effects, and locality effects between overt movement and covert movement at the LF representations. Below, I will briefly review each of these arguments.

2.3.1 Selectional requirement

Since in-situ wh-phrases in language like Chinese, Japanese and Korean undergo movement at LF, the selectional requirement of the verbs are satisfied in the same way as in English. In English the scope of a wh-phrase is determined by overtly moving the wh-element to the Spec-CP. For example, the wh-phrase has matrix scope in (6a and 8a) and embedded scope in (7b and 8b) (Huang, 1995: 149).

6. a. What does John think Mary bought t?

   b. *John thinks what Mary bought t?

7. a. *What does John wonder Mary bought t?

   b. John wonders what Mary bought t.

8. a. What does John remember Mary bought t?

   b. John remembers what Mary bought t.

The difference in grammaticality among sentences in (6-8) is attributed to the selectional properties of the matrix verbs: *think*-type verbs select declarative clauses, *wonder*-type verbs select questions, and *remember*-type verbs select either, as their complements. The same scope interpretation has been observed in the corresponding cases of Chinese where in-situ wh-phrases are assumed to move to Spec-CP at LF as in (9 and 11).

9. Zhangsan yiwei Lisi mai-le shenme?

   Zhangsan think Lisi bought what

   "What does Zhangsan think Lisi bought?"

10. Zhangsan xiang-zhidao Lisi mai-le shenme?

   Zhangsan wonder Lisi bought what

   "Zhangsan wonders what Lisi bought"

11. Zhangsan jide Lisi mai-le shenme?

Zhangsan remember Lisi bought what

"Zhangsan remembers what Lisi bought"

"What does Zhangsan remember Lisi bought?"

In sentence (9), *shenme* 'what' has the matrix scope as a direct question, while in (10) it has embedded scope as a statement containing embedded question, and in (11) it has either. These restrictions are the same restrictions observed with the English sentences (6-8). The only difference is that whereas the restrictions are observed as a matter of *form* (i.e. grammaticality) in English, they seem to be a matter of *interpretation* (e.g., presence or absence of ambiguity) in Chinese. If wh-phrases move to the Spec-CP at LF as they do in overt syntax, the following structures (12-14) may be derived from (9-11).

12. a. [shenme_i [Zhangsan yiwei [[Lisi mai-le t_i]]]]

for which x: x is thing, Zhangsan thinks Lisi bought x

b. *[[Zhangsan yiwei [shenme_i [Lisi mai-le t_i]]]]

Zhangsan thinks [for which x; x is thing, Lisi bought x]

13. a. *[shenme_i [Zhangsan xiang-zhidao [[Lisi mai-le t_i]]]]

for which x: x is thing, Zhangsan wonders Lisi bought x

b. [[Zhangsan xiang-zhidao [shenme_i [Lisi mai-le t_i]]]]

Zhangsan wonders [for which x; x is thing, Lisi bought x]

14. a. [shenme_i [Zhangsan jide [[Lisi mai-le t_i]]]]

for which x: x is thing, Zhangsan remembers Lisi bought x

b [[Zhangsan jide [shenme_i [Lisi mai-le t_i]]]]

Zhangsan remembers [for which x; x is thing, Lisi bought x]

The assumption is that the selectional restrictions that account for (6-8) apply also at the level of LF, therefore, of these sentences above, (12b) and (13a) are ruled out as ill-formed LF structures. This leaves (12a), (13b), and (14a-b) as well-formed, representing the only possible interpretations of (9-11). This suggests that the selectional restrictions are applicable to English-type languages as well as to Chinese type languages. The two types of languages simply differ in whether wh-movement takes place in overt syntax or at LF.

2.3.2 Scope taking properties

Another evidence for LF movement approach comes from scope-taking properties of in-situ wh-phrases similar to quantifier NPs at LF. The multiple wh-questions in (15) in English can be answered by (15a) or (15b) (Huang, 1995, p. 151).

15. Who remembers where we bought what?

a. John does. John remembers where we bought what.

b. John remembers where we bought the pencils, and Mary remembers where we bought the pens

The ambiguity is one of scope, and it arises, under the LF movement hypothesis due to the possibility of moving the in-situ wh-phrase what to the matrix or to the embedded Spec-CP. (15a) is an appropriate answer to (15) as a singular question containing an embedded multiple questions, where the matrix wh-phrase ranges over individuals and the embedded wh-phrase ranges over pairings of places and things. In contrast, (15b) is an answer in which *what* is answered, indicating wide-scope of the in-situ wh-phrases. The wide-scope reading of the in-situ wh-phrase in (13c) can be accounted for if the in-situ wh-phrase undergoes movement to the matrix Spec-CP, yielding a multiple question associated with the matrix wh-phrases.

This property has also been observed in Mandarin Chinese as in (16) (Cheng, 2009, p.771).

16. Mei-ge xuesheng dou xuanle na yi-men ke?
every-CL student all choose-PREF which one-CL course
"Which course did every student choose?"

In (16), the wh-phrase takes the scope over the universal quantifier. The interpretation of the question is that: *which course is such that every student chooses it*. This shows that though the wh-phrase is in situ, it takes wide scope.

2.3.3 Weak crossover effects

Another parallel between overt wh-movement and covert wh-movement is the manifestation of the Weak Crossover effect (WCO; Chomsky. 1976) which states that a variable cannot be antecedent to a pronoun to its left. Overt wh-movement shows weak crossover violation as shown in (17).

17. *Who$_i$ does his$_i$ mother like t$_i$?

In (17), *who*, which is coindexed with *his* and the base trace position is moved to the Spec-CP position causing the ungrammaticality in the sentence. Similarly, covert movement shows weak crossover violations like quantifiers at LF as shown in (18) (Aoun & Li,, 1993, p.201)

18. a. *Xihuan ta$_i$ de ren kandao shei$_i$?
like he DE man saw who
"Who did the person that likes him see?"

In (18) the in-situ wh-phrase *shei* 'who' in Chinese, which is coindexed with *ta* 'he' and the base trace position is moved to the Spec CP at LF as in (19).

19. a. [shei$_i$ [[[xihuan ta$_i$ de] ren] kandao t$_i$]]

who like he DE man saw

The ungrammaticality of (18) can be accounted for by the weak crossover violation at LF (19) and suggests that in-situ wh-phrases undergo movement at LF like quantifier phrases.

2.3.4. Locality effects

The most important evidence for LF movement comes from the fact that the interpretation of in-situ wh-phrases is subject to syntactic constraints on overt wh-movement. One of these constraints is the Empty Category Principle (ECP; Chomsky, 1981; Lasnik & Saito, 1992) which requires a trace to be properly governed, i.e. either lexically governed or antecedent-governed. A complement to a lexical category is lexically governed, but a subject is not, so a subject trace needs to be antecedent-governed, but object trace need not to. Therefore, long-distance subject extractions yields ungrammatical questions, but long-distance object extractions do not. In other words, the ECP allows arguments to be raised out of islands but not adjuncts as in (20a-b) and (21a-b), respectively (Lasnik & Saito, 1992).

20. a. ??What do you wonder whether John bought?

b. ?*Who$_i$ do you believe the claim that John said Mary saw t$_i$?

21. a. *Why$_i$ do you wonder whether John left t$_i$?

b. *Why$_i$ do you believe the claim that John said Bill left t$_i$?

The traces in adjunct extractions (21 a and b) are not lexically governed either in VP or NP, so they must be antecedent-governed and thus they cannot be moved out of the wh-islands or complex NP islands. The same pattern of contrast can also be found with wh-in-situ multiple questions in English-type languages. That is, we cannot leave wh-adjuncts in-situ in an island, as in (22 a-b), but wh-arguments can stay in-situ as in (23 a-b).

22. a. *Who wonders whether John left why?

b. *Who believes the claim that John said Bill left why?

23. a. ??Who wonders whether John bought what?

b. ??Who believe the claim that John said Mary saw who?

At LF, the sentences (22a–b) are excluded by the ECP because the LF-created traces of *why* fail to be properly governed.

The same argument-adjunct asymmetry has been observed in the movement of in-situ wh-phrases at LF in languages like Chinese and Japanese (Huang, 1982). Arguments can be extracted out of islands, but subject and adjuncts cannot. Huang (1982) makes an influential proposal that extractions out of subjects and adjuncts are violations of Condition on Extraction Domain (CED). The CED states that 'A phrase A may be extracted out of a domain B only if B is properly governed' (Huang, 1982). For example, (24) in Chinese can be interpreted as an indirect question (24a) or as a matrix clause (with an indirect question as well) (24b). However, though it can be interpreted as a matrix question, it can only be a matrix question asking for which person, but not for which reason (as can be seen from the ungrammaticality of (24c). Example (25) is a typical (ungrammatical example of an adjunct in-situ in a complex NP island (Cheng, 2009: p. 769).

24. Hufei xiang-zhidao shei weishenme sehngqi (?)

Hufei want to know who why get-angry

a. Hufei wonders who gets angry why.

b. ‘for which x, x a person, Hufei wonder why x gets angry’

c. ‘*what is the reason x, Hufei wonders who gets angry for x’.

25. *Qiaofeng xihuan Botong weishenme xie de shu?

Qiaofeng like Botong why write the book

‘For what reason x such that Qiaofong likes the book that Botong wrote for x.’

The same argument-adjunct asymmetry holds for other island conditions such as sentential subject island and adjunct islands (see Lasnik & Saito, 1992). According to Nishiaguchi (1990), Choe (1987), and Pesetsky (1987), LF movement is entirely parallel to overt movement, obeying the same constraints. In contrast to Huang, they claim that LF movement is also subject to subjacency. They account for the argument and adjunct asymmetry by the Pied-Piping hypothesis in which they propose that what undergoes LF movement in the Japanese question in (26) is not in-situ wh-phrase *dare,* but that the entire complex NP [dare-ga kaita] *hon*,which is pied-piped as in (27) (Watanabe, 2003: 205).

26. kare-wa [dare-ga kaita] hon-o yonde-iru no?

He-TOP who-NOM wrote book-ACC read-PROG Q

“Is he reading a book that who wrote?”

27. [[dare-ga kaita] hon_i-o [kare-wa $t_i$yonde-iru] no] (LF epresentation)

Who-NOM wrote book-ACC he-TOP read-PROG Q

Since the movement of the complex NP does not cross an island, (26) is acceptable. In the case of adjuncts, pied-piping is not possible with certain adjuncts like *naze* "why" as in (27).

27. *kare-wa [John-ga naze kaita] hon-o yonde-iru no?

he-TOP John-NOM why wrote book-ACC read-PROG Q

"Is he reading a book that John wrote why?

Although the strong unacceptability of (27) is attributed to the subjacency effect, it can also be attributed to the ECP violation.

In sum, this section presented main arguments for the proposals that in-situ wh-phrases in languages like Chinese, Japanese and Korean undergo movement at LF, providing evidence from the parallelism between overt movement and covert movement at LF representations. The parallels between overt and covert movement show that in-situ wh-phrases can take wide and narrow scope, and exhibit selectional requirements of the verbs, weak crossover effects and locality effects at LF.

In the following section, I will discuss syntactic island constraints (e.g., Subjacency) that prevent movement out of certain constructions known as islands such as complex NP island, relative clause island, adjunct island, subject island and the *that*-trace effect in English.

2.4. Island constraints and Subjacency in English

There are constrains on wh-movements imposed by Bounding Theory, which dates back to work of Chomsky (1964) and Ross (1967). Ross identified a number of constructions as islands, out of which wh-phrases cannot be extracted such as relative clauses, clausal adjuncts, wh-clauses, coordinate structures, sentential subjects. The movement out of these structures results in ungrammaticality, which has been called *island effects.* Chomsky (1973) offered a unified account for the island constraint, the Subjacency Principle. Subjacency defines the boundaries for movement and thus determines how far a wh-phrase can be moved from its base position. More specifically, it states that movement cannot cross more than one bounding node at a time, where bounding nodes are IPs and NPs in English. This principle has been revised within the framework of Barriers (Chomsky, 1986) in order to account for crosslinguistic differences in island phenomenon (e.g., parametric differences in bounding nodes (Rizzi, 1982)) and some shortcomings in explaining a number of island conditions. Also, Chomsky incorporates Huang's CED phenomena into the Barriers under the general principle of Subjacency. In his Barriers monograph, Chomsky has proposed that certain constructions become barriers to movement because of not what they are but because of where they sit in a structure. Just like the original CED, the Barriers account made the complement/noncomplement distinction. For example, complements are not barriers to movement due to their special property called L-marking- that is, theta-marking by a lexical head.(i.e., complements are related to lexical heads in that they are selected by them, thus they are L-marked). Whereas, noncomplement such as subjects and adjuncts are not L-marked, therefore they are barriers to movement. A barrier is an XP that is not a complement (Chomsky & Lasnik, 1993, p. 79). Within this framework, a distinction has been made between strong and weak islands via the number of barriers that has to be crossed to move out of an island: crossing one barrier results in a weak subjacency violation, whereas crossing two barriers or one barrier along with a violation of the ECP yields strong violation. According to this definition, subjects, adjuncts and relative clauses are strong islands, out of which movement results in strong violation. Weak islands, on the other hand, are complements of verbs or nouns (i.e., wh-islands, Complex NP islands) and movement out of these forms does not yield strong violation.

A simpler approach to boundedness, proposed by Rizzi (1990) is called Relativized Minimality, which claims that 'a moved constituent moves to the nearest appropriate position where what is an appropriate position is relative to the type of constituent being moved' (Radford 2000, p.526). In other words, all movements should be to the nearest relevant position. The relevant position here depends on the moved element. For example, if it is a head that is moving, then the nearest relevant position is the head position. If an argument is undergoing A-movement, then the nearest relevant position is the nearest A-position; and if an element is undergoing A'- movement, then the relevant movement is the nearest A'-bar position. Within the framework of the MP (Chomsky, 1995), the Relativized Minimality Principle has been interpreted under the Minimal Link Condition, which also favors shorter movements over longer ones, thus fitting with the idea of economy: the more distance covered by a movement, the costlier it is and so there is a pressure to keep the links between elements in a movement chain to a minimum (Cook & Newson, 2007).

Below, I will discuss four types of islands which are relevant for the present investigation, namely the complex NP islands, relative clause islands, adjunct islands and subject islands along with that-trace effect in English.

2.4.1. Complex NP Island (CNPI)

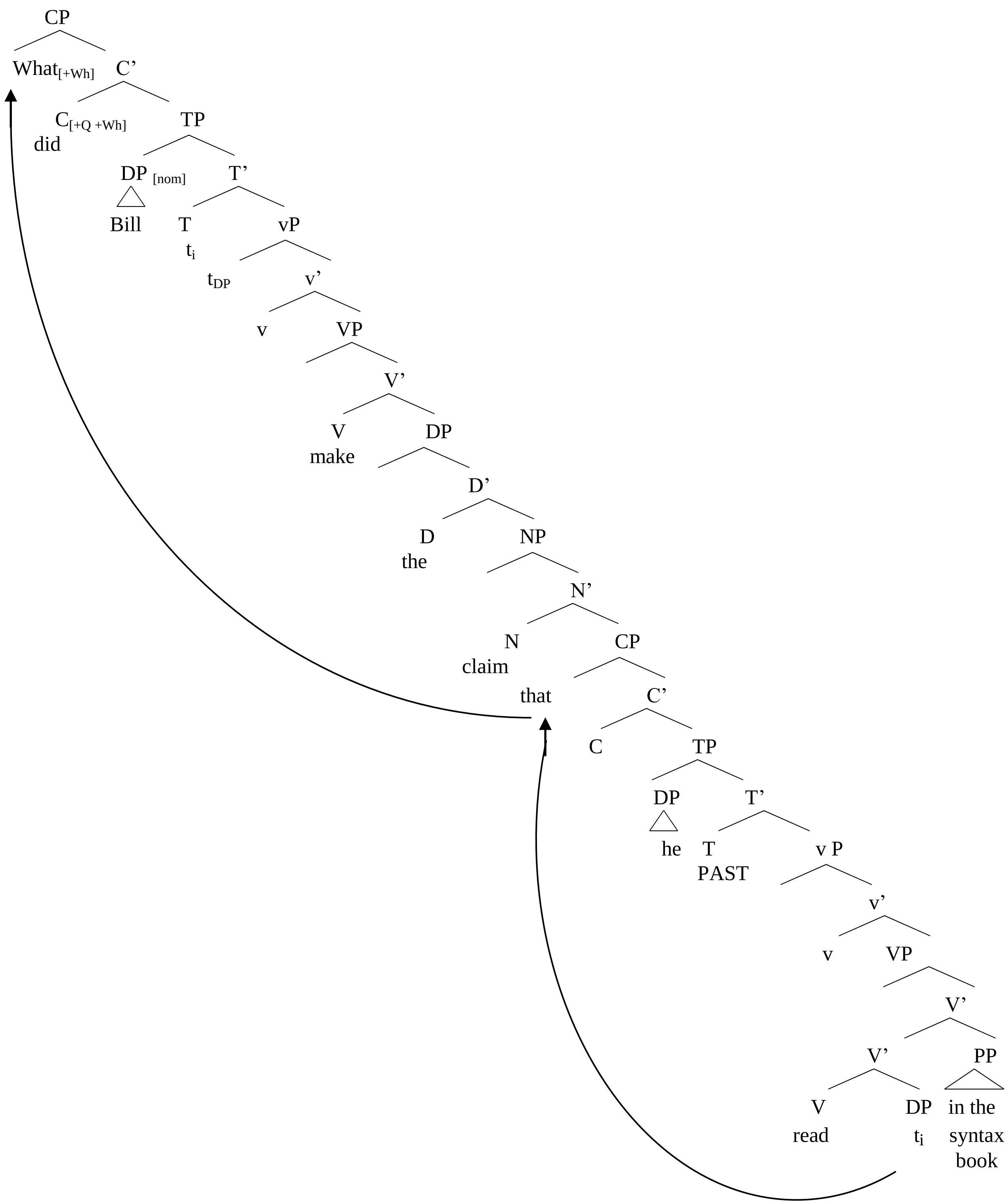
The complex NP/DP constraint states that a CP which is dominated by a NP is an island for movement. In other words, no element can be extracted out of a complex NP/DP (i.e., an NP/DP[1] consisting an N, and a complement clause. The sentence (28) illustrates complex NP[1] island (Carnie, 2007: 334).

28. a. *What did Bill make the claim that he read in the syntax book?

b. [$_{CP1}$ What$_i$ did [$_{TP1}$ Bill make [$_{DP}$ the claim [$_{CP2}$ t$_i$ that [$_{TP2}$ he read t$_i$ in the syntax book]]]]]?

The wh-phrase *what* is extracted out of a CP that is dominated by a DP. The head N (*claim*) takes a CP as a sentential complement According to Subjacency Principle, the ungrammaticality of (28) is due to long distance movement of the wh-word, *what* from its base position in the embedded clause to the Spec-CP of the main clause. This movement, crossing two bounding nodes (TP1 and DP) at a time violates subjacency and the Shortest Movement Principle, since the Spec-CP position of the main clause is not the nearest Spec-CP position above the (VP- complement) position in which *what* originates. The Spec-CP position in the complement clause occupied by *that* is the nearest Spec-CP position which blocks successive cyclic movement. The following tree diagram shows the violation of the Complex DP constraint (Carnie, 2007: 334).

[1] From now onward, I will use NP and IP as DP, and TP respectively.


CP
What[+Wh]
C'
C[+Q +Wh]
did
TP
DP [nom]
Bill
T'
T
ti
vP
tDP
v'
v
VP
V'
V
make
DP
D'
D
the
NP
N'
N
claim
CP
that
C'
C
TP
DP
he
T'
T
PAST
v P
v'
v
VP
V'
V'
V
read
DP
ti
PP
in the
syntax
book

2.4.2. Relative Clause Island (RCI)

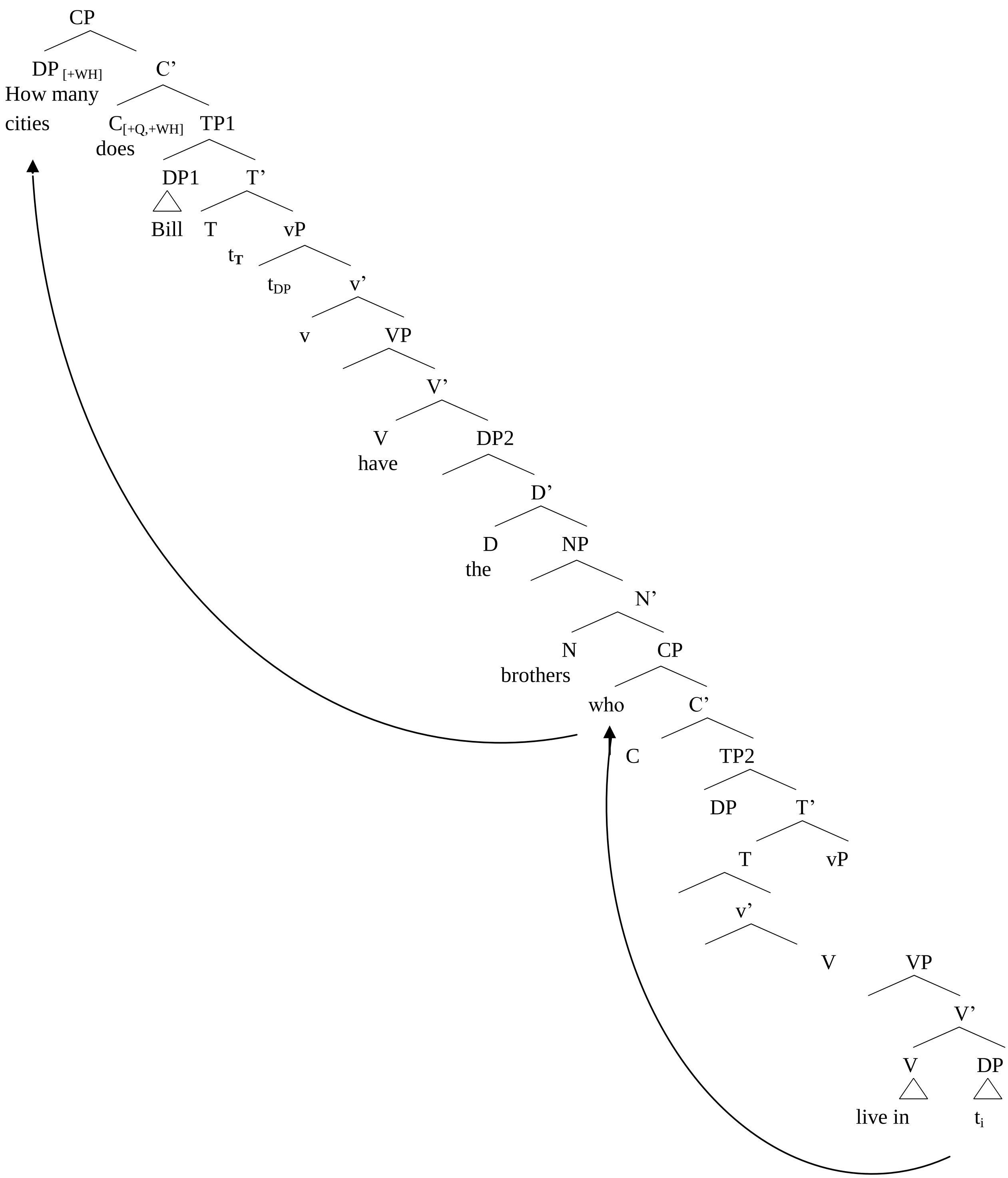
Like complex DPs, relative causes are islands for movements. Wh-phrases cannot move out of relative clause islands as in (29).

29. a. *How many cities does Bill have brothers who live in?

b. *[$_{CP1}$ How many cities does [$_{TP1}$ [$_{DP1}$ Bill have [$_{DP2}$ brothers [$_{CP2}$ t_i who [$_{TP2}$ live in

t_i]]]]]?

In sentence (29), the wh-phrase, *how many cities* moves out of a CP dominated by a DP. The CP is a relative clause (i.e., an adjunct to the N). This movement violates subjacency because the wh-word, *how many cities* crosses two bounding nodes (TP1 and DP2) in one step. It also violates the Shortest Movement Principle because the wh-word *how many cities* cannot move to the nearest CP because the nearest complement clause Spec-CP is filled by another wh-word *who*, which blocks successive cyclic movement. The following tree diagram shows relative clause island violation:


CP
DP [+WH]
How many
cities
C'
C[+Q,+WH]
does
TP1
DP1
Bill
T'
T
tT
vP
tDP
v'
v
VP
V'
V
have
DP2
D'
D
the
NP
N'
N
brothers
CP
who
C'
C
TP2
DP
T'
T
vP
v'
V
VP
V'
V
live in
DP
ti

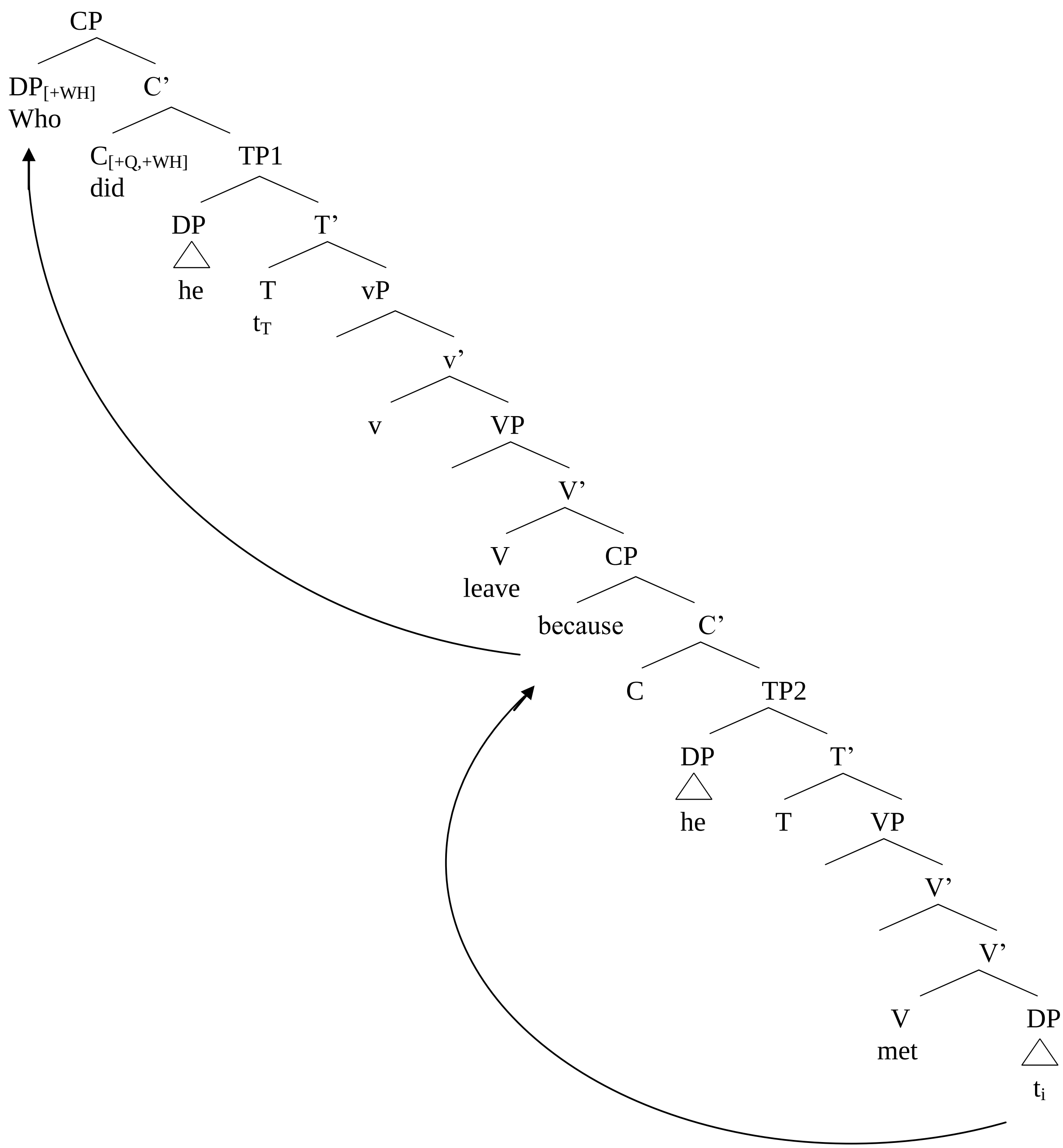
2.4.3. Adjunct Island (AI)

Adjuncts are another kind of Islands, out of which wh-phrases cannot be extracted as in (30) (Cook & Newson, 2007:143).

30. *Who did he leave because he met?

*[$_{CP1}$ Who did [$_{TP1}$ he leave [$_{CP2}$ t_i because [$_{TP2}$ he met t_i]]]]?

As in the previous constraints, this movement violates subjacency by long-distance movement of the wh-word *who* to the main clause Spec-CP, which takes two bounding nodes (TP2 and TP1) to cross in one step. It violates the Shortest Movement Principle in that the nearest complement clause Spec-CP is filled by *because*, which blocks successive cyclic movement of the wh-word, *who*. Therefore, *who* moves to the main clause Spec-CP. The following tree diagram shows the ungrammatical wh-movement out of an adjunct clause.

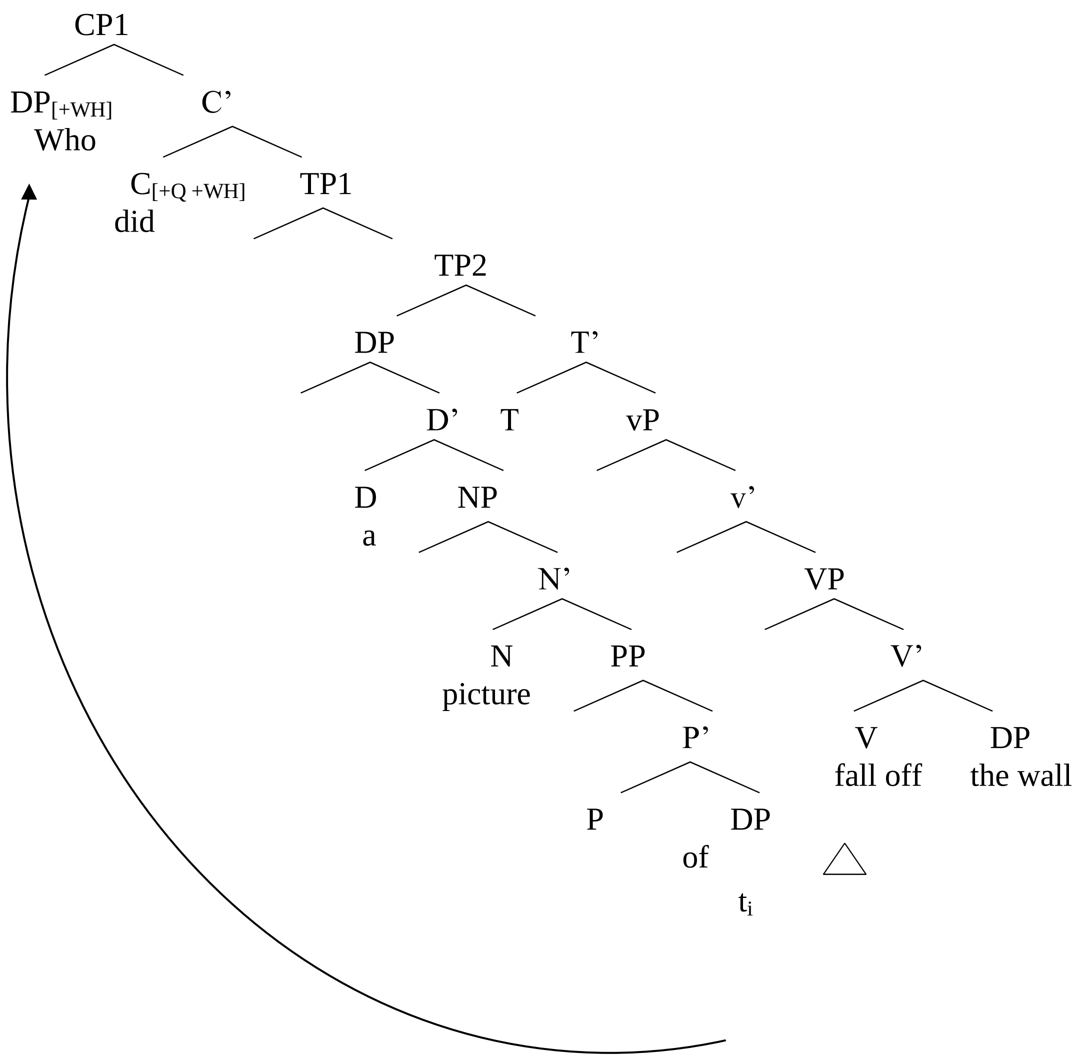

CP
DP[+WH]
Who
C'
C[+Q,+WH]
did
TP1
DP
he
T'
T
tT
vP
v'
v
VP
V'
V
leave
CP
because
C'
C
TP2
DP
he
T'
T
VP
V'
V'
V
met
DP
ti

2.4.4. Subject Island (SI)

According to the Subject Island Constraint, a wh-phrase cannot be extracted out of a subject island in the subject position as in (31). The movement crosses a DP and a TP at a time, which violates subjacency.

31. *Who did a picture of fall off the wall?

* [$_{CP1}$ Who$_i$ did [$_{TP1}$ [$_{DP}$ a picture [$_{PP}$ of t$_i$] fall off the wall]]?

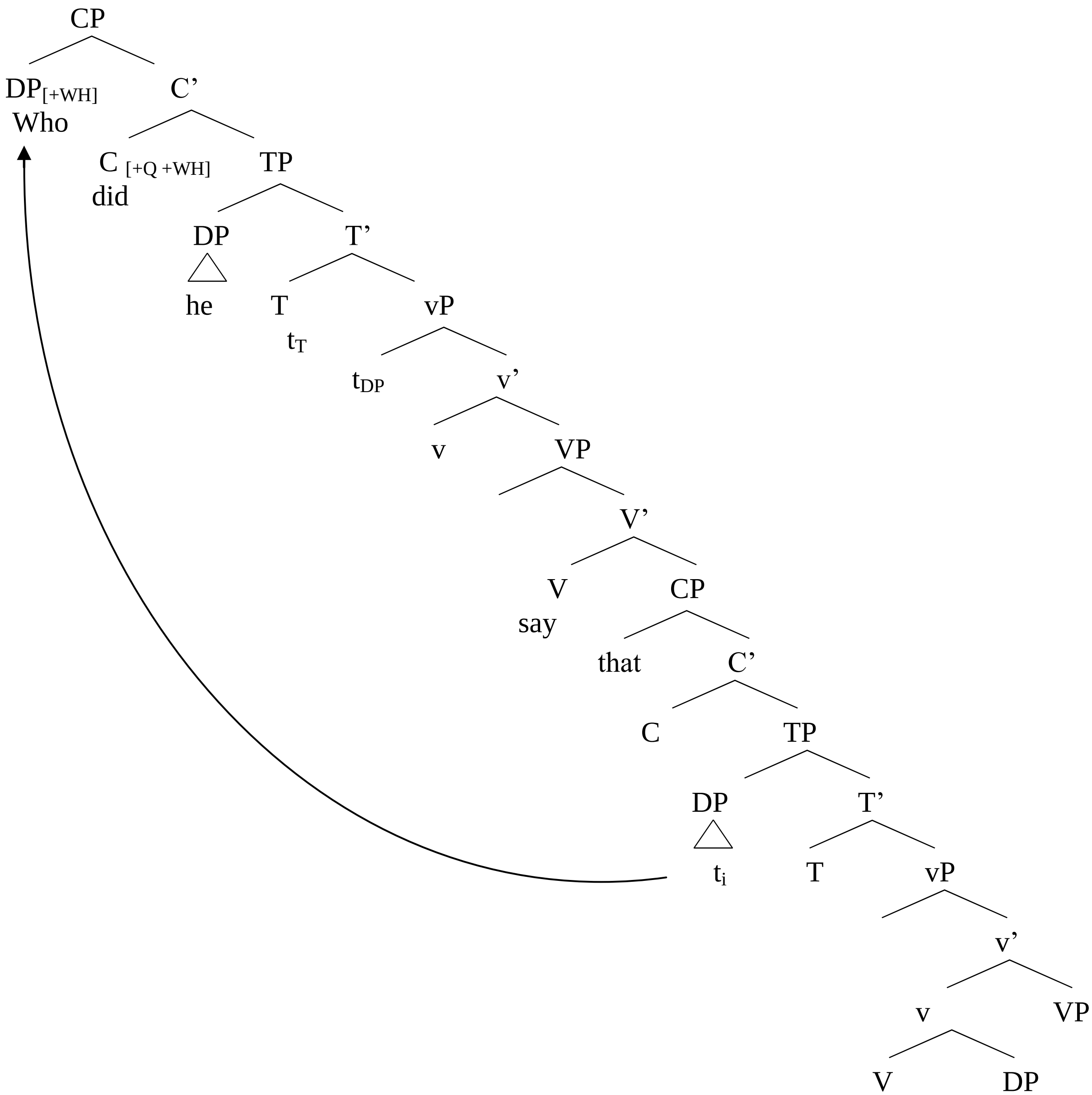


2.4.5. *That*–trace effect

Another constraint on wh-movement in English is the so-called *that*-trace effect. Long distance movement of subjects which are immediately preceded by an overt complementizer *that* is not grammatical. This phenomenon was first noted by Chomsky and Lasnik (1977) and termed as the *that-trace effect*. The *that*-trace phenomenon indicates that traces are licensed in the object position freely, whereas they are licensed in the subject position only in the absence of a complementizer in English. In GB theory, empty categories are licensed by a notion of proper government due to the ECP, which states that traces must be properly governed. Proper government is: 'α properly governs β if only if: (i) α governs β and (ii) α is lexical or an antecedent'. Traces are 'licensed' by being governed by either a lexical head or an antecedent. Thus, traces in object positions are always licensed as they are an object of lexical head. Traces in subject positions are never head-governed and therefore must be antecedent-governed. This gives rise to the that-trace effect as the antecedent government is blocked by the presence of an overt complementizer (Cook & Newson, 2007: 177). The question in (32) violates the ECP since the subject trace fails to be theta-governed: its theta role assigner, the verb is too low in the structure to govern it, and its antecedent, the intermediate trace is blocked from governing its subject trace by the intervening complementizer *that* (Cook & Newson, 2007:175)

32. *Who did he say that wanted a beer?

*[$_{CP}$ Who$_i$ did [$_{TP}$ he say [$_{CP}$ that [$_{TP}$ t$_i$ wanted a beer]]]]?



CP
DP[+WH]
Who
C’
C [+Q +WH]
did
TP
DP
he
T’
T
tT
vP
tDP
v’
v
VP
V’
V
say
CP
that
C’
C
TP
DP
ti
T’
T
vP
v’
v
VP
V
DP

2.5. Wh-movement in Spanish

Like English, Spanish exhibits overt wh-movement in wh-questions. That is, a wh-phrase moves from its canonical to sentence-initial position (i.e., Spec-CP), but the position of the verb is restricted in certain ways. The sentence-initial position of wh-phrases is illustrated in (33a-b) (Zagona, 2002, p.242):

33. Juan leyò ese libro. (Declarative sentence)

Juan read-PAST that book

$[_{CP}$ Juan $[_{TP}$ leyò ese libro]]

a. ¿Què libro leyò Juan? (Direct Question)

Which book read-PAST Juan

‘Which book did Juan read?’

b. María no sabe [qué libro leyó Juan]. (Indirect Question)

María not know-PRES [which book read-PAST Juan].

“Maria doesn’t know which book Juan read.”

To have an interrogative reading the wh-phrase *qué libro* “which book” moves from canonical object position to sentence initial position in the direct question (33a) and the indirect question (33b).

However, in Spanish, the subject-verb order must be inverted in wh-questions when the wh-element is an object. This inversion rule is called verb preposing (V-preposing) by Torrego (1984:106), who states that in Spanish, a wh-word in the complement position of a tensed clause triggers obligatory inversion in both main and embedded clauses. She argues that obligatory verb inversion in Spanish is similar (though not identical) to Subject-Auxiliary Inversion in English. The following sentence (34) shows that a subject NP is inverted with the main verb, and it cannot generally appear between the wh-phrase and the verb as in (35) (Suňer, 1994: 336).

34. ¿Qué compró Mara ayer?

what buy-PAST Mara yesterday

"What did Mara buy yesterday?"

35. * ¿ Qué Mara compró ayer ?

In the direct question (34), the main verb *compró* 'appears to the left of the subject. According to Rizzi (1996), the order of wh-phrases and verbs in questions follows from the Wh-criterion, a universal constraint on question formation that may be satisfied overtly or covertly.

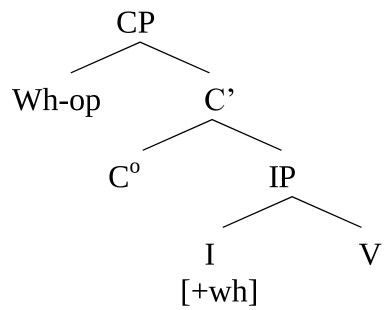
Wh-criterion (Rizzi, 1996, p.64):

a. A wh-operator must be in a Spec-Head configuration with a [+wh] X^0 .

b. A [+wh] X^0 must be in a Spec-Head configuration with a wh-operator.

It ensures that when [+wh] is present in a clause, a wh-operator (wh-phrase) will occur in the Spec CP, licensing both the operator and [+wh] head. The structure required to satisfy the Wh-criterion is (36). In languages with overt movement, the Wh-criterion applies at S-structure accounting for the appearance of a wh-phrase in sentence-initial position. In languages like Spanish, English and Italian, the [+wh] feature is in Infl, and the verb must then move from I to C to satisfy the Wh-criterion, but unlike English, Spanish has the option of leaving the subject in postverbal position or the option of moving it to a left dislocated position. As a result of I-to-C movement and the availability of a postverbal subject position in Spanish, there is a subject-verb inversion in questions (Zagona, 2002. p.94).

36.



The position of the verb in interrogatives may follow from the Wh-Criterion. To satisfy the Wh-Criterion, C^0 must have a [+WH] feature. Rizzi proposes that in main clauses, the [+WH] feature originates in INFL and moves to COMP via the INFL-to-C movement to satisfy the Wh-Criterion. This movement is similar to the subject-auxiliary inversion in the main clauses in languages like English which is also explained by the assumption that [+wh] is generated on INFL in main clauses. Rizzi notes that languages like Italian and Spanish, the subject–verb inversion takes place in the complement clauses as in (37a) as well (Zagona, 2002:244).

37. a. No sabía qué querían esos dos.

not know-PAST what want-PAST those two

"I didn't know what those two wanted."

b *No sabía qué esos dos querían.

not know-PAST what those two want-PAST

"I didn't know what those two wanted."

In the next section, I will discuss island constraints and subjacency in Spanish.

2.6. Island Constraints and Subjacency in Spanish

In his influential work, Rizzi (1982) observes that although there is evidence of the application of Subjacency in Italian in the case of complex NPs, Italian shows systematic differences from English in the realization of the wh-island and subject conditions. Rizzi proposes that the distribution of island effects in Italian may be accounted for by the choice of bounding nodes for Subjacency. If bounding nodes for Italian are S'/CP and NP rather than S/TP and NP/DP as in English, then this can explain the difference between Italian and English in the application of subjacency. This variation has been referred as the Subjacency Parameter (Goodluck, & Rochemont, 1992:1).

Torrego (1984) suggests that like English and Italian (Rizzi, 1982), Spanish exhibits subjacency, but Spanish patterns with Italian in the choice of bounding nodes for Subjacency. The bounding nodes in English are IP and NP (Chomsky, 1973), whereas in Spanish, they are CP and NP. Torrego proposes that V-preposing provides straightforward evidence that Subjacency allows one step wh-movement over the boundary [S´/CP [-wh] S/TP[ ]] in Spanish as in (41b) (Torrego, 1984, 114):

The configurations presented in (38a-b) are the possible derivations for wh-movement allowed by Subjacency in Spanish:

38. a. S'/CP[[wh-phrase$_i$ S/TP […….S'/CP[e$_i$ S/TP[…..S'/CP[e$_i$ S[…e$_i$ ….]]]]]]

b. S'/CP[[wh-phrase$_i$ S/TP […….S'/CP[e$_i$ S/TP[…..S'/CP[S/TP[e$_i$ ….]]]]]]

In English extractions out of wh-islands results in ungrammatical wh-questions as in (39) because wh- phrase moves more than one bounding node (TP) in one step (Cebreiros, 1996, p. 3):

39. *[$_{CP1}$ What book$_i$ don't [$_{TP1}$ you know [$_{CP2}$ if [$_{TP2}$ Pepe has read t$_i$]]]]?

In (39), the wh-phrase *what book* crosses two bounding nodes (IP1 and IP2) at once, due to the fact that the nearest landing side (CP2) is already filled by *if.* In Spanish, however, (40) the equivalent sentence to the ungrammatical (39), is grammatical because TP is not a bounding node in that language, so the wh-phrases can move across two TPs in a single step.

40. [$_{CP1}$ Qué libro$_i$ [$_{TP}$ no sabes [$_{CP2}$ si [$_{TP}$ Pepe ha leido t$_i$]]]
what book not know-PREST-2PR if Pepe has read
*'What book don't you know if Pepe has read?'

Torrego underlies that in such extractions, wh-movement skips one CP-cycle (CP2). Then, V-Preposing is missing in that particular CP-cycle where wh-phrase originates and allows the subject to precede the verb.

To sum up, except for difference in bounding nodes between Spanish and English, Spanish is assumed to exhibit the same island constraints as those observed on wh-movement in English. In the following section, I will present complex NP island, relative clause island, adjunct island, subject island violations along with the *that*-trace effect in Spanish.

2.6.1 Complex NP Island (CNPI)

In Spanish, extracting a wh-phrase out of a NP island violates subjacency as in (41) (Cebreiros, 1996:3). The wh-phrase *qué* crosses two bounding nodes (CP and DP) in a single step to move to the Spec-CP of the main clause, this long movement results in an ungrammatical wh-sentence in Spanish.

41. ¿* [$_{CP}$ Qué$_i$ [$_{TP}$ no sabías [$_{DP}$ el hecho [$_{CP}$ que [$_{TP}$ te dieron t_i]]]]

what not know-PAST-2SG the book that you give-PAST-3 PL

*What didn't you know the book that they gave you t_j?'

2.6.2. Relative Clause Island (RCI)

Similarly, extracting a wh-phrase out of a relative clause island yields an ungrammatical question in Spanish as in (42)[2].

42. *[$_{CP1}$ [$_{PP}$ En cuántas ciudades]$_i$'tiene [$_{TP1}$ [Bill [$_{NP2}$ hermanos [$_{CP2}$ que [$_{TP2}$ viven t_i]]]]]?

in how -many cities has Bill brothers that live

*How many cities does Bill have brothers who live in?

In (42), the wh-phrase *en cuántas ciudades* 'how many cities' moves to the Spec, CP, crossing NP2, and CP1 nodes at once. It also violates the Shortest Movement Principle because the wh-word *en cuántas* ciudades cannot move to the nearest CP because the nearest complement clause Spec-CP is filled by another wh-word *who*, which blocks successive cyclic movement.

[2] I thank Sergio Baauw for providing these examples to me.

2.6.3. Adjunct Island (AI)

Extracting a wh-phrase out of an adjunct island results in an ungrammatical wh-question as in (43) (Montrul, Foote & Perpignan, 2008:94) because it violates subjacency in Spanish.

43. ¿*[$_{CP1}$ A quièn [$_{TP1}$ hablò Josè con María [$_{CP2}$ t_i despuès de [$_{TP2}$ ver t_i]?

who speak-PAST Jose with Maria after seeing?

*Who did Jose speak with Mary after seeing?

In (43), the wh-phrase, *quièn* moves out of an adjunct island (*despuès de ver ‘*after seeing’), to the Spec-CP1, crossing two bounding nodes (CP2 and TP1) in a single step. The shortest movement is not possible due to CP2, which has already been filled by *after*. This blocks a successive cyclic movement in the wh-sentence.

2.6.4. Subject Island (SI)

In Spanish, extraction out of a subject island results in an ungrammatical question as in (44)[3] .

44. *¿[$_{CP1}$ De quién$_i$ [$_{TP1}$ cree [$_{NP1}$ el profesor [$_{CP2}$ que [$_{TP2}$ [$_{NP2}$ una historia t_i [divierte a los niños]]]]]]?

of whom$_i$ believes the teacher that a story t_i amuses ACC. the children

*" Who does the teacher think a story by _ amuses the children?"

In (44), d*e quién* ‘of whom’ moves out of a subject island to the Spec-CP, crossing two bounding nodes (CP2 and DP1) in a single step. This movement violates the subjacency principle.

[3]I thank Sergio Baauw for providing the above examples to me.

2.6.5 *That*-trace effect in Spanish

In English, subject extraction from an embedded clause following complementizer 'that' violates the ECP (see section 2. 4.5 above). However, in object extraction, the complementizer *that* is optional and does cause ungrammaticality since it is an intermediate trace of successive cyclic wh-movement (Kayne, 1980) as in (45 a-b).

45. a. Who do you think (that) Sue met __?

b. Who do you think (*that) __ met Sue?

In contrast, in Spanish the complementizer *that* is typically obligatory in both object and subject extractions as in (46 a-b). Moreover, subject extraction from embedded clauses following the complementizer '*that*' does not create an ECP violation (Torrego, 1984). Torrego accounts for the *that*-trace effect in Spanish by verb-preposing in questions. According to Torrego, once the verb is preposed, it no longer governs a trace that is in the VP. Instead, it properly governs the subject position to its right as in (46b) (Montrul et al, 2008, p. 95).

46. a. ¿Con Quièn piensa-s que Sue reunió?

whom think-PRES-2SG that Sue meet-PAST

"Who do you think that Sue met __?"

b. ¿Quièn piensa María que t_i es de Argentina?

Who think-PRES Maria that is from Argentina

*Who$_i$ does Maria think that t_i is from Argentina?

In (46), the complementizer *que* allows subject extractions from the embedded clause. In contrast the complementizer *that* does not allow subject extraction from the subject position of the embedded clause in English, blocking proper government of subject trace by its antecedent.

In sum, this section shows that as in English, wh-phrases overtly move to the Spec- CP in direct and indirect questions in Spanish, and that they exhibit the same island constraints on overt movement as those in English. However, Spanish diverges from English in terms of bounding nodes (i.e., the bounding nodes in Spanish are CP and DP, whereas they are TP and DP in English). Also, subject extraction from embedded clauses with complementizer *that* does not create that-trace effect in Spanish, but it does in English. In the following section, I will discuss wh-movement in Turkish.

2.7. Wh-movement in Turkish

Turkish is a wh-in-situ language with a basic SOV word order, in which wh-phrases stay in their base-generated positions in both main and embedded questions. That is, wh-phrases such as *kim* 'who', *nereye* 'where', *neden* 'why', *hangisi* 'which', etc, do not overtly raise to the Spec-CP position. Wh-phrases are also inflected with case markers[4] according to their grammatical function in the structure (Arslan, 1999). (47a) and (47b) exemplify declarative and interrogative forms in Turkish, respectively (Arslan, 1999:3).

47. a. Zeynep Ali-yi gör-dü.
   Zeynep-NOM Ali-ACC see-PAST
   Zeynep saw Ali.
  b. Zeynep kim-i gör-dü?
   Zeynep-NOM who-ACC see-PAST
   Who did Zeynep see?

In (47b) wh-phrase *kim-i* stays in-situ at the preverbal syntactic position as the internal argument of the verb *see* and is marked with accusative case. However, in-situ wh-phrases in Turkish move to Spec-CP position in two ways: (1) they undergo movement at LF (covert movement), and (2) can overtly move via scrambling. I will first discuss covert wh-movement in Turkish, providing examples for the main arguments and

[4] **Ø** Nominative, **–(y)I** Accusative, **-(y)A** Dative, **-DA** Locative, **-Dan** Ablative, **-(n)In** Genitive, **-(y)IA** Comitative)

motivations for covert wh-movement (i.e., in scope and selection, weak crossover effects, and locality conditions). Then, I will present arguments for overt movement in Turkish via scrambling.

2.7. 1. Covert movement in Turkish

Following Huang's (1982) LF-raising analysis of the in-situ wh-phrases in Chinese, it has been assumed that in-situ wh-phrases in Turkish as in (48b) move to Spec-CP at LF to derive interrogative interpretation (48a), but the movement is not phonetically observable (e.g., Özsoy, 1996; 2009).

48. a. Zeynep kim-i gör-dü?
Zeynep-NOM who-ACC see-PAST
b. [$_{CP}$ Kim-i$_i$ [$_{IP}$ Zeynep [$_{vP}$ t$_i$ [$_{VP}$ t$_i$ gör-dü]]]] ? (LF)
who-ACC Zeynep-NOM see-PAST
'Who did Zeynep see?'

As in other wh-in-situ languages such as Chinese and Japanese, the LF movement in Turkish has been accounted for on the basis of observations such as parallels in scope and selection, weak crossover effects, and locality conditions (Özsoy, 1990; 1996; Kornfilt, 1996).

2.7. 1.1. Selectional requirement

It has been proposed that in Turkish, selectional properties of the matrix verbs determine the wide or narrow scope of in-situ wh-phrases as in (49a-b) (Arslan, 1999: 19). In English, the wh-phrase in (49a) takes scope over embedded clauses, whereas (49b) it has matrix scope.

49. a. [He wonders [what$_i$ [you bought x$_i$]]].

b. [What$_i$ does [he think [you bought x$_i$]]]

The same scope interpretation obtains in the corresponding cases of Turkish where the wh-element is in-situ as in (50a-b).

50. a. [[Sen-in ne al-dığ-ın] - ı merak ed-iyor].

you-GEN what buy-NOM-2SG.POSS-ACC wonder-PROG

"He wonders what you bought".

b. [[Sen-in ne al-dığ-ın]- ı sanıyor]?

you-GEN what buy-NOM-2SG.POSS-ACC think-PROG

"What does he think you bought?

At LF, after wh-raising has applied, (50a-b) will yield the following representations in (51a-b) similar to those of (49a-b).

51. a. [$_{CP2}$ [$_{CP1}$ ne$_i$ [$_{IP}$ sen-in x$_i$ al-dığ-ın]]- ı merak ediyor].

b. [$_{CP2}$ ne$_i$ [$_{CP1}$ [$_{IP}$ sen-in x$_i$ al-dığ-ın]]- ı sanıyor]?

In (51a) wh-phrase *ne* 'what' moves to Spec-CP of the embedded clause and has scope over the embedded question. In (51b), it is at Spec-CP of the matrix clause and takes scope over the matrix clause. This shows that matrix verbs like *merak et-* 'wonder' and *san*-'think' in Turkish are similar in terms of selectional restrictions, and that in-situ wh-phrases undergo movement at LF.

2.7. 1.2. Scope taking properties

Another argument for the LF movement in Turkish comes from the scope interaction between the quantifiers and the wh-elements in Turkish as in (52-53) (Arslan, 1999).

52. Herkes ne gör-dü?
 everyone what see-PAST
 "What did everyone see?

53. Kim herşey-i gör-dü?
 who everything-ACC see-PAST
 "Who saw everything?"

(52) is ambiguous in that the universal quantifier has both a collective and distributive reading. In the former case, it is a singular question, in which the thing seen by everyone is one and the same thing. In the distributive reading, is a set of questions asking for each person x, what x saw. Sentence (53) is not ambiguous. It has only the collective reading in which an answer like "John saw everything" would be appropriate. This contrast has been accounted for by the Path Containment Condition (PCC) proposed by Pesetsky (1982), who demonstrates that grammatical contrasts observed in overt wh-movements can be naturally accounted for by observing the interaction of paths that such movement creates. An A'-path is a set of successively dominating nodes leading from a trace to its

c-commanding A'-binder. The PCC provides that if two A'-paths intersects, and then one must be properly contained in the other. Overlapping but non-nesting paths are ill-formed. May (1985) has shown that the contrast between (52) and (53) can be seen as an effect of the PCC at LF if quantifiers are subject to QR. The result of applying QR to (52) is (54) (Arslan, 1999, p.37).

54. [$_{CP}$ ne$_j$ [$_{TP}$ herkes$_i$ [$_{TP}$ x$_i$ [$_{VP}$ t$_j$ [$_{VP}$ x$_j$ gör-dü]]]]]

The path connecting *ne* 'what' and its trace t_j consists of {VP, TP, TP, CP}, and the path connecting *herkes* 'everyone' and its LF trace x_i is {TP, TP}. The latter path is properly contained in the former, so the path structure of this LF representation is well-formed with respect to the PCC. In such a structure, where *ne* and *herkes* are in a mutual government relation, either operator may be interpreted as having wider scope than the other, so the ambiguity of (52) arises. The LF representation of (53) is (55).

55. [$_{CP}$ kim$_i$ [$_{TP}$ x$_i$ [$_{VP}$ herşey-i$_j$ [$_{VP}$ x$_j$ [$_{VP}$ x$_j$ gör-dü]]]]]

Here the two paths {TP, CP} and {VP, VP}, do not overlap, so the structure is well-formed with respect to the PCC. In this structure *herşey* 'everything' does not govern *kim* 'who', and so cannot have scope wider than kim 'who'. Therefore, (53) is predicted to be unambiguous. This provides evidence for the application of the PCC at LF in Turkish.

2.7. 1.3. Weak Crossover effects

Further evidence for LF movement of in-situ wh-phrases in Turkish comes from the Weak crossover effects (WCO). The following examples (56a-b) taken from Ikizoglu (2007, p. 40) illustrates WCO effects observed in Turkish.

56. a. *[$_{CP}$ pro$_i$ [$_{TP}$ Anne-si$_i$ [$_{vP}$ kim-i$_i$ sev-iyor]]]?

mother-POSS who-ACC love-PROG

b. *[$_{CP}$ Who$_i$ [$_{TP}$ does his$_i$ mother [$_{VP}$ love [$_{NP}$ t$_i$]]]]?

Although, in (56a) the wh-phrase *kim-i* 'who-ACC' stays at its base position, the sentence is ungrammatical. This shows that (56a) in-situ wh-phrase *kim-i* undergoes movement at LF and is coindexed with *pro anne-si* 'mother-POSS' and its base position. This movement results in the WCO violation as in English (56b).

2.7. 1.4. Locality effects

A final evidence for LF movement of in-situ wh-phrases in Turkish comes from locality effects (e.g. Özsoy, 1996; 2009; Kornfilt, 1996). It has been observed that Turkish, like English, has island effects, however, it also exhibits the argument-adjunct asymmetry in wh-extractions in the same way as the other in-situ languages like Japanese and Chinese do. Arguments in Turkish, unlike English, can be extracted out of islands except for certain types of coordinated structures, but adjuncts cannot. Sentences (57a-b) below illustrate the argument-adjunct asymmetry in wh-extractions (Görgülü, 2006:64).

57. a. Cem [[kim-i sev-en] adam]-ı tanı-yor?

Cem-NOM who-ACC love-REL man-ACC know-PROG

'Who is (x) such that Cem knows the man who loves x?'

*Who$_i$ does Cem know the man who loves t_i?

b. *Cem [Ali'nin nasıl yaz- dığ-ı] mektub-u] oku-du?

Cem-NOM Ali-GEN how-write-NOM-POSS letter-ACC read-PAST

'How is such that Cem read the letter that Ali wrote x?'

*How did Cem read the letter which Ali wrote *t*?

In (57a), the wh-phrase *kim-i* 'who-ACC' is inside the complex NP island (CNPI). The grammaticality of (57a) indicates that unlike English, a wh-phrase as an internal argument of the verb inside the CNPI in Turkish takes the matrix scope without yielding an ungrammatical wh-question. This shows that arguments are not subject to island effects. In (57b), on the other hand, the adjunct wh-phrase *nasıl* 'how' within the CNPI cannot take matrix scope and its extraction results in ungrammaticality. This indicates that adjuncts are sensitive to island effects.

In Turkish, there is also overt movement via scrambling. Below, I will briefly discuss scrambling and island constrainsts on movement via scrambling.

2. 7. 2. Scrambling in Turkish

Turkish with a basic SOV order is known to have scrambling, which leads relatively free word order. Scrambling in Turkish derives from the movement of constituents leftward into various specifier positions. (58b) and (59b) illustrate local and long distance scrambling of wh-words in Turkish, respectively (Özsoy, 2009:223).

58. a. Ayşe kim-i gör-müş?

Ayşe who-ACC see-HS-3SG

'Who has Ayşe seen?'

b. Kim-i$_i$ Ayşe t_i gör-müş?

Who-ACC Ayşe see-HS-3SG

'Who has Ayşe seen?

59. a. Aylin Melis-in ne-yi beğen-diğ-i-ni duy-muş?

Aylin-NOM Melis-GEN what-ACC like-NOM-3SP.POSS-ACC hear-HS-3SG

' What has Aylin heard Melis likes/d?'

b. [$_{CP}$ Ne-yi$_i$ [$_{TP}$Aylin [$_{CP}$ [$_{TP}$Melis-in t_i [$_{vP}$[$_{VP}$ t_i beğen-diğ-i-ni] duy-muş]]]]]?

what-ACC Aylin-NOM Melis-GEN like-NOM-NOM-3SG.POSS-ACC hear-HS-3SG

'What has Aylin heard Melis likes/d?'

In (58b) the wh-phrase *kim-i* (who-ACC) moves from its merge position to the Spec-CP position and takes scope over the matrix clause. In (59b) *ne-yi* (what-ACC), on the other hand, raises to the Spec-CP position of the embedded clause and takes scope over the embedded cause. Following Mahajan (1990), Kural (1993) suggests that wh-scrambling in (58b) and (59b) are syntactic operations constrained by syntactic principles.

In his reformulation of the scrambling phenomena within the generative framework, Mahajan (1990) has shown that leftward movements via scrambling are subject to regular syntactic constraints such as locality constraints on movement, the WCO effect, reconstruction effects and binding properties associated with overt movement. Mahajan characterizes scrambling in two distinct operations: an argument shift operation and an adjunction to XP operation. The former is an L-movement rule and involves substitution into an L-position. It shows binding properties associated with A-movement. The latter is an adjunction operation and shows properties associated with A-bar movement (i.e., it does not provide new binders, it is subject to WCO effects and reconstruction effects).

Similar arguments have been hold for scrambling in Turkish Kural (e.g., Kural, 1993; Aygen, 2000; Öztürk, 2005) suggests that long distance scrambling in Turkish is an A-bar movement because it is subject to locality constraints and WCO effects; it allows reconstruction and it does not provide new binders. Local scrambling, on the other hand, is an A-movement; therefore it does not exhibit such effects. It provides new binders. I will briefly discuss a few of them below:

2. 7. 2. 1. Reconstruction effect

Mahajan (1990) has demonstrated that long distance leftward scrambling in Hindi exhibits reconstruction effects similar to that in A-bar movement in English. The examples in (60) and (61) illustrate reconstruction effect in A-bar movement and A-movement in English (Haegeman, 1994:523), respectively.

60. Which pictures of $himself_i$ does $John_i$ think that Jane will sell?

61. *[This picture of himsef]$_i$ is liked by $John_k$

(60) shows that *himself* is bound by the subject of matrix clause *John* because the antecedent of himself must be the NP with the feature [+Masculine]. The antecedent John does not c-command himself at s-structure, but (60) is grammatical. The binding relation required for (60) is assumed to be achieved by reconstruction, which is a process by which a moved wh-phrase is placed back to a previous movement site. The wh-phrase *which pictures of himself* has moved via the lower [Spec-CP2] to reach the higher [Spec-CP1]. To establish a c-command relation between *John* and *himself*, the wh-phrase is reconstructed to the position of the intermediate trace in order to ensure that the reflexive can be bound by the antecedent *John*. (62) illustrates the s-structure of configuration with all relevant traces.

62. $[_{CP1}$ $[_{IP1}$ $John_i$ does think $[_{CP2}$ [which pictures of $himself_i$]$_i$ that $[_{IP2}$ Jane will sell tj]]]].

This confirms reconstruction effect in an A-bar movement. The preposed phrase is moved, but behaves as if it had not i.e., as if it were there in the position it moved from, which means the preposed phrase is reconstructed into its trace. The ungrammaticality of (61), on the other hand confirms the claim that an A-movement does not allow reconstruction. That is, a moved NP cannot be reconstructed in its trace position to establish a binding relation between *John* and *himself*.

Similarly, in Turkish, Kural (1993) has demonstrated that reconstruction is possible in A-bar scrambling (A-bar movement), while it is not in A-scrambling (A-movement). (63) and (64) illustrate A-bar and A-scrambling in Turkish, respectively (Kural, 1993, p.8).

63. Ahmet adam-lar-ı$_i$ birbirleri-ne$_i$ göster-di.
Ahment-NOM man-PL-ACC each other-DAT see-PAST
"Ahmet showed the men to each other"

64. *Ahmet birbirleri-ne$_i$ adam-lar-ı$_i$ göster-di.
Ahment-NOM each other-DAT man-PL-ACC see-PAST
"Ahmet showed the men to each other".

(63) shows that the subject NP *adam-lar* 'the men' binds the object anaphor *birbirleri-ne* 'each other-ACC', however, in (64) the leftward scrambling of the object anaphor birbirleri-ne 'each other-ACC' cannot bind the subject NP *adam-lar* 'the men', thus the structure is ungrammatical. The ungrammaticality of (64) is due to fact that the object NP moved to sentence initial position via scrambling cannot be reconstructed in its trace position at LF, and thus, cannot be in a binding relation with its antecedent.

2. 7. 2. 2. Weak Crossover effects

Mahajan (1990:23) accounts for the WCO effects by the "Weak Crossover Filter", which states that to be construed as a bound variable, a pronoun must be c-commanded by a binder and its variable (if there is one) at s-structure. That is, a pronoun that is not c-commanded by a binder at s-structure cannot be construed as a bound variable. This accounts for (65b) where the wh-phrase does not c-command the pronoun at the s-structure. It also accounts for (65a) where the wh-phrase does c-command the pronoun but its variable does not. However, the following sentences do not exhibit the WCO effect because the pronoun can be a bound variable.

65. a. Who$_i$ t_i saw his$_i$ mother?

b. [Kim-i$_i$ [$_{TP}$ *pro*$_i$ anne-si [$_{VP}$ t_i sev-iyor]]]?

who-ACC mother-POSS love-PROG

In (65a) both the wh-phrase and its variable c-command the pronoun which is coindexed with *his* and the base trace position is moved to the Spec-CP, while in (65b) *kim-i* 'who-ACC' c-commands the pronoun (pro anne-si). The traces left behind do not affect the WCO filter and that NP movement can provide new binders for a pronoun.

2. 8. Scrambling and Island constraints in Turkish

Long distance movement through scrambling in Turkish exhibits island constraints (i.e., complex NP island, relative clause island, adjunct island, and sentential subject island). This is inline with Boeckx (2008), who states that island effects exists in all languages, but there is some variation in the patterns of extractions that may be difficult to explain within a purely configurational view of locality. In Turkish, Meral (2010: 124) proposes that 'the island status of a linguistic unit is not something which is intrinsic to the unit, but is closely related to the clausal architecture and the grammatical operations applying in the clause".

In Turkish, there is an asymmetry in the extraction of arguments and adjuncts as well as case-marked and bare adjuncts (İkizoğlu, 2007). Arguments do not exhibit island effects except for certain types of coordinate structure. Adjuncts show a different pattern of extractions. Case-marked adjuncts can be extracted without causing ungrammaticality, whereas bare adjuncts cannot. The sentences (66a) and (66b) demonstrate the asymmetry between bare and case marked adjuncts in Turkish, respectively (Ikizoglu, 2007:42).

66. a. *Ahmet-in ne zaman okul-da kim-le konuş-tuğ-u-nu gör-dü-n?

Ahmet-GEN when school-LOC who-COM talk-NOM-3SG.POSS-ACC see-PAST-2SG

'*When did you see whom Ahmet was talking to at school?'

b. Ahmet-in dün nere-de kim-le konuş-tuğ-u-nu gör-dü-n?

Ahmet-GEN yesterday where-LOC who-COM talk-NOM-POSS-ACC see-PAST-2SG

'* Where did you see whom Ahmet was talking to yesterday?'

In (66a), the extraction of the bare adjunct *ne zaman* 'when' from the wh-island results in an ungrammatical wh-question, whereas the extraction of case marked adjunct *nerede* 'where' in (66b) does not cause ungrammaticality in Turkish.

Öztürk (2010, personal communication) argues that in Turkish, long distance movement via scrambling exhibits locality constraints, providing evidence for island constraints in Uyghur (a language from the Turkic language family). Öztürk (in press) proposes that in languages like Turkish and Uyghur, extraction of long-distance scrambled wh-phrases out of islands results in ungrammatical structures. She also notes that there is a subject-object asymmetry in extraction out of islands. For example, a genitive-marked subject can move out of islands without causing any ungrammaticality, whereas objects cannot. The following examples for relative clause island, and adjunct island, are translated for Turkish from Öztürk's study (in press) on Uyghur and Khalkha

2.8.1 Complex NP Constraint

Complex NPs[2] followed by a complement clause behave as islands, but only object extraction exhibits island effect as in (67b). Subjects with genitive case marker can be extracted without causing ungrammaticality as in (67a).

67. a. Ayşe-nin$_i$ Ali [*t*$_i$ kitab-ı oku-duğ-u iddia-sı]-nı yalanla-dı

Ayşe-GEN Ali-NOM book-ACC read- NOM-POSS claim-POSS–ACC deny-PAST-3SG

*'What did Ali deny the claim that Ayşe read?'

b. *Kitab-ı Ali [Ayşe-nin *t*$_i$ oku-duğ-u iddia-sı]-nı yalanla-dı.

book-ACC Ali-NOM Ayşe-GEN read-NOM-POSS claim-POSS–ACC deny-PAST-3SG

* 'What did Ali deny the claim that Ayşe read?'

In (67a), extraction of subject NP *Ayşe-nin* ' Ayşe-GEN' out of the complex NP island does not yield an ungrammatical sentence but extraction of object NP *kitab-ı* 'book-ACC' results in an ungrammatical sentence.

2.8.2. Relative Clause Island (RCI)

In Turkish, relative clauses are islands for long distance extractions. Subjects can be extracted out of relative clause islands as in (68b)[5] but objects fail to do so (68c) .

68. a. Ben [Zemire-nin kitab-ı- ver-diğ-i kız]-ı bil-iyor-um.

I Zemire-GEN book-ACC give-NOM–POSS girl-ACC know-PRG-1SG

I know the girl to whom Zemire gave her book.

b. Zemire-nin$_i$ ben [t_i kitab-ı-nı ver-diğ-i kız]-ı bil-iyor-um.

Zemire-GEN I book-POSS-ACC give-NOM-POSS girl-3SG-ACC know-PRG-1SG

I know the girl to whom Zemire gave the book.

c. *Kitab-ın-ı$_i$ ben [Zemire-nin t_i ver-diğ-i kız]-ı bil-iyor-um.

book-ACC I Zemire-GEN give-NOM girl-3SG-ACC know-PRG-1SG

I know the girl to whom Zemire gave the book.

Extraction of subject with genitive case marker *Zemire-GEN* 'Zemire-GEN' (68b) results in a grammatical sentence, but object extraction *kitab-ın-ı* 'book-ACC' yields an ungrammatical sentence. This shows that subjects with genitive case markers do not exhibit relative clause island, but objects do.

[5] Examples for RCI were translated from Öztürk's (in press) study on Uyghur and Khalkha into Turkish.

2.8.3 Adjunct Island (AI)

Adjuncts are also islands in Turkish. Therefore extraction out of adjunct clauses is not allowed. Since subjects are not marked for genitive, they also fail to move out of adjunct clauses as in the case of non-subjects (69b-c)[6].

69. a. Sen [Zemire film-i seyred-iyor-ken] kitap oku-yor-du-n

you Zemire film-ACC see-PRG-PART book read-PRG-PAST-2SG

You were reading (a) book while Zemire was watching the film.

b. *Zemire$_i$ sen [t_i film-i seyred-iyor-ken] kitap oku-yor-du-n

Zemire you film-ACC see-PROG-PART book read-PRG-PAST-2SG

You were reading (a) book while Zemire was watching the film.

c. *Film-i$_i$ sen [Zemire t_i seyred-iyor-ken] kitap oku-yor-du-n

film-ACC you Zemire see-PRG-PART book read-PRG-PAST-2SG

You were reading (a) book while Zemire was watching the film.

In (69a) and (69b) extraction of both subject and object out of adjunct clauses result in ungrammatical structure, showing that extraction of subjects without genitive case markers and objects exhibit adjunct island effects.

[6] Examples for AI were translated from Öztürk's (in press) study on Uyghur and Khalkha into Turkish.

2.8.4. Subject Island (SI)

Subject island constraint as in (*Who$_i$ does the teacher think a story by t$_i$ amuses the children?) is inapplicable in Turkish because subjects do not form islands in Turkish.

2.8.5. *That*-trace effect

Another property that is investigated in this study is *that*-trace effects. However, *that*-trace effect seems irrelevant for Turkish, because unlike English and Spanish, Turkish does not have an overt complementizer *that* in complement clauses. Therefore, it does not exhibit *that*-trace effect at the trace site, where embedded subject is extracted. The examples (70a-b) illustrate *that*-trace effect in English and Turkish, respectively.

70. a. *Who did he say that wanted a beer?

*[$_{CP}$ Who$_i$ did [$_{TP}$ he say [$_{CP}$ that [$_{TP}$ t$_i$ wanted a beer]]]]?

b. Kim-in$_i$ [*t*$_i$ bir bira iste-diğ-i-ni] söyle-di?

Who-GEN a bear want-NOM-3SG.POSS-ACC say-PAST

*Who did he say that wanted a beer?

(70b) shows that subject extraction from the embedded clauses does not result in ungrammaticality in Turkish because there is not an intervening complementizer *that*, which prevents the subject trace to be properly governed.

In sum, examples for covert movement of in-situ wh-phrases at LF and for overt-movement in scrambling indicate that although Turkish is a wh-in-situ language, it has covert movement at LF but allows overt movement via scrambling. This movement is similar to that of overt movement in languages like English with respect to island constraints.

In the next chapter, I will discuss theories of sentence processing in the first and the second language and the role of working memory in sentence processing.

CHAPTER 3

SENTENCE PROCESSING IN THE FIRST AND THE SECOND LANGUAGE

3.1. Introduction

This chapter will provide an overview of theories of L1 and L2 sentence processing. In the first section, I will first discuss previous research studies in relation to current theories of L1 sentence processing. I will then focus on the processing of wh-dependencies with reference to working memory (WM) resources in L1 sentence processing. In the subsequent section, I will focus on L2 studies that examine processing of wh-dependencies and will discuss the role of WM resources in L2 sentence processing.

3.2. Theories of L1 sentence processing

Understanding how people compose the meanings of sentences from words they hear or read is of great importance for both psychologists and linguists because it would give them deep insights into the nature of human cognitive functioning, and inform them about why language takes the shape it does (Frazier, 1995). For almost more than four decades, research in L1 sentence processing has been seeking an explanatory theory of language comprehension which will provide a detailed description of the comprehension process with its underlying principles and mechanisms together with a specification of the grammar of a particular language. Much of the progress that has been made toward a potentially adequate theory has come from examining how human sentence processing (or parsing) mechanism responds to the local or temporary structural ambiguities that are so common in sentences such as (1) (Frazier & Clifton, 1996, p.20):

1. The horse raced past the barn fell.

The sentence in (1) is temporarily ambiguous because the verb ‘raced’ can be analyzed as a past participle or a tense-marked main verb. When the parser reaches the end of the sentence, it goes back to reread and to reanalyze the sentence. Following the Minimal Attachment Principle (see section below), the parser opts for the simplest syntactic analysis, which contains fewer phrase-structure nodes. Therefore, the parser initially analyzes the verb ‘raced’ as the main verb instead of past participle in this reduced relative clause (i.e, *the horse which was raced past the barn fell*). Nevertheless, the correct relative clause analysis becomes available upon encountering the main verb *fell* at the end of the sentence. These observations in Garden Path sentences have provided evidence for assumption that people compute the grammatical structure of sentences incrementally (e.g., Rayner, Carlson & Frazier, 1983; Ferreira & Clifton, 1986; Clifton, Traxler, Mohamed, Williams, Morris, & Rayner, 2003).

Such findings are compatible with a serial processing account in which, people initially adopt one analysis (the main clause analysis). Difficulty occurs when people realize this inconsistency and reanalyze (i.e., adopt a different analysis). They are also compatible with parallel processing accounts in which people adopt more than one analysis, but rank one higher than any others. Difficulty occurs when later information causes people re-rank their analyses (e.g., Trueswell, Tanenhaus & Garnsey, 1994; MacDonald, Pearlmutter & Seidenberg, 1994; MacDonald, 1994). In the literature of L1 sentence processing, the former is commonly termed as ‘two-stage’ theories, while the latter is known as ‘interactive’ theories.

Let's us now look at each of these theories and the predictions they make for L1 sentence processing in general.

3. 2.1. Modular (two-stage) theories in L1 sentence processing

Modular (two-stage) theories date back to Bever's idea of 'Perceptual Strategies' (1970) and Kimball's 'Seven Principles of Parsing' (1973), and were strongly influenced by the 'information encapsulation' and the 'modularity hypothesis' (Fodor, 1983). Encapsulation is the extent to which different knowledge sources are formally separated. It is a critical property of modularity which claims that the mind is organized into two distinct types of systems: a number of separate, specialized input systems (modules), and a central processing system. Language is one of these modules. Following Fodor, Frazier (1985; 1987) suggests that the human sentence processing mechanism consists of informationally encapsulated modules, with the syntactic module being informationally encapsulated from the others (i.e., there is no non-syntactic influence on syntactic decision-making). The basic assumption of all two-stage theories for ambiguity resolution during initial processing is that the human sentence processor (or parser) is modular, using syntactic information before it uses other information in resolving ambiguity and it processes information serially, retaining exactly one analysis at the initial stage of parsing. The parser eventually applies non-syntactic information, such as semantic or pragmatic information or discourse context.

There are many two-stage theories which are very similar in their basic assumptions, but they differ in many details (e.g., Abney, 1989; Pritchett, 1992; Crocker, 1995; Inoue & Fodor, 1995). For example, most of these models propose that the parser adopts an analysis in which thematic information is applied to a new constituent (e.g., Gibson, 1991; Pritchett, 1991; 1992; Weinberg, 1999). The Garden-Path Theory proposed by Frazier and her colleagues (e.g., Frazier & Fodor, 1978; Rayner, Carlson & Frazier, 1983; Ferreira & Clifton, 1986; Frazier, 1987) is one of the modular models and maybe the most influential one. The Garden-Path Theory claims that the parser assigns a single immediate analysis to an ambiguous fragment of a temporarily or permanently ambiguous sentence, using purely encapsulated structural information (i.e. phrase structure rules). All other sources of information such as thematic roles, discourse context, semantic plausibility, and lexical frequency are ignored during this initial stage. In the second stage, other sources of information are applied to reanalyze the sentence. Frazier (1987) developed two principles the parser applies when faced with syntactic ambiguities: (1) Minimal Attachment, and (2) Late Closure. Both of these principles were basically motivated in terms of reducing memory load. That is, humans must quickly structure materials to preserve them in a limited-capacity memory and that structured materials are better held in immediate memory than the unstructured ones.

The *Minimal Attachment (MA) principle* states that we do not postulate any unnecessary nodes (Frazier & Clifton, 1996, p. 9). That is, in the case of an ambiguity, the MA principle dictates that the parser will adopt the analysis that requires the simple structure rather than the analysis which requires a complex structure. In example (1), the structure associated with the least preferred analysis (i.e., the reduced relative clause interpretation) is more complex as it contains more branching nodes than the structure associated with the preferred analysis (i.e., the main verb interpretation). The LC principle, on the other hand, states that if grammatically permissible, we attach new items into the clause or phrase currently being processed (Frazier & Clifton, 1996, p.9). The LC principle applies to ambiguous sentences in which each interpretation contains the same number of nodes, and predicts that a phrase will form part of the current constituent rather than the start of a new constituent.

A wide range of sentential ambiguities has been analyzed in terms of these principles. Among these are Complement Clause-Relative Clause ambiguities (e.g., Ferreira & Henderson, 1991; Mitchell, Cuetos, Corley, & Brysbaert, 1995); Main Clause-Reduced Relative Clause ambiguities (e.g., Frazier &Rayner, 1982; Rayner et al., 1983; Ferreira & Clifton, 1986; Ferreira & Henderson, 1991); Prepositional Phrase (PP) attachment ambiguities (e.g., Rayner et al., 1983; Ferreira & Clifton, 1986; Clifton, Speer, & Abney, 1991); and Noun Phrase-Sentential Complement ambiguities (e.g., Rayner & Frazier, 1987; Ferreira & Henderson, 1991). The outcome of those experiments clearly confirmed the predictions of these parsing principles and thus supported the Garden-Path Theory of sentence comprehension. For example, in an eye-tracking experiment, Ferreira and Clifton (1986) examined whether semantic or pragmatic information influences the initial syntactic analysis assigned to the sentences with main verb/reduced relative clause ambiguity. They manipulated the plausibility of the initial NP as the agent of the main verb as in (2a-b) (Ferreira & Clifton, 1986, p. 352):

2. a. The evidence examined by the lawyer turned out to be unreliable.

b. The defendant examined by the lawyer turned out to be unreliable.

Sentences in (2) have a temporary syntactic ambiguity that is essentially disambiguated by the *by* phrase. Using the Minimal Attachment Principle, the parser will take the first verb *examine* to be the main verb of the sentence. This will lead to processing breakdown at the point of *by*-phrase where the parser realizes it is led down the garden-path. The sentence (2a) is different from the sentence (2b). The first verb *examine* requires an animate agent. The subject of (2a) *evidence* is not a semantically potential agent, but the subject of (2b) *defendant* is. If the parser can use animacy information to guide the initial parse, the parser will be faster on *by*-phrase in (2a) than on the *by*-phrase in (2b). However, if it initially uses syntactic information alone, the parser will slow down on the *by–phrase* in both sentences to reanalyze the initial misinterpretation. Ferreira and Clifton (1986) found that participants spent significantly longer reading times at the disambiguating *by*-phrase region in both types of sentences, showing garden-pathing emerged due to the MA principle preference. Ferreira and Clifton interpreted these results as compelling evidence for two-stage theories, such that readers do not use semantic category information to guide their initial syntactic analysis, but use it to revise the first analysis.

Although the principles of MA and LC are assumed universal, research studies which examined the universality of these principles, found that they do not always apply in an equivalent manner across different languages. For example, Cuetos and Mitchell (1988), Mitchell and Cuetos (1991) examined how Spanish speakers resolve PP attachment in sentences such as "*Someone shot the servant of the actress who was on the balcony."* They found that, in contrast to what is predicted by the principle of LC, Spanish speakers prefer high-attachment (*the servant)* for the ambiguous PP *on the balcony* rather than the low-attachment (*the actress*). They account for this by the 'Tuning Hypothesis' (Mitchell, Cuetos & Corley, 1992), which states that frequency of usage determines parsing decisions and that high attachment of a RC is more common in

Spanish than in English. The high-attachment preference has also been found in French (Mitchell, 1994; Zagar, Pynte, & Rativeau, 1997), Italian (de Vincenzi & Job, 1993), and Dutch (Brysbaert & Mitchell, 1997). Within the refined Garden-Path Theory, these attachment biases violating the LC principle have been accounted for in terms of a discourse-based revision mechanism, which is assumed to act very rapidly, reversing the preliminary tree-structure-based choice (see Frazier & Clifton, 1996; de Vincenzi & Job, 1993).

3.2.2 Interactive (constraint-based/satisfaction) theories in L1 sentence processing

Current interactive theories were derived from earlier interactive theories (e.g., Marslen-Wilson & Tyler, 1987; McClelland, 1989) and are also termed as *constraint-based/satisfaction theories.* In contrast to modular (two-stage) theories, interactive theories assume that all potentially relevant sources of information, such as syntactic and semantic information, discourse context, and lexical frequency, are used immediately in parallel during the initial stage of sentence processing (e.g., MacDonald, Pearlmutter & Seidenberg, 1994; Trueswell, Tanenhaus & Garnsey, 1994; Trueswell, 1996; Garnsey, Pearlmutter, & Lotocky, 1997; McRae, Spivey-Knowlton & Tanenhaus, 1998). In other words, it is assumed that the parser activates all possible analyses of a sentence in parallel, and that the activation of these analyses depends on the amount of support they receive from the various sources of information. When one particular analysis receives the highest support compared to its alternatives, processing is easy, but when two or more analyses receive equal support, processing difficulty occurs (Pickering & Gompel, 2006, p. 460).

Although they differ from each other in specific details, most constraint-based/satisfaction theories have a lexicalist approach to sentence processing. According to these models, lexical representations contain syntactically relevant information such as verb-argument structure information, lexical category information, and morphological information such as tense and number; therefore, many syntactic ambiguities are associated with lexical ambiguities and resolved in a similar way to lexical ambiguities (e.g., MacDonald et al., 1994; Trueswell, 1996).

Support for the constraint-based/satisfaction theories comes from a substantial number of studies that have examined the role of a variety of factors in ambiguity resolution such as plausibility information (e.g., Trueswell et al., 1993; 1994; MacDonald et al., 1994; Mak, Vonk, & Schriefers, 2002; Clifton et al., 2003; Kim & Osterhout, 2005), lexical frequency (e.g., Trueswell, 1996; Garnsey et al., 1997; Van Gompel & Pickering, 2001), and discourse context (e.g., Altman & Steedman,1988; MacDonald et al., 1994; Tanenhaus, Spivey, Eberhard, & Sedivy, 1995; Spivey & Tanenhaus, 1998; Spivey, Tanenhaus Eberhard, & Sedivy, 2002; Chambers, Tanenhaus, & Magnuson, 2004).

For example, following Ferreira and Clifton's (1986) study, Trueswell et al., (1994) examined sentences having Main Verb/Reduced Relative Clause ambiguity like (2a-b) given above. They used items that Burgess (1991) developed using sentence completion norms, in which they ensured the semantic fit between subject NPs (i.e. *the evidence* versus *the defendant*) and verb (i.e. *examine*) (e.g., inanimate NPs as poor Agent, but good Patient/Theme of the verb as in '*The evidence examined'*). Their assumption was that the semantic fit between the initial NPs and their thematic roles would activate (or support) the past participle form (the reduced relative clause interpretation) rather than past tense form (main verb interpretation). Animacy information as just one of the constraints might be strong enough to override the main verb interpretation, but if there is a strong frequency asymmetry between the main verb and the reduced relative interpretations, it may not completely override the frequency constraint. They reported that participants did not show an initial preference for the main verb analysis in sentences with inanimate NP followed by a verb. In other words, participants did not spend longer RTs on *by* phrase in processing sentences with an inanimate noun plus verb (i.e., *The evidence examined*…), but they spent longer RTs on *by* phrase in sentences with an animate noun plus verb (i.e., *The defendant examined*…) (cf. Ferreira and Clifton, 1986). These results were interpreted as evidence for the immediate use of plausibility (i.e. animacy) and lexical frequency in resolving ambiguity in on-line sentence processing and as support for constraint-based theories. These results were confirmed by subsequent studies of Trueswell (1996) and Garnsey et al. (1997), who examined the effects of the frequency of syntactic constructions during the initial stage of sentence processing.

Similar arguments have been forwarded for the role of discourse context effects (e.g., Spivey-Knowlton et al., 2002) in sentence processing. Many studies have tested the prediction of the referential theory (e.g., Crain & Steedman, 1985; Altman & Steedman, 1988), which suggests that different syntactic analyses are developed in parallel but the structure whose pragmatic presuppositions best satisfied by the discourse is then rapidly selected. Thus, it is predicted that appropriate discourse contexts can override syntactic preferences during the initial stage of parsing. Studies such as Spivey-Knowlton et al. (2002), Altman et al. (1994), have found evidence for immediate effects of the discourse context thus provide support for constraint based theories.

To sum up, there is a good amount of evidence for both modular and interactive theories of sentence processing. However, results from modular (two-stage) and interactive (constraint-based/satisfaction) theories are not yet conclusive. It appears that all theories agree on the view that all sources of information are used during sentence processing, but they disagree on the time course of the application of these sources. There are good reasons to believe that syntactic information is used initially and cannot be reduced to a set of weakly interacting constraints, whereas research has shown that frequency, plausibility, and discourse context play important roles during sentence processing. In the next section, I will discuss the processing of wh-dependencies in L1 sentence processing in the light of processing theories discussed above.

3.3. Processing *wh*-dependencies in the L1

Although wh-dependencies are not syntactically ambiguous, they might still be difficult to process as they involve empty categories (ECs). An EC poses at least two problems for the parser. The first one is that as it is not present at the s-structure; its existence has to be inferred indirectly. The other is that in order to receive an interpretation, it has to be associated with an antecedent phrase, which is usually not adjacent to it and is often quite distant. This spoils the strictly local character of the processes that are otherwise sufficient to construct a representation for a sentence (Fodor, 1989, p. 156).

Research on processing of *wh*-dependencies has mainly focused on whether or not the parser makes use of ECs in decoding wh-dependencies. So far, there have been two competing groups of processing accounts which make different assumptions about how *wh*-dependencies are analyzed by the human sentence processing mechanism. One of them is 'gap-based' accounts, which suggests that *wh*-dependencies involve the construction of gaps in canonical argument positions (e.g., Crain & Fodor, 1985; Stowe, 1986; Frazier and Clifton, 1989; Nicol & Swinney, 1989; de Vincenzi, 1991; Nicol, 1993; Gibson & Hickok, 1993; Nakano, Felser & Clahsen, 2002). The other one is the 'gap-free' accounts in which the formation of wh-dependencies involves direct dependency between the wh-filler and a verb (Pickering & Berry, 1991; Pickering et al, 1994; Sag & Fodor, 1994; Steedman, 2000).

The following section includes a detailed discussion of the gap-based and gap-free accounts in reference to studies examining processing of *wh*-dependencies.

3.3.1. Gap-based accounts in processing of wh-dependencies

The gap-based accounts derive from transformational generative grammar, in which *wh*-dependencies are formed through the movement of an argument of the verb from its canonical position to the sentence-initial position, leaving a trace (or gap) at its original location (e.g., Chomsky, 1981; 1995). The major question addressed in the gap-based accounts is how the parser identifies the gap site during real-time sentence processing. Some important early set of studies demonstrated that the parser actively predicts the gap sites as the sentence unfolds (filler-driven parsing) (Fodor, 1978; Crain and Fodor, 1985; Frazier, 1987; Frazier and Flores D'Arcais, 1989), rather than waiting to identify an empty argument position before positing a gap (gap-driven parsing) (Jackendoff & Culicover, 1971; Wanner & Maratsos, 1978).

Stowe, (1986) observed a *filled gap effect*[1] at the direct object gap position of the embedded verb in (3b) with a fronted wh-phrase and this is reflected in slower reading times for the pronoun *us*. But no such effect was found in the control condition that did not involve a fronted wh-phrase (3a). This slowdown is expected if the parser actively posits a direct object gap in (3b) as soon as it encounters the transitive verb *bring*, and hence experiences difficulty when it finds an overt pronoun in the direct object position. The slowdown is unexpected if the parser waits to identify an empty argument position before positing a gap (Stowe, 1986, p. 234).

3. a. My brother wanted to know if Ruth will bring us home to Mom at Christmas.

   b. My brother wanted to know who Ruth will bring us home to _at Christmas.

[1]*Filled gap effect*: a surprise effect that is elicited when readers encounter an overt NP in a post-verbal position where a gap is anticipated.

In other words, according to the gap-driven (also known Gap as a Last Resort) strategy, the parser waits until a sequence of words is encountered which would not be syntactically well-formed unless a gap is posited and the filler is inserted at that position. For example, in (3b) the parser might read …*Ruth will bring us home to at* and realize that the sequence *bring us home to at* cannot be parsed unless the filler *who is* inserted after *to*. In contrast, the filler-driven strategy suggests that the parser actively predicts a gap site as soon as it identifies a *wh*-filler (e.g., *who* in 3b). Having identified the wh-filler *who*, the parser processes *Ruth will bring* and posits a gap after *bring* and immediately interprets *who* as the direct object of the verb. On encountering *us* this analysis has to be cancelled and the parser has to reanalyze the sentence and posit a gap to the next potential gap position that is licensed by the grammar. In the above study, Stowe showed that readers slow down after the verb *bring* , which is a potential but not realized gap position, suggesting that they are forced to reanalyze the object gap analysis at this point, and provide evidence for claim that the native speakers of English adopt the filler-driven strategy. Related evidence for the filler-driven approaches has also been found in many languages such as Dutch (Frazier, 1987; Frazier & Flores d'Arcais, 1989; Kaan, 1997), Russian (Sekerina, 2003), Hungarian (Radó, 1999), Italian (De Vincenzi, 1991), German (Schlesewsky, Fanselow Kliegl, Krems, 2000; Fiebach, Schlesewsky & Friederici, 2002), and Japanese (Nakano, Felser & Clahsen, 2002; Miyamoto & Takahashi, 2002; Aoshima, Phillips, & Weinberg, 2004).

Also, a set of priming studies in L1 sentence processing research has provided empirical evidence for the psychological reality of the gaps (which have grammatical but not phonetic features) and for reactivation of fillers at the gap position (e.g., Nicol &

Swinney, 1989; Nicol, 1993; Love & Swinney, 1996; Clahsen & Featherston, 1999; Nakano, Felser, & Clahsen, 2002). These studies reported that gaps are covert anaphoric elements, which produce experimental priming of their antecedents (filler) at the gap site in the same way as the overt pronouns and other overt anaphoric elements do. Such findings have first been accounted for by a well-known principle within the framework of the Garden-Path Theory, namely the Active Filler Strategy (AFS) proposed by Frazier and Clifton (1989, p.95):

Active Filler Strategy: "When a filler has been identified, rank the option of assigning it to a gap above all other options."

The AFS stipulates that as soon as the parser encounters a wh-filler, and identifies it, using phrase structure rules (i.e., the syntactic category of a phrase) alone, it posits a gap that has the same phrasal category as that of the filler in a syntactic phrase marker. In the example (3b) given above, as soon as the parser encounters the wh-filler *who,* it posits a gap in the argument position of the earliest verb *bring*. However, this assignment turns out to be wrong once the parser encounters an overt NP *us* in the argument position of the verb *bring,* where it predicts a gap. This surprise results in a filled gap at this point. Then, the parser reanalyzes this first parse. The AFS assumes that working memory resources are necessary to establish the dependency between the filler and its gap in such a way that the wh-filler is actively kept in working memory until the gap is identified.

To deal with the data from null-subject languages like Italian, an alternative version of the AFS has been offered by De Vincenzi (1991) as the Minimal Chain Principle (MCP) (De Vincenzi, I991, p. 199):

"Avoid postulating unnecessary chain members at the s-structure, but do not delay required chain members."

Based on the insights that chains are computationally complex, the MCP suggests that the parser will always posit the shortest and fewest chains when determining the decisions made at ambiguous points. For example, when a filler, which requires a chain with an empty category, is found, the parser tries to complete the chain, postulating an NP trace at the first available empty NP position. Consider the sentences (4a-b) below (De Vincenzi, 1991, p. 201). A structural ambiguity in Italian occurs in a sentence like (4) in which a wh- or the postverbal NP can be the subject of the sentence since in Italian, the subject can freely appear in the postverbal position. The parser prefers associating wh-filler with the subject gap as in (4a) rather than the object gap (4b) because it is the shortest chain.

4. Chi (*t*) ha chiamato (*t*) Giovanni?

" Who has called Giovanni?"

a. Chii ei ha chiamato Giovanni?

"Who has called Giovanni?"

b. Chii ej ha chiamato ei Giovanni?

"Who has Giovanni called?"

Another alternative to the AFS is that the parser posits a gap as soon as an appropriate verb is identified. This approach is well represented in both the modular parsing theories (e.g., Gibson, 1991; Pritchett, 1991; 1992; Gibson, Hickok, & Schütze, 1994; Weinberg, 1999) and the constraint-based lexicalist theories (e.g., MacDonald, 1994). For example, Pritchett's Theta Attachment Constraint states that "*The theta* *criterion attempts to apply at every point during processing* …" (1988, p.542). This suggests that as soon as a wh-filler is identified, an active search is made to link it to an argument role. In line with Pritchett's views, Gibson & Hickok (1993) propose that gaps can be posited as soon as their positions are licensed by the grammar. According to their gap-positing algorithm, they assume that given a wh-NP, an associated gap will be posited in the direct object position of an English transitive verb as soon as that verb is encountered, because the direct-object position is fully licensed by the verb under X-bar theory, theta theory and Case theory (p. 152).

This approach has challenged the claim that the parser initially applies syntactic information in an encapsulated manner (i.e., independent of issues of verb-specific constraints such as the thematic fit between a filler and a verb) during the formation of filler-gap dependencies. In head-initial languages like English, evidence for dependency formation at the verb position comes from a number of different sources, including filled-gap effects in reading-time studies (Crain & Fodor, 1985; Stowe, 1986), eye-movement studies of implausibility detection (Traxler & Pickering, 1996), antecedent reactivation effects (Nicol, Fodor & Swinney, 1994; but cf. McKoon , Ratcliff &Ward, 1994), event-related potential (ERP) measures (Garnsey, Tanenhaus, & Chapman, 1989; Kaan, Harris, Gibson, & Holcomb, 2000; Felser, Clahsen, & Münte, 2003; Phillips, Kazanina, & Abada, 2005) and eye-tracking (Sussman & Sedivy, 2003).

Moreover, a number of studies showed that the active positing of gap site is filtered by subcategorization constraints (i.e., the lexical argument structure requirements of the verb) (e.g., Stowe, Tanenhaus, & Carlson, 1991; Boland, Tanenhaus, Garnsey, & Carlson, 1995, but see Pickering & Traxler, 2003). For example, using a stop-making-sense task, Boland et al., (1995) showed that the fit between the verb and the argument may reflect selectional as well as simple argument structure constraints. They found that a *filled gap effect* disappeared when the filler was an implausible direct object of a transitive verb that allowed an additional gap site inside its complement, as in (5a) as opposed to (5b), in which a *filled gap effect* emerged since the filler was a plausible direct object of the verb *visit* (Boland et al., 1995, p. 781).

5. a. Which prize did the salesman visit ……?

   b. Which client did the salesman visit…….?

Another line of work has examined whether the parser respects island constraints[3] during the real-time construction of filler-gap dependencies. A number of different studies in recent years have examined the sensitivity of the parser for islands in a number of different experimental measures; however, they have obtained conflicting results.

Some studies have shown that the parser is sensitive to island constraints during the initial stage of parsing (Stowe, 1986; Bourdages, 1992; Pickering, Barton, Shillcock, 1994, (exp. 2); McElree & Griffith, 1998; Yoshida, Aoshima, & Phillips, 2004). For example, in a self-paced experiment, Stowe (1986) showed that the filled-gap effect observed at licit gap sites was not found in a subject-island context.

[3] Island constraints involve limitations on the movement of a wh-phrase from its base position. The term *island* introduced by Ross (1967) refers to constructions that do not allow a wh- phrase to escape from them. Ross identified a number of island constructions such as complex NP island, relative clause island, and adjunct island. For example, movement of a wh-phrase out of a relative clauses yields an ungrammatical wh-question as in (*Who_i *did the candidate read a book that praised* t_i?) (Phillips, 2006, p. 796). See Chapter 3 for a detailed analysis and more examples of island constraints.

She examined subject islands, such as the bracketed phrase in (6a), to test whether gaps are incorrectly posited in positions where they cannot grammatically occur. A gap cannot grammatically occur in a prepositional complement to a subject NP as in (6a), although a NP may appear in the corresponding condition in an *if*-clause (6b). She predicted that if participants incorrectly expect an illicit gap inside the subject island, they will experience a filled-gap effect (i.e., a local increase in processing load) at the NP *Greg's* in the wh-clause (6a), but not in the corresponding if-clause (6b).

6. a. The teacher asked $what_i$ [the silly story about Greg's older brother] was supposed to mean t_i.

   b. The teacher asked if the silly story about Greg's older brother was supposed to mean anything.

She found that the NP Greg's in example (6a) which is a part of complex subject NP, was read just as quickly as the one in (6b). In other words, no filled-gap effect was observed at the NP Greg's in (6a), suggesting that no gap site is posited after the preposition *about* and that the parser applies island constraints during the initial stage of the construction of filler-gap dependencies.

However, in contrast to the findings of immediate island sensitivity, some studies have argued that the parser is able to construct representations that violate island constraints (e.g., Freedman and Forster, 1985; Clifton and Frazier, 1989; Kurtzman and Crawford, 1991; Pickering et al., 1994, (exp.1)). For example, in an eye-tracking and self-paced reading experiment, Pickering et al. (1994, Exp.1), tested the question of whether or not the parser posits a gap site inside a relative clause island. They used sentences with illicit wh-extractions out of relative clause islands as in (7a) and sentences with legal extraction without island constraint as in (7b).

7. a. *I realize what the artist who painted *(_) the large mural ate_ today?

b. I realize what the artist painted (_) the large mural with_ today?

Pickering et al., (1994) found longer reading times from both eye-tracking and self-paced reading at the critical region (i.e. *painted*) in the both extraction conditions (7a-b). This suggests that the participants were forming dependencies at this position. Since the results from the island condition (7a) resemble those from the legal extraction condition (7b), this supports the view that the parser does not use island constraints during real-time construction of wh-dependencies.

To sum up, there is a consensus among gap-driven processing accounts involving either a syntactically-driven strategy or lexically-driven strategy that the mental representation of wh-dependencies involves the formation of a filler-gap. Experimental investigations provide substantial evidence for the construction of filler-gap dependencies during parsing. However, there is still no consensus on the type of information the parser applies in the initial filler-gap formation.

3.3.2. Gap-free (traceless) account of wh-processing

Emerged from non-transformational theories such as Categorial Grammar (e.g., Pickering, 1993; Steedman; 2000), Lexical-Functional Grammar (Kaplan and Zaenen, 1988), and Head Driven Phrase-Structure Grammar (Pollard and Sag, 1993), gap-free accounts do not assume gaps (traces) in the formation of wh-dependencies. The rationale behind this view is that unbounded dependencies are not derived from canonical sentences or representations, and they do not require the existence of extraction site where empty categories are located (Pickering & Barry, 1991). However, similar to gap-based accounts, they expect an association between the extracted element and the subcategorizer. In other words, the parser has to associate the filler with the subcategorizer directly in the formation of the unbounded dependencies, but there is no need to assume the existence of phonologically null intermediary form during the construction of unbounded dependencies.

The Direct Association Hypothesis (DAH) (Pickering & Barry, 1991; Pickering, 1993; Pickering, Barton, & Shillcock, 1994) is developed within gap-free accounts. The DAH states that when the parser encounters a wh-phrase, it stores the filler in memory and associates it directly with its subcategorizer as in (8) (Pickering & Barry, 1991, p. 231).

8. [Which man]$_i$ do you think Mary [loves]$_i$?

Nevertheless, in a sentence like (8), it is difficult to see any empirical difference between the two classes of theories (gap-based versus gap-free accounts) because the trace in this type of sentences is adjacent to the subcategorizer *love* as in (9) (Pickering & Barry, 1991. p. 230):

9. [Which man]$_i$ do you think Mary loves t_i ?

In this context, it has been argued in the experimental literature that evidence used to explain the existence of gaps in the formation of wh-dependencies can very well be used to explain the direct association of the filler with its subcategorizer as well. We will come back to this issue later. Now, let us examine the arguments for gap-free accounts in sentence processing.

Pickering and Barry (1991) provide arguments against traces using two sorts of sentences: those with an extraction site remote from its subcategorizering verb, and those with multiple embeddings (see Gibson & Hickok, 1993; and Gorrell, 1993 for counter arguments, Pickering 1993 for the reply). For example, they contrast sentences such as (10a), in which an argument PP is extracted to sentences such as (10b), in which an object of the argument PP is extracted (Pickering & Barry, 1991, p. 232-233).

10. a. In which box did you put the cake?

    b. Which box did you put the cake?

The verb *put* canonically takes two post-verbal arguments in the order NP PP, as in *put the cakes in the box*, and the PP-argument has been extracted as in (10a). Pickering and Barry (1991) hypothesize that according to gap-based accounts, the filler *in which box* has to be held in the memory until the end of the sentence, because it is impossible to associate it with *put* until the gap is located as in (11a). Recall that in gap-based accounts, first, a dependency is constructed between the filler and its gap, and then, the filler is linked to its subcategorizer. This means that there can be no interpretation of *in which box did you put*. This seems wrong because *in which box* is an argument of *put.* In contrast, the DAH associates *in which box* with *put* directly as in (11b).

11. a. [In which box]$_i$ did you put the cake t_i ?

    b. [In which box]$_i$ did you [put]$_i$ the cake?

Based on this analysis, Pickering and Barry claim that the DAH can explain a range of processing phenomena for which the gap-based accounts make incorrect predictions.

In the case of (10b), *which box* is not itself an argument of *put*, but instead is subcategorized for by the preposition *in.* Therefore the gap will be after *in* as in (12a) and the DAH assumes an association between *which box* and *in* as in (12b).

12. a. [Which box]$_i$ did you put the cake in t_i ?

    b. [Which box]$_i$ did you put the cake [in]$_i$?

Since the gap location is adjacent to the subcategorizer, the gap-based and the DAH do not show a clear difference. Extending the distance between the gap location and the verb *put* by replacing *the cake* with a longer NP as in (13a-b), they try to give a clearer distinction between these two accounts.

13. a. In which box did you put the very large and beautifully decorated wedding cake bought from the expensive bakery?

    b. Which box did you put the very large and beautifully decorated wedding cake bought from the expensive bakery in?

It has been suggested that (13b) is intuitively more awkward than (13a) for most speakers. Pickering & Barry attribute this to the fact that the filler *which box* (that needs to be associated with the sentence-final proposition) must be held in memory while the complex NP is being processed. Assuming the DAH, this is not true for (13a) as the preposed PP is directly associated with the verb *put.* The gap-based analysis posits a gap in both sentences after postverbal NP and therefore it cannot predict the comparative awkwardness of (13b).

However, Gibson & Hickok (1993) and Gorrell (1993) have demonstrated that the distance between the filler and its subcategorizer is an inadequate predictor of the complexity. They argue that Pickering and Barry's bottom-up incremental parser assumes that the gap cannot be posited until the preceding constituent has been entirely processed. For example, Gibson and Hickok (1993) suggest that within the gap-based accounts, the parser can posit a gap on encountering a verb (i.e. as soon as the filler is licensed by a subcategorizer) and before the purported gap site is reached (see Gibson and Hickok, 1993 for more explanation). Nevertheless, these researchers admit that distinguishing these two accounts is not possible in most cases due to the fact that in available data in head-initial languages, gap sites are adjacent to subcategorizers.

For example, associative (or antecedent) priming effects found by priming studies in head–initial languages (e.g., Swinney, Ford, Frauenfelder, & Bresnan, 1988; Nicol & Swinney, 1989; Nicol, 1993; Nicol, Fodor, & Swinney, 1994), for the filler at the offset of its subcategorizer in sentences like (14) has been assumed due to a gap which causes reactivation of the properties of the filler (or antecedent). In the sentence (14) taken from Pickering & Barry (1991, p. 231), Swinney et al., (1988) found that associates/antecedents of *boy* were primed after *accused*.

14. The policeman saw the boy that the crowd at the party accused of the crime.

This was interpreted as antecedent reactivation effect at the gap sites as in (15). Within the DAH, it could be interpreted equally well by direct association as in (16).

15. The policeman saw the boy [that]$_i$ the crowd at the party accused t_i of the crime.

16. The policeman saw the boy [that]$_i$ the crowd at the party [accused]$_i$ of the crime.

Similar assumptions have been hold for other experimental results such as the filled-gap-effect found by a set of studies (e.g. Crain & Fodor, 1985; Stowe, 1986) at the potential gap sites which were adjacent to subcatorizers, as in (17a-b)) (Stowe, 1986, p. 234).

17. a. My brother wanted to know if Ruth will bring us home to Mom at Christmas.

b. My brother wanted to know who Ruth will bring us home to at Christmas.

As discussed in the section 3.3.1, Stowe (1986) attributes longer reading times found at the NP *us* in (17b) (examples 3a and 3b are repeated here as 17a and 17b) to the filled-gap effect parser experienced. The parser initially posited an object gap after *bring* to link with the filler *who*, but this analysis turned out to be incorrect once the parser encounters an overt NP at the gap site following the verb *bring*. This forces the parser to revise this incorrect analysis. According to the DAH, the filler *who* is immediately associated with the verb *bring* directly. When the parser encounters *us*, it was forced to undo this association.

Empirical evidence to ensure the role of gaps in sentence processing has been provided by the studies in verb-final languages in which it is possible to show an experimental effect at an assumed gap location before the subcategorizing verb, which cannot be explained in terms of the DAH (Clahsen & Featherston, 1999; Miyamoto & Takashi, 2002; Nakano et al., 2002; Fiebach et al., 2004; Aoshima et al., 2004). There is also evidence form studies investigating the processing of indirect-dependencies (Nicol, 1993), and subject-relative clauses (Swinney & Zurif, 1995; Lee, 2004), or dependencies spanning more than one clause (Gibson & Warren, 2004). For example, Aoshima et al. (2004) found a filled-gap-effect similar to that of Stowe (1986) in the formation of wh-dependencies in a scrambled condition in Japanese. This effect was assumed to occur due to an attempt that the parser makes to create a gap site before it encounters the subcategorizing verb.

To sum up, in real-time processing of wh-dependencies, gap-based accounts including the AFS, assume that the parser makes use of gaps to construct filler-gap dependencies. In contrasts, gap-free accounts like the DAH, suggests that the parser directly associates the filler with its subcategorizer. However, both accounts share the common assumption that the filler is held in memory until the parser encounters its gap or subcategorizer.

In the following section, I will discuss ‘working memory’ (WM) as another important factor determining the speed and accuracy in sentence processing. I will first start with a definition of WM and its constructs and then discuss, in the subsequent sections, the L1 studies measuring the role of WM capacity in sentence processing.

3. 4. Working memory and L1 sentence processing

3.4.1. Introduction

In cognitive science, WM has been defined as a mental construct, which involves the temporary storage and manipulation of information that is assumed to be necessary for a wide range of cognitive activities such as reasoning, problem solving, decision making, and language comprehension (Baddeley & Hitch, 1974; Gathercole & Baddeley, 1993; Baddeley & Hitch, 1994; see Miyake & Shah, 1999). In that sense, WM is viewed as both a storage space and a processing ground for performing computations. Indeed, the concept of WM has emerged from the traditional view of short-term memory (STM), which refers to a limited-capacity input memory, which can retain information for approximately 18 seconds without rehearsal (Petersen and Peterson, 1959); and which has a rehearsal loop to keep a trace of the information in the STM and to transfer it to the long-term memory (LTM) (Waugh and Norman, 1965).

The STM has later been incorporated into broader models to develop the concept of WM as a complex construct. The best known of these broader models is Baddeley and Hitch's (1974) *multi-component model*, which incorporates an earlier conceptualization of the STM into their concept of slave systems, consisting of *the phonological loop* (for storing and rehearsing verbal information) and *the visuospatial sketchpad* (for storing and rehearsing visuospatial-based information). These slave systems are dependent on a limited capacity attentional system termed as *the central-executive component*, which is responsible for allocating attention, planning, inhibiting irrelevant responses and coordinating resources demanded by concurrent tasks. Baddeley (2000) proposed a fourth component of the working memory system, namely, the *episodic buffer*, which stores and binds information from a number of different sources into chunks or episodes.

One of the critical questions discussed in relation to the nature and the functional organization of WM in the mind is whether WM is a unitary or a modular construct. More specifically, the question is: are all higher level cognitive activities supported by a single pool of general purpose WM resources (e.g., Just & Carpenter, 1992). Or are there separate pools of resources dedicated to supporting different processes and representations (e.g., Caplan & Waters, 1996) In other words, is the WM system domain-general in nature, in the sense that the same neural/cognitive resources are used for multiple cognitive functions (e.g., language comprehension, musical processing, counting, mental arithmetic, and face perception), or are there separate cognitive modules, which are subserved by highly specialized neural structures dedicated to specific cognitive functions and if so, how many pools of resources/modules are there?

With respect to the issue of modularity of verbal WM resources underlying language processing, different views have been proposed. For example, Just and Carpenter (1992) assume a single WM for language comprehension. They argue that during sentence processing, we use the items in the WM to construct syntactic and semantic representations to comprehend the sentence. In other words, there is one single pool of resources for all verbally-mediated tasks.

Unlike Just and Carpenter (1992), Waters and Caplan (1996) and Caplan and Waters (1999) have proposed that the verbal WM resources can be divided into two separate pools: (1) verbal WM for linguistic processes (i.e., for on-line interpretive processes), which involve automatic first-pass language processing; and (2) verbal WM for non-linguistic cognitive tasks (i.e., for off-line post-interpretive processes), which contain conscious processing of the propositional content of a sentence and use of it to accomplish tasks such as reasoning, planning actions, storing information in the long-term semantic memory.

Before, moving into a detailed discussion of further differences between these two views referred to as the Single-Resource (SR) theory and the Separate Sentence Interpretation Resource (SSIR) theory, I will first briefly note, in the following section, how the role of verbal WM resources has been perceived in L1 sentence processing in general. I will then return to the discussion on the theories of SR and SSIR. In the final part of this chapter, I will discuss WM span tasks (e.g., reading span task, operation span task) that are commonly used to measure WM capacity. Within this context, I will also discuss some statistical and methodological issues related to different measures of WM. Two common analytical approaches, namely the correlational studies and the individual differences approach, which are used to assess the relationship between WM and language comprehension will be relevant in this discussion.

3.4.2. Working memory constraints in L1 sentence processing

Sentence comprehension is a cognitively complex and demanding task. Results from behavioral research into sentence processing have shown that sentence comprehension requires a moment-by-moment integration of different information sources, constrained by computational and memory resources (e.g., MacDonald et al., 1994; Trueswell et al., 1994; Trueswell, 1996; Gibson, 1998; 2000; Roberts & Gibson 2003). Early accounts of memory constraints on sentence comprehension go back to

Miller and Chomsky (1963), who observed that center-embedded sentences as in (18) are grammatically acceptable but are nearly impossible to understand.

18. The mouse that the cat that the dog hated chased got caught in the trap.

They have argued that difficulty in understanding such a sentence is due to the fact that each of initial NPs has to be held in WM until it is associated with its appropriate predicate and that more than two NPs exceed the capacity of most people's WM. This assumption was supported by Kimball (1973), who proposed that sentence comprehension was clause-based, such that, at most two partially processed clauses could be maintained in WM at a time. This proposal has accounted for the difficulty associated with center-embedded structures as in the example above.

This proposal has been challenged by Gibson's (1998; 2000) Syntactic Prediction Locality Theory (SPLT), which suggests that the difficulty that people have in processing center-embedded structures might stem from two factors: (1) performing structural integration by connecting a word into the structure for the input; and (2) keeping the structure in memory, which requires keeping track of incomplete dependencies. Gibson and his colleagues (e.g., Gibson, 1998; Warren and Gibson, 2002; Felser et al., 2003; Nakatani & Gibson, 2003; Hsiao & Gibson, 2003; Chen, Gibson, & Wolf, 2005; Gibson et al., 2005; Grodner & Gibson, 2005) have tested participants on center-embedded subject and object relative clauses as in (19a-b) (Gibson, 1998, p. 4).

19. a. The reporter that attacked the senator disliked the editor.

b. The reporter that the senator attacked disliked the editor.

Their results provided empirical evidence for the SPLT. That is, object extracted relative clauses were found to be more difficult than subject-extracted relative clauses due to storage and integration cost of the NP *the reporter* with the verb phrase *attacked*.

The role of WM constraints in sentence comprehension has also been implicated in the context of decoding Garden Path sentences (i.e., The *Minimal Attachment and Late Closure Principles*) (e.g., Frazier, 1987; Frazier & Clifton, 1989), The *Recency/Locality Principle* (e.g., Pearlmutter & Gibson, 2001), the *Active Filler Strategy*, (e.g., Frazier and d'Arcais, 1989), *The Minimal-Chain Principle* (De Vincenzi, 1991) and *The Direct Association Hypothesis* (e.g., Pickering and Barry, 1991).

Recall that in Garden-Path theory of Two-Stage models, it has been suggested that the parser computes only a single and the simplest analysis of an input sentence at a time, using the Minimal Attachment strategy because holding more than one analysis in WM can tax more memory resources. The Late Closure, Recency, and Locality Principles favor attaching an ambiguous modifier to the most recent possible site because distant sites require more memory resources. In the same vein, The Active Filler and Minimal-Chain Principles favor shorter wh-dependency over longer wh-dependencies and postulate a gap for a dislocated wh-filler as soon as the wh-filler is identified.

The nature and the functions of WM resources have been explored in various studies that were couched in a different theory of WM. In the following section, I will review some of the studies in reference to the theoretical assumptions they carry for the relation between WM and sentence processing.

3.4.3. The nature of verbal WM resources: i. Single Resource Theory versus Separate Sentence Interpretation Resource Theory

As a cognitive task/function, language comprehension requires WM because, regardless of whether language is written or spoken, it requires temporary storage of discontinuous parts of the input and processing (or computation) of information for language to be understood. The concept of WM resources or capacity for temporary storage and processing of information in language comprehension has been discussed within two dichotomous theories: (1) the Single Resource (SR) theory (King & Just, 1991; Just & Carpenter, 1992; Miyake, Carpenter, & Just, 1994), and (2) Separate Language Interpretation Resources (SLIR) theory (Caplan & Waters, 1995; 1996; Waters & Caplan, 1996; Caplan, Waters, & Dede, 2007). The SR theory considers WM a unitary construct (as originally defined by Baddeley &Hitch, 1974) in which language processing relies on the same or overlapping pool of resources as other cognitive tasks. That is, there is one single pool of resources for all verbally-mediated tasks. The SSIR theory, on the other hand, claims that the verbal WM system is a modular construct composed of at least two subsystems, which are devoted to verbal tasks: (1) a specialized verbal WM system that supports specific aspect of language processing which is referred to interpretive processing, and (2) a more general working memory system that supports other aspects of language processing, which is referred to post-interpretive processing. The interpretive processing includes the processes of recognizing words, and appreciating their meanings and syntactic features; constructing syntactic and prosodic features; and assigning thematic roles, and other aspects of propositional and discourse-level semantics. The post-interpretive processing includes the use of the extracted meaning to accomplish other tasks such as storing information in the long-term semantic memory, reasoning, planning actions and other functions (Caplan & Waters, 1999, p. 78).

Studies testing these theories used two approaches in their investigations: (1) an individual-differences approach, and (2) a dual-task approach. In the individual-differences approach, participants are divided into two or more groups on the basis of their performance on some form of verbal WM task (e.g., reading span task), and tested on syntactic structures of varying complexity (e.g., subject- and object-extracted relative clauses). In the dual-task paradigm, on the other hand, participants perform two tasks simultaneously: (1) processing sentences of high and low syntactic complexity (e.g., subject- and object-extracted sentences), and (2) at the same time, they are asked to remember a set of nouns or arithmetic digits (i.e., external memory load).

The underlying assumption of the individual-difference approach is that syntactic complexity should interact with the group-type (i.e., low-, medium-, high-span) if there is one single generic pool of resources for all verbally-mediated tasks (King & Just, 1991; Just and Carpenter, 1992). That is, low-span subjects should have more difficulty in processing sentences of high complexity than high-span subjects because complex sentences require more WM resources.

In the case of the dual-task approach, syntactic complexity should interact with the difficulty of the external load (e.g., storing a list of words) if both tasks rely on the same pool/overlapping pools of verbal WM (Just and Carpenter, 1992). If there are at least two pools of resources, one of which is dedicated to syntactic processing alone, and the other one dedicated to post-interpretive processing, then, there should be no interaction between the syntactic complexity and the group-type or the difficulty of the external load (Caplan and Waters, 1999).

The first suggestive evidence for the SR theory comes from King and Just (1991), who divided participants into high-and-low span groups based on a reading span task developed by Daneman and Carpenter (1980) and tested them on self-paced word-by-word reading task involving relative clause sentences with subject and object extractions as in (20a-b) respectively (Just & Carpenter, 1991, p.584)

20. a. The reporter that attacked the senator admitted the error publicly after the hearing.

b. The reporter that the senator attacked admitted the error publicly after the hearing.

They reported a main effect of span (low-spans were slower than high-spans), and a main effect of extraction type (object-extraction is slower than subject extraction, but there was no interaction between syntactic complexity and the group-type and the region where the syntactic processing difficulty increased. However, this study has been questioned due to insufficient statistical report.

In the same vein, Just and Carpenter (1992) replicated Ferreira and Clifton's (1986) study with high-and low-span participants, using sentences involving Main-verb/Reduced-relative clause ambiguity as in (21a-b) (Just & Carpenter, 1992, p. 127) to test the assumption that people with low WM capacity may not have the capacity to entertain non-syntactic information during syntactic computations, but those with high WM capacity use non-syntactic information in initially interpreting syntactic ambiguity.

21. a. The evidence (that was) examined by the lawyer shocked the jury.

b. The defendant (that was) examined by the lawyer shocked the jury.

Recall that Ferreira and Clifton (1986) reported slow reading times for all participants at the disambiguating *by*-phrase independent of the plausibility information (the animacy of subject NP) and suggested that animacy information was not used during the initial stage of parsing. This was interpreted as evidence for the modularity of syntactic information (cf. Trueswell et al., 1994). Just and Carpenter (1992) have found that in reduced relative clauses, high-span participants were faster than low-span participants in reading the *by*-phrase in the sentence with inanimate subject as in (21a). This result was interpreted as evidence for the claim that high-span readers are sensitive to the pragmatic information during the first-past syntactic analysis, but low-span participants are not. In other words, their finding confirmed the assumption that people with high WM capacity can activate both syntactic and non-syntactic information in parallel during the initial analysis of syntactic ambiguity. These results were also questioned due to a lack of interaction among syntactic complexity, group-span and the critical region.

Within the dual-task approach, King and Just (1991) tested the assumption that comprehending more complex sentences and maintaining an external load such as a series of words or digits in memory require more processing resources from the same pool. According to the SR theory, the difficulty that low-span participants will experience in processing complex sentences will be exacerbated by a concurrent external memory load. In a self-paced reading experiment, high-, medium-, and low- span subjects read object-and subject- extracted relative clause as in (20a-b) above, while retaining one, two or three sentence-final words in memory. For the recall task, King and Just reported significant interaction between the group and size of memory load but no interaction among group, syntactic complexity and memory load. The high-span subjects were found to be significantly better than the low-span subjects in accuracy. Overall, object extraction was more difficult than subject extraction. However, none of the interaction was significant.

Caplan and Waters (1999) replicated King and Just's (1991) and Just and Carpenter's (1992) study, using a variety of methods and large participant pools, but they were not able to demonstrate the main effect or the required interaction among the group-type, sentence type and the syntactically critical region.

Furthermore, MacDonald, Just, and Carpenter (1992) tested the SR hypothesis, conducting an experiment with sentences containing a main verb and a reduced relative clause ambiguity such as (22a-b) (MacDonald et. al., 1992).

22. a. The experienced soldiers warned about the dangers before the midnight raid.

b. The experienced soldiers warned about the dangers conducted the midnight raid. They found ambiguity effect for reduced relative clauses (as in 22b), compared to unambiguous control sentences (as in 22a), but there were no group effects for low-, mid- and high span participants. They unexpectedly found a memory span effect for main verb continuation, such that high-span participants were slower than low-span participants during the last word of the sentence "raid". This effect, though observed at an unexpected place beyond the ambiguous region, is interpreted as a result of high-span participants' maintaining multiple interpretations in parallel. Holding multiple interpretations in parallel takes longer time, Caplan and Waters (1999) replicated this study but could not find the same effect. Caplan and Waters (1996) also conducted a series of experiment to test the SR theory using a dual-task approach, where participants were asked to perform self-paced reading/listening while maintaining a memory load (usually, a string of digits). However, they did not observe on-line interactions or main effects for group-spans, syntactic complexity and memory load. Based on these results, they proposed that there is an independent pool of verbal WM resources dedicated to on-line sentence processing (see for Caplan and Waters, 1999; Caplan et al., 2007 for a review).

Caplan and Waters (1999) reported some data from neuropsychological studies conducted with various patients' populations such as patients with short term memory disorders (Waters, Caplan, & Hildebrandt, 1991), and patients with limitations in central the executive (i.e.., patients with dementia of the Alzheimer type (DAT)) (Rochon, Caplan, & Waters, 1994; Waters, Caplan, & Rochon, 1995), patients with Parkinson's disease (PD) (Waters et al., 1995), and patients with Broca's aphasic, who have reduced resources for syntactic processing (Caplan & Waters, 1996). For example, one of the patients who has an intact long-term memory, but a specific auditory verbal short-term memory impairment (i.e., rehearsal and storage are impaired) was tested on garden-path and non-garden-path sentences. She demonstrated a reading span of 1 when she was tested on the Daneman and Carpenter task, but she had no difficulty in understanding the sentences used in the task, when they were presented in isolation. She performed as well as normal individuals on garden-path sentences.

In another study, 22 DAT patients, who have intact functioning of the rehearsal and phonological storage, but have impairments on tasks requiring central executive functions were tested on syntactically complex sentences in a sentence–picture matching task (Rochon et al., 1994). The prediction was that if the SR is correct, then the DAT patients should have particular difficulty with syntactically more complex sentences, and they should perform differentially poorly on the more difficult syntactic structures when a concurrent memory load is imposed. Rochon et al. reported a main effect for group (i.e., the DAT group performed more poorly than controls that consist of education and age-matched healthy individuals) and a group-by-sentence-type interaction. Analysis of interaction showed that the DAT patients did not perform more poorly on the syntactically more complex sentences, but rather, that their performance was poorer than controls on sentences with two propositions. Similar results were obtained by Waters et al., (1995) with PD patients having limitations of central executive tested in a similar sentence-picture-matching task. The performance of PD patents differed from that of controls on the sentence comprehension task, but only on sentences containing more propositions. These results suggest that the DAT and PD patients do not have impairments in structuring sentences syntactically. In addition, the number-of-proposition effects arises at a stage of processing that shares resources with WM tests such as reading span (Caplan & Waters, 1999; p. 91). This suggests that such an effect is observed at the post-interpretive stage of sentence processing.

The last study to note comes from Broca's aphasics with reduced syntactic processing (Caplan & Waters, 1996). The SR theory assumes that syntactic complexity effects will be increased in aphasic patients under a concurrent verbal memory load because of shared WM resources. In contrast, the SSIR theory suggests that these pools are separate and therefore aphasic patients will not show an increase in syntactic complexity effects under a concurrent memory load condition. In their sentence-picture matching task, they tested 10 aphasic patients in no-interference and concurrent load conditions (storing digits). They reported that aphasic patients showed larger effects of syntactic complexity in a condition without a concurrent external load. However, this effect was not exacerbated by the addition of memory load. They found that in the digit-recall data, there was an effect of number of propositions but not of syntactic complexity.

In sum, Waters and his colleagues' findings from neuropsychological studies favor the SLIR theory, which suggests two separate pools of resources for verbal WM.

More recent studies by Gordon, Hendrick, & Johnson (2001), Gordon, Hendrick, &Levine (2002), and Fedorenko, Gibson, &Rohde (2006; 2007) have provided evidence for the SR theory in a set of dual-task experiments in which the external load was manipulated. For example, Gordon et al. (2002) argued that the load manipulation used in the previous dual-task experiments (e.g., increasing the number of memory-items in the digit span task) was not the right one for the purposes of assessing the nature of verbal WM resources in sentence comprehension. They suggested that WM capacity in language processing should be conceptualized not in terms of the number of items that must be kept active in memory during the comprehension process, as has been suggested by Daneman and Carpenter (1980), King and Just, (1991) and Lewis (1996), but rather in terms of the amount of the interference produced by the items that must be kept active in memory. Gordon et al. tested the SR theory of verbal WM for sentence comprehension using a novel dual-task paradigm in which participants read sentences of high and low complexity (e.g., subject- and object-extracted cleft sentences), which contain either occupations as in (23a-b) or personal names as in (24a-b) (Gordon et al., 2002, p. 427)

23. a. It was the dancer that liked the fireman before the argument began.

b. It was the dancer that the fireman liked before the argument began.

24. a. It was Tony that liked Joey before the argument began.

b. It was Tony that Joey liked before the argument began.

At the same time, participants were asked to remember a set size of three words which could be either occupations (e.g., poet, cartoonist, and voter) or a personal name (e.g., Greg, Andy, and David). This design yielded two match conditions: memory-nouns and sentence-nouns in the same category, and memory-nouns and sentence-nouns from different categories. At the end of each sentence, participants were asked to answer a comprehension question about the content of the sentence and recall the words from the memory task. Gordon et al. hypothesized that the similarity between the memory-nouns and the sentence-nouns might affect the more complex sentences (object-extracted clefts) to a large extent. Within a similarity-based interference[5] framework, Gordon et al. reported that in cases where memory traces of the memory-nouns are similar to the memory traces of the relevant antecedent, interference takes place. That is, it is harder to identify the relevant antecedent among all the available memory traces. They further hypothesized that these effects might be larger in object-extracted clefts due to a higher memory demand posed by these structures, compared to subject-extracted clefts.

[5]Similarity-based interference (e.g., Hintzman, 1986) refers to interference effects in sentence comprehension which apply at the retrieval stage of the memory processes and are conceptualized in terms of an overlap in retrieval cues. Specially, it has been argued that with an increase in the overlap in retrieval cues for different items in memory, the cue-to–target strength for any individual item decreases.

Gordon et al. found a significant interaction between syntactic complexity and the noun type match in comprehension accuracy. That is, there was a larger difference between subject-and object-extracted clefts for the matched conditions than for the non-matched conditions. They interpreted these results as evidence for the SR theory.

This was the first report of an interaction between syntactic complexity and memory load in a dual-task paradigm in which the two tasks did not interrupt each other.

Fedorenko et al., (2007) reported similar findings from a series of dual-task experiments (Exp. 1 and 2) in which participants read syntactically complex sentences and at the same time, they were required to perform a series of simple additions. The on-line addition task was similar to on-line sentence comprehension in that an incoming element– a number- must be integrated into the representation constructed thus far: the working sum. The results of these experiments showed a significant interaction between syntactic complexity and arithmetic complexity in the critical region of the linguistic materials, where syntactic complexity was manipulated between subject-and object-extracted relative clauses. This finding was interpreted as evidence for a WM capacity in which syntactic integration and arithmetic integration rely on overlapping resource pools.

While the debate between the SR theories and the SLIR theory still continues, MacDonald and Christiansen (2002) proposed an alternative view of WM in language processing, which was inspired by a connectionist view of language comprehension. According to this view, individual differences in comprehension stems from differences in skills and experience with language. They argue that WM tasks which measure participants' WM capacity are simply different measures of language processing because there is no linguistic WM capacity separate from linguistic representations and processes.

Within the connectionist framework, the processing of input is achieved through the passing of activation through a multi-layer network. In this framework, the network's capacity to process information varies as a function of input (e.g.., whether the material is complex or simple), the properties of the network (how activation is passed through weights, etc.) and the interaction of these properties- how much the network has experienced similar input before. WM is assumed as the network itself. It is not a separate entity that can vary independently of the architecture and of experience that governs the network's processing efficiency. Individual differences emerge from experience rather than variation in separate capacity (see MacDonald and Christiansen, 2002 for a detailed review, but cf. Caplan & Waters, 1999; Roberts and Gibson, 2002).

In sum, the SR, SSIR and the connectionist framework make different predictions as to the relationship between language (e.g., syntactic) knowledge and WM capacity and individual differences observed in WM capacity. According the SR theory developed by Just and Carpenter (1992), having a low WMC will reduce the resources available for sentence processing and make it less efficient, therefore the SR theory predicts significant correlations between measures of WMC and measures of sentence processing efficiency. In research which uses the individual differences approach in which performance on WM task serves to divide subjects into high-, medium-, and low-WMC, and sentence-processing performance is measured, the SR theory predicts that there will be a main effect of WMC in experiments of this type, with high-capacity subjects performing better on the sentence processing task than low-span subjects. There will be an interaction between syntactic complexity and WMC: sentence complexity will affect low-capacity subjects more than high-capacity subjects, and the differences between performances of low- and high-capacity subjects with complex sentences will be greater than the differences between the groups with simple sentences. Like the SR theory, the connectionist theory of MacDonald and Christiansen predict significant correlations between measures of WMC and measures of linguistic tasks, or/and a main effect and interaction between syntactic complexity and WMC, due to the differences between high- and low-capacity subjects in skills and experience with language.

According to the SLIR theory of Waters and Caplan (1999), performance on general verbal WM tasks will significantly correlated with language processing efficiency. There may be a main effect of group in dual-task experiments, due to other aspects of tasks demands (i.e., difficulties low-capacity subjects have in dividing their attention in the task), but there will be no interaction between syntactic complexity and WMC: the differences between performances of low- and high-capacity subjects with complex sentences will not be greater than the differences between the groups with simple sentences.

Improving the methods to measure WM has been the focus of recent research into WM due to the presence of inconsistent results obtained from earlier studies (see Roberts and Gibson, 2002; Conway, Kane, Bunting, Hambrick, Wilhelm, & Engle, 2005 a detailed review). Below, I will briefly discuss some of the methods used to measure WM.

3.4.4. Measuring Working Memory Capacity: Working Memory Span Tasks

There are several WM tasks, which are commonly used to assess WM capacity. These are reading span (Daneman & Carpenter, 1980; Waters & Caplan, 1996), operation span, (Turner & Engle, 1989), counting span (Case, Kurland, & Goldberg, 1982), spatial

WM span (Shah & Miyake, 1996; Kane, Hambrick, Tuholski, Wilhelm, Payne & Engle, 2004). Most of the WM tasks such as reading span, operation span or spatial-rotation tasks were designed on the basis of Baddeley and Hitch's WM model (1974), which involves storing various intermediate products of a computation while simultaneously processing new information. That is, WM keeps task-relevant information active and accessible in memory during the execution of complex cognitive tasks such as decoding of a complex sentence, mentally rotating an unfamiliar geometric figure, or solving a difficult reasoning problem. WM Capacity measures were, therefore, constructed in such a way that they not only measure information storage and rehearsal (as do simple measures of short term memory capacity, such as digit span or word span), but also the simultaneous processing of additional information (e.g., Case et al., 1982; Daneman and Carpenter, 1980; Turner and Engle, 1989). These WM span tasks interleave the presentation of to-be-remembered target stimuli, such as digits, words and letters, with the presentation of a demanding secondary task, such as comprehending sentences verifying equations or enumerating an array of shapes. WM span tasks predict complex cognitive behavior across domains, such as reading comprehension problem solving, and reasoning.

The reading span task was the first task developed to measure the storage and processing functions of WM (Daneman and Carpenter, 1980). In this task, participants read sentences and verify the logical accuracy of the sentences, while trying to remember words, either the last word of each sentence presented, or one independent word for each sentence presented. The sentences are presented in groups that typically range in size from two to six (a group of sentences is referred to as one item). Word recall is prompted at the completion of an item. Later versions of this task have included either an independent word or a letter as the target item to-be- remembered. In addition, Turner and Engle (1989) have showed that they could predict reading ability with a WM span task which involves solving mathematical operations while trying to remember words. This is called an *operation span task*. It is similar to a *reading span task* in terms of task demands. Nevertheless, it includes mathematical equations to be solved, instead of sentences to be comprehended (see Engle et al, 1992).

In the original version of the Daneman and Carpenters (1980)' reading span task, subjects are required to read aloud, at their own pace, the sentences presented on index cards. At the same time, they try to keep in mind the last word of each sentence for later recall. After a series of sentences, the subject recalls the target words in the order in which they are presented. There are 15 items, each consisting of two, three, four, five and six sentences that are composed of 13-16 words in length. They are presented in ascending order (i.e., from the smallest to the largest). Increasingly longer items are presented until the subject fails to recall all 3 items of a given size. At this point, the experiment is terminated. A subject's reading span is the level at which s/he can correctly recall 2 of the 3 items. For example, if a subject were to successfully recall 2 out of 3 two-word items, the experimenter would continue for the subject to attempt 3 two-word items. If the subject were then to successfully recall only 1 out of 3 three-word items, the experiment would terminate, and then the subject's reading span would be 2.

As mentioned earlier, there are several WM tasks used to assess WM capacity such as reading span, operation span, counting span, and spatial WM span tasks. It is not always easy to decide which one needs to be used. In some recent studies, it has been suggested that scores on commonly used WM tests may not in fact be highly correlated with each other and that the different tests may be differentially sensitive to processes such as storage, response time, rapid stimulus manipulation time and other central functions (e.g., Lehto, 1996; Waters & Caplan, 1996). Thus it was suggested that different WM tests are sensitive to different components of WM.

It has been argued that all of the above-mentioned measures of WM capacity, including operation span, reading span, and counting span, suffer from the fact that no single task is a perfect measure of the construct it ostensibly represents (Conway et al., 2005, p. 780). For instance, the operation span task measures WM capacity but, most likely, also taps mathematical ability, motivation and world knowledge among other things. Similarly, the reading span measures WM capacity, but certainly also verbal ability. Also, any of these WM tasks are not perfectly reliable (see Conway et al., for a detailed discussion). Thus, despite their validity and reliability in assessing WM capacity, WM span tasks are still not perfect or process pure. Therefore, the best research strategy is to administer multiple WM span tasks and then use the average of all scores on all tasks as the measure of WM capacity. That means a latent variable from the common variance among counting span, operation span and reading span need to be obtained (Conway, Kane, & Engle, 2003; Kane et al., 2004). The latent variable represents only the variance that is common among the tasks and removes task-specific factors.

Another concern is the methods which are used to examine the relation between WM and language comprehension. There are two approaches to this, namely, the correlation and the individual difference approach.

In correlation studies (e.g., Daneman & Carpenter, 1980; Waters & Caplan, 1996), participants are given a variety of tasks of WM. Correlations, multiple regression, and factor analysis are used to determine what variables are related to the cognitive function of interest. This approach has the advantage of identifying relations among large numbers of variables, and the extent to which different tasks contribute to common and unique variance to the measures of interest. However, it has some disadvantages in the sense that multiple comparisons require large numbers of participants in order to be reliable; and also with large numbers of participants, statistically significant correlations may account for only a small amount of variance on a given test.

In the individual differences approach (e.g., Just and Carpenter, 1992), participants are given a test designed to measure WM capacity, such as reading span, and then divided into three groups on the basis of their scores (i.e., a high-, medium-, and low-span groups). Usually the medium-span group is omitted from further analyses. The high-and low-span groups are compared, using ANOVAs, on other measures of interest, such as reading speed and comprehension. The groups are treated as though they are independent and homogenous and are compared to see whether they perform differently on the secondary task. This approach has some significant disadvantages. One of them as that it ignores large amount data due to the exclusion of the data coming from the medium-span group. Another problem is the overestimation of the relation between variables due to the selection of two extreme groups, which eliminates participants with scores near the mean of the sample, however, keeping participants whose scores have larger deviations from the mean. As a result, the correlation coefficient is likely to be larger with the extreme groups. Third, the choice of a cut-off point seems arbitrary if the inspection of a scatter plot does not suggest any natural grouping of the data. Some studies use a median-split design for dividing participants into groups. However, this has two problems, one of which is that there is no reason to categorize participants when the entire continuum has been sampled. The other problem is that miss-classification of participants is more likely in a median split design than in an extreme group design (Conway et al., 2005).

Among the methods of analyses, the correlational approach should be preferred over the extreme-group design due to significant problems associated with the latter. Therefore, as can be seen in the next chapter, in this study, the correlational approach has been used to test the relationship between WMC and sentence processing ability.

In the next section, I will first give an overview of the theories in L2 sentence processing, by making reference to the studies conducted to test them and discussion of findings regarding the processing of wh-dependencies in the L2. Finally, I will discuss the role WM in L2 sentence processing in general and L2 processing of wh-dependencies in particular.

3.5. Theories of L2 sentence processing

3.5. 1. Introduction

Research into L1 sentence processing has provided critical information about the human sentence processing mechanism (i.e., the parser) and helped us identify the relationship between the parser and the grammar as well as the type of information resources and processing strategies it applies to processing L1 input in real time. L1 sentence processing research has also paved the way for a new framework in L2 acquisition research, which integrates psycholinguistic methods to examine on-line parsing in the L2 to ultimately identify potential native-non-native differences in sentence processing.

L2 input has always been assumed to play a crucial role in the acquisition of native-like competence in the L2 (e.g., Krashen, 1987; Sharwood-Simth, 1993; White, 1991; Van Patten, 1996, 2004). For example, within the Generative framework, White, (1991) argues that some input-processing mechanisms may indeed be required for L2 acquisition. In the same vein, Fodor (1999) notes that parameter-resetting in the L2 takes place only after significant linguistic analysis. However, little is known about how L2 learners process L2 input in real time. If this is clearly identified, much insight can be gathered not only for understanding the L2 acquisition phenomenon but also for L2 instruction/teaching.

In recent years, there is an increase in the number of L2 studies that use on-line behavioral measures (e.g. self-paced reading and eye-tracking, cross-modal priming) and neurophysiological measures (e.g., ERP, functional magnetic resonance imaging (fMRI), and PET) to identify how L2 learners process L2 input in real-time. Some of the major questions that L2 researchers have investigated involve; (1) what are the architectures, mechanisms and representations that underlie L2 processing? (2) What types of processing strategies and information sources do L2 learners use in on-line processing of L2 input? (3) How does the on-line processing of two or more languages take place with respect to each other? (4) Are there qualitative and quantitative differences between processing sentences in the L1 and the L2? (5) What role (if any) does the L1 play in L2 sentence processing?

The following discussion will start with L2 studies examining syntactic ambiguity resolution and then it will focus on studies that explore L2 processing of wh-dependencies. The role of WM in L2 sentence processing will also be relevant in this discussion. Therefore, the final section of this section will look at the role of WM in L2 sentence processing in general and processing wh-dependencies in particular.

3.5.2. Studies in L2 sentence processing

Much recent L2 research has focused on differences between native and non-native speakers in ultimate success in L2 acquisition. Within the generative framework, a considerable number of off-line studies have so far tested L2 learners of different languages on a number of syntactic structures assumed to be part of Universal Grammar (UG) to examine whether UG is available in adult L2 acquisition (e.g., Bley-Vroman, Felix, & Ioup, 1988; Bley-Vroman, 1989; Schachter, 1989; 1990; Johnson & Newport, 1991; White, 1991; Martohardjono & Gair, 1993; White & Juffs, 1998). Some of these studies seem to find evidence for the claim that the UG remains accessible to L2 learners, whereas others have provided evidence against this claim. Thus, the UG-availability issue is still disputed. Nevertheless, the very same issue is now being discussed in the context of on-line language comprehension and processing to identify whether L2 learners can ultimately reach native-like processing in the L2 and to explore what implications would this have for L2 grammatical representation.

One important line of initial L2 sentence processing research includes studies that have examined structural ambiguity resolution in the L2. Identifying potential differences between native and non-native speakers in terms of the strategies they employ in resolving ambiguities is one of the main aims of these studies. Similarly, the question of whether or not processing strategies used by L2 learners better suited for L1 input than for L2 input is also an important issue discussed in L2 sentence processing research (e.g., Frenck-Mestre &Pynte, 1997; Fernandez, 1999; Dussias, 2003; 2008; Frenck-Mestre, 2002; 2005; Juffs & Harrington, 1996).

To this aim, L2 studies, following L1 processing research, examined the structures that include syntactically ambiguous constructions such as main verb/reduced relative ambiguities (e.g., *The horse raced past the barn fell*) (e.g., Juffs, 1998); relative clause and PP attachment ambiguities involving complex *genitive* (NP-of-NP) antecedents (e.g., *Someone saw the servant of the actress who was on the balcony*) (e.g., Frenck-Mestre & Pynte, 1997; Dussias, 2003; Felser et al., 2003; Papadopoulou and Clahsen, 2003); and direct object/sentential complement ambiguities as (e.g. *After Bill drank the water proved to be poisoned*) (e.g., Juffs and Harrington, 1996; Juffs, 1998a; 1998b; Felser & Roberts, 2004; Juffs, 2004; Dussias, 2008).

Unfortunately, these previous on-line studies have not provided conclusive evidence concerning the question of L1-L2 differences in processing strategies and information sources used in on-line L2 ambiguity resolution (see Clahsen & Felser, 2006 for an overview). Some studies have shown that L2 learners use processing strategies similar to those used by native speakers, whereas there are also studies which suggest that the L2 parsing strategies are different from those of L1 (see Clahsen & Felser, 2006; Frenck-Mestre & Pynte, 1997 (Exp.1); Fernandez, 1999; Papadopoulou and Clahsen, 2003). These studies mostly concern with lexical biases in on-line L2 relative clause (RC) attachment, which is also a controversial issue in L1 sentence processing.

Differences between native and nonnative speakers have been discussed under the *Shallow Processing Hypothesis*' proposed recently by Clahsen and Felser (2006), who argue that L1 sentence processing is fundamentally (and permanently) different from L2 sentence processing in that syntactic representations that L2 learners construct while processing L2 input are shallower and less detailed than those constructed by L1 speakers. Furthermore, in this view, it is suggested that native speakers use structure-driven strategies and syntactic information during sentence processing, while L2 learners employ lexical-semantic and pragmatic information. However, this claim has been found to be too premature and too strong in a developing field, which has a limited number of research studies on a small range of syntactic structures (e.g., Dussias & Cramer Scaltz, 2009).

Due to a limited number of on-line studies, we do not yet have conclusive evidence for or against the claim that L2 speakers are fundamentally different from native speakers even in ultimate L2 state. Available data are not yet sufficient for us to clearly identify possible processing problems (and their reasons) that L2 learners might have in processing sentences in the L2. Therefore, it is believed that the present study will make a significant contribution to the field by offering new set of on-line data from L1 Turkish and L1 Spanish learners' processing of wh-questions in L2 English. To set up a background for the present study, I now will give an overview of L2 studies which examined the processing of wh-dependencies in L2.

3.5.3. Processing wh-dependencies in the L2

Like ambiguity resolution, wh-dependencies have been the subject of the recent studies in L2 sentence processing. Nevertheless, these sentence processing studies were inspired by earlier off-line work on the acquisition of wh-movement. For example, Schachter & Yip (1990) compared, through off-line tasks, native and non-native judgments on long distance wh-dependencies (e.g. Schachter &Yip, 1990) and results revealed that both groups demonstrated more difficulty in comprehending subject extractions from nonfinite clauses, compared to object extraction.

The first attempt in studying the same phenomenon in on-line sentence processing has come from Juffs and his colleagues (e.g., Juffs &Harrington, 1995; 1996; Juffs, 2005; White & Juffs, 1998) to examine whether the difference between the native and nonnative speakers in ultimate attainment are due to a processing problem (i.e., a subject and object asymmetry in grammatical wh-extractions) or a deficit in L2 grammar.

Using a self-paced word-by-word reading method, Juffs and Harrington (1995; 1996) have tested L2 wh-processing in an on-line grammaticality judgment task in advanced Chinese L2 learners, who do not have overt wh-movement in their L1. The task involved a set of grammatical subject- and object- extractions from finite and nonfinite clauses, and a set of ungrammatical wh-questions violating constraints on subjacency as in (25-32) (Juffs & Harrington, 1995, p. 496):

*25. Who does Sam deny the story that he kissed? (Complex NP island)

*26. What does Sam see the man who stole? (RC island)

*27. Who did you meet Tom after you saw? (Adjunct island)

*28. Who did a story by please the children? (Subject island)

29. Who did Ann say likes her friend? (finite, subject extraction)

30. Which man did Jane say her friend likes? (finite, object extraction)

31. Who does Tom expect to fire the manager? (nonfinite, subject extraction)

32. Who does Tom expect to fire? (non-finite, object extraction)

Following Pritchett's Generalized Theta Attachment Theory (GTA) (Pritchett, 1992), which states that "every principle of the syntax attempts to be maximally satisfied at every point during parsing" (p.155). The authors' assumption is that Chinese learners of English will have more difficulty with subject extractions (particularly with subject extractions from nonfinite clauses) compared to object extractions because long-distance subject extraction requires more changes in theta roles and Case assignment. In finite clauses, participants are expected to experience a garden path due to encountering a finite verb as in *likes* following the main verb *say* as in (29), which will be reflected by longer RTs at the embedded verb. In nonfinite clauses, the wh-phrase *who* can be interpreted as the antecedent of three potential traces (i.e., the object trace of the matrix clause, PRO and object trace of the embedded clause), therefore, it will at least require three reanalyses before reaching the correct analysis of *who* as the subject of embedded nonfinite clause finite as in (31).

Accuracy results from the full-sentence condition confirmed these predictions in that L2 speakers were as accurate as native speakers in rejecting ungrammatical extractions with island violations as in (25-28), but they were significantly less accurate than native speakers in accepting grammatical extractions. More specifically, unlike native speakers, they were significantly less accurate on subject extractions from grammatical finite clauses (29), compared to object extraction from finite clauses (30). In nonfinite grammatical extractions, both L2 speakers and English native speakers were significantly less accurate on subject extraction than object extraction. Results from RTs showed that L2 learners were significantly slower than English native speakers across all structures. Also, all participants, in particular Chinese learners, spent significantly longer RTs for subject extraction from nonfinite clauses, suggesting that subject extraction from nonfinite clauses is the most difficult of all types. Results from accuracy and RTs suggest that L2 learners have difficulty in processing subject extraction from finite as well as nonfinite clauses while English native speakers have difficulty with subject extraction from nonfinite clauses. The results of self-paced word-by-word reading task revealed that the locus of the difficulty in finite clauses was the embedded verb '*likes*' (29), where L2 learners expected an object gap, but found a finite verb and thus experienced a filled-gap effect. In nonfinite clauses, it was the embedded object NP '*the manager*" (31), where they expected an object trace. Juffs and Harrington argue that these results support the claim that advanced L2 learners have the abstract knowledge of overt wh-movement with constraints imposed by the Subjacency Principle, despite the fact that they do not have overt wh-movement in their L1. Also, they confirm previous finding of an asymmetry between subject and object extraction, a pattern observed both in L2 learners and native speakers. The difference observed between native and the Chinese learners has been accounted for by L1 influence. Chinese learners do not have overt wh-movement in the L1 and are, therefore, never required to reanalyze empty categories in wh-constructions and juggle theta roles and Case in A-bar chains at S-structure (Juffs & Harrington, 1995, p. 510). Overall, in line with the predictions of the GTA theory, these results show that a gap reanalysis is problematic even for native speakers. Thus, these findings support the view that it is parsing performance that underlies the inaccuracy of Chinese learners on grammatical sentences. However, this study has a major methodological limitation with respect to the number of L2 groups.

In a follow-up study, Juffs and Harrington (1996) compared L2 learners' processing of wh-traces (embedded finite and nonfinite clauses) and garden path sentences. They assumed that garden path sentences share with wh-extractions important underlying grammatical characteristics that are manifested in processing difficulties experienced by both L1 and L2 participants. That is, GP sentences are often ambiguous and therefore cause native speakers to fail to get the correct reading and consequently leading them to reject the sentences as ungrammatical. Like GP sentences, wh-extraction, in particular subject extraction, forces the parser to make a similar reanalysis in sentence parsing.

Juffs and Harrington tested the participants in Juffs and Harrington's (1995) study on grammatical subject and object extractions from finite and nonfinite clauses in (29-32 above) and 'GP Optional Transitive' sentences in (33) and non-GP Intransitive sentences in (34) as the control condition. Also, they included GP sentences as in (35), having verbs with two internal arguments, goal and preposition, along with a non-GP control condition as in (36) (Juffs & Harrington, 1996, p. 297).

33. ¿After Bill drank the water proved to be poisoned.

34. After Sam arrived the guests began to eat and drink.

35. ¿Sam warned the student cheated on the exam.

36. Jane knew her mother hated Tom.

Findings showed that L2 learners dramatically failed to correctly accept subject extraction from nonfinite clauses. Results from error data showed that there is a similarity between reading profile of L2 speakers, who incorrectly rejected grammatical subject extractions and that of the native speakers, who correctly accepted these sentences. Both groups spent longer RTs at the object NP following the embedded verb in the embedded clause, where they expected an object trace to associate with the wh-phrase. However, L2 learners' RTs continued to increase after the object NP while those of native speakers' decreased. This suggests that L2 learners did a reanalysis, which led them to make an incorrect judgment. The authors speculate that the Chinese learners who executed the most reanalyses also made the most errors.

Results from the GP and the non-GP sentences indicated that L2 learners and native speakers were more accurate on the non-GP sentences than the GP sentences, and that they experienced similar difficulties with the GP sentences. There were no significant differences between the groups on the GP structures. RTs for error data in the self-paced reading condition revealed that the locus of the difficulty in the GP sentences for both native speakers and L2 learners was the matrix verb which is the disambiguating region. This shows that L2 learners who reanalyzed the structure at the correct region were most successful at judging the GP sentences. Juffs and Harrington argue that RTs on embedded verbs whose subjects have been extracted via wh-movement and verbs that cause GP effects give rise to very similar reading profiles. The difficulty both accurate and inaccurate participants had with GP structures probably derives from the GP effect-a performance problem, because the sentences are grammatically correct. They conclude that the parallels in accuracy and word-by-word reading profiles in subject extraction and GP sentences, and parallels in structural reanalyses support the hypothesis that a parsing problem rather than a competence deficit underlies the difference between native and nonnative speakers in subject wh-extraction. In addition, they assume that Chinese learners are not used to multiple reanalyses of A-bar chain in which theta roles and case of traces change, which might increase the amount of difficulty L2 learners have in subject extraction.

However, Juffs (2005) has reported that the major concern of these studies is that the authors tested less than two L2 groups, which casts doubt on the evidence that the absence of wh-movement in Chinese was really the cause of the problems Chinese speakers were having with subject extractions from finite and nonfinite clauses. Comparison groups of L2 speakers with an L1 having overt wh-movement are also necessary because such speakers should not have as great a problem processing wh-movement. If they do, then the influence of the L1 will be less convincing as explain.

To explore the role of L1 in the processing of wh-dependencies, most recently, Juffs (2005) replicated Juffs and Harrington's (1995) study, with three L2 groups having typologically different L1s such as Japanese, Chinese, and Spanish. More specifically, the author examined whether the presence or absence of *wh*-movement in the L1 affects

L2 learners performance in processing of wh-extractions in the L2. In Chinese and Japanese, wh-phrases stay in situ whereas in Spanish they overtly move to the sentence initial position as in English. Juffs made the assumption that the L1 grammar and L1 processing strategies are closely linked. The L1 grammar transfers as well as L1 processing strategies, and that both can influence L2 processing. Using a self-paced reading time method, Juffs tested all the participants on an on-line grammaticality judgment task involving ungrammatical wh-extractions violating the Subjacency Principle as in (37a-c), and grammatical subject and object extraction from finite and nonfinite clauses as in (37d-g) (Juffs, 2005, p. 129).

37 a. *Who did Tom believe the claim that Ann saw_ at school? (Complex NP Island)

b. *Who did Tom hear the woman who saw _ on TV? (Relative Clause Island)

c. *Who did Ann meet the teacher after she saw _ last week? (Adjunct Island)

d. Who does the nurse know _ saw the patient at the hospital? (Finite subject)

e. Who does the nurse know the doctor saw _ in his office? (Finite Object)

f. Who does the boss expect _to meet the customers next Monday? (Nonfinite Subject)

g. Who does the boss expect to meet _ next Monday? (Nonfinite Object)

Results from accuracy data showed that L2 speakers correctly rejected ungrammatical wh-extractions at a level above chance, but they were significantly less accurate than English native speakers. The Japanese and Chinese learners had the lowest accuracy compared to Spanish learners, but they were not significantly different from each other. The accuracy results for grammatical wh-extractions indicated that all groups correctly accepted grammatical wh-movement beyond a chance level. Similar to the accuracy results of ungrammatical wh-extractions; all L2 groups were significantly less accurate than English native speakers. The Japanese and Chinese learners were less accurate than Spanish learners and English native speakers, but not different from each other. Among the four types of sentences, subject extraction from a finite clause was significantly more difficult than object extraction. No significant difference was observed between subject and object extraction from a nonfinite clause.

Results from RTs indicated that the locus of the difficulty in subject extraction from a finite clause is the embedded verb (e.g., *saw* as in (37d)) where the participants expected an object trace to integrate with the wh-phrase. Juffs reported that all groups including Chinese and Japanese, were significantly different from each other. Among the L2 groups, Japanese spent the longest RTs for the embedded verb. This suggests that they had the most difficulty with subject extractions. They were followed by Chinese and Spanish. In the case of nonfinite clauses, no significant difference was found either between extraction types, or among language groups except for the lower performance of the Japanese group. Juffs notes that this might be due to the fact that the Japanese group may not match with the other two groups in proficiency.

Juffs interprets these results as evidence for the claim that L2 learners have the knowledge of wh-movement in the L2, but they have processing problem with subject extraction from a finite clause but not from a nonfinite clause. L2 leaners'L1 background affects L2 learners' accuracy for grammatical wh-extractions, but marginally so for the ungrammatical extractions. That is, the absence of overt wh-movement in the L1 provides a disadvantage for all learners while the presence of wh-movement provides a clear advantage for judging wh-extractions in the L2 English. However, all of them have similar difficulties with parsing subject extraction. Juffs accounts for this difficulty triggered by seeing two finite verbs appearing next to each other in a sentence (37d) rather than an inability to make multiple reanalysis subject traces as suggested by Juffs and Harrington (1995). He proposes that the source of the processing problem at the juxtaposition of two tensed verbs (i.e., *know saw*) as in (37d) in embedded finite clauses might be finiteness because tense plays a role in Case assignment to a wh-chain and is marked by overt morphology in the verb in English. Juffs notes that these results imply that L2 learners have an ability to reassign features to wh-chains during processing performance.

Nevertheless, White and Juffs' (1998) similar study on on-line sentence processing revealed that L2 learners have more difficulty with the comprehension of subject extractions from both finite and nonfinite clauses. L1 speakers, on the other hand, have more difficulty with subject extractions from nonfinite clauses.

In another reading time study, Williams, Möbius, and Kim (2001) examined L2 learners' processing strategies in the formation of wh-dependencies in the L2. Using a 'stop making sense task' (Boland, Tanenhaus, Garnsey & Carlson, 1995), Williams et al., (2001) tested advanced Korean, Chinese, German learners and English native speakers on wh-dependencies in (38-39) (Williams et al., 2001, p. 536). They examined whether the filled gap effect is affected by the plausibility of the initial filler-gap assignment or not.

38. Which girl did the man push the bike into late last night? (Plausible-at-V)

39. Which river did the man push the bike into late last night? (Implausible-at-V)

In both sentence there is a potential gap after the verb *push*. According to the Filler-driven strategy, participants should hypothesize *girl* as the direct object of *push* in (38) and *river* as the direct object of *push* in (39). If plausibility of these assignments is computed, all participants should exhibit a filled-gap effect (a tendency to slow down at the post-verbal region) because the gap that they originally hypothesized turns out to be filled by an unexpected NP, in this case *the bike*. Participants in this self-paced reading experiment, read the sentences presented on a computer screen in a word-by-word fashion and pressed a stop button to indicate the point at which they thought the sentence had stopped making sense.

Results indicated that all groups made more ‘stop making sense decisions’ at the verb in the Implausible-at-V condition (39) than in the Plausible-at-V condition (38). They showed longer reading times on the post-verbal noun when the initial filler-gap assignment was plausible (38). These results show that all groups were analyzing the filler as the direct object of the verb and computing the plausibility of the filler as the direct object. This suggests that both native speakers and L2 learners use a filler-driven strategy and apply plausibility information during on-line processing. The similar results observed in native speakers and L2 learners also suggest that L2 learners’ L1 background does not affect their processing performance in the L2.

The authors observed that the native speakers showed a plausibility effect at the post-verbal determiner, the implausibility-at-V condition being slower than the plausible-at-V condition. They assume that the filler-gap implausibility and the disambiguating syntactic cue provided by the determiner might have triggered rapid syntactic reanalysis in the implausible-at-V condition. However, L2 learners did not show the same effect at the post-verbal determiner in the implausible-at-V condition that will trigger syntactic reanalysis in this condition. This difference between native speakers and L2 learners was interpreted as a slight delay in the use of plausibility information in L2 processing. In a follow-up off-line acceptability judgment task, they examined whether or not native speakers initiate reanalysis more rapidly than nonnative speakers. They found that L2 learners judge the plausible-at-V sentences unacceptable significantly more often than the implausible at-at-V sentences. They concluded that L2 learners demonstrate more difficulty recovering from an initial misanalysis than native speakers, particularly when this analysis is plausible.

In another ‘stop making sense’ task, Williams (2006) replicated the study of Williams et al., (2001) with Spanish, Italian, and Chinese learners to examine whether plausibility effects in nonnative sentence processing are delayed or whether there is some fundamental difference in the way in which plausibility affects performance though argument competition or a direct interaction with syntactic processing. He found that like native speakers, nonnative speakers, regardless of the nature of the L1, appeared to immediately compute the plausibility of the potential filler-gap relationship at the verb. Williams suggests that L2 processing is remarkably native-like.

Results from studies exploring the processing of wh-dependencies in the L2 provide evidence for on-line filler-gap integration, but are not conclusive in terms of whether L2 learners make direct association between a wh-filler and its subcategorizing verb or they associate wh-filler with its gap as suggest by filler-driven accounts (see sections 3.2.31 and 3.2.3.2) (e.g., Juffs and Harrington, 1995; Juffs, 2005; Williams et al., 2001; Williams, 2006). Marinis, Roberts, Felser, and Clahsen, (2005) argue that the nature of Juffs and Harrington’s or Williams et al.’s materials, does not allow us to distinguish between theoretical approaches that analyze wh-dependencies via wh-gaps as suggested by the Active Filler Hypothesis (e.g., Frazier & Clifton, 1989) or without wh-gaps as suggested by the Direct Association Hypothesis (e.g., Pickering & Barry, 1991). This is because in English, the trace is adjacent to the subcategorizing verb. Therefore, the slowdown observed in the post-gap region may also be due to the learners' attempt to link the wh-phrase directly to its subcategorizer in accordance with the Direct Association Hypothesis.

Nevertheless, Marinis et al., (2005) replicated Gibson and Warren's (2004) study, which looked at whether English native speakers make use of intermediate gaps during the processing of long-distance wh-dependencies and found evidence for the Active Filler Hypothesis, which suggest that a filler is reactivated cyclically to break up long distance dependencies into series of shorter dependencies. Marinis et al. tested advanced Chinese, Japanese, German, and Greek learners of L2 English on a self-paced reading experiment to examine whether L2 learners, like native speakers, postulate gaps during the processing of long distance wh-dependencies as in (40 and 41) (Marinis et al., 2005, p. 61).

40. The nurse who_i the doctor argued $é_i$ that the rude patient had angered $é_i$ is refusing to work late.

41. The nurse who_i the doctor's argument about the rude patient had angered é is refusing to work late.

Within the framework of gap-based accounts, the first sentence involves intermediate landing sites *(é)* for the fronted wh-pronoun, but the second sentence involves extraction across an NP and does not offer landing sites. The authors predicted that participants who postulate intermediate gaps should find it easier to integrate the wh-filler with its subcategorizer in the sentences as in (40). This should be reflected in their shorter reading times on the wh-filler's subcategorizer *'angered'* in (40) than in (41). RTs for the complementizer *that* in (40) should be longer than in the corresponding non-extraction condition. Results indicated that all participants, regardless of the nature of their L1, slowed down at the verb '*had angered'* in the two extraction condition as in (40), showing the extra processing cost in relation to filler-gap/subcategorizer integration at this point. However, only native speakers showed a significant interaction between extraction and phrase type at this region. That is, they spent shorter RTs for '*had angered'* in the sentence with intermediate traces (40), than in the corresponding condition (41), suggesting that the presence of intermediate traces facilitates filler-gap integration in long-distance wh- dependencies. The native speakers spent longer RTs on the complementizer *that* in (40) than in the corresponding nonextraction condition, which confirms the hypothesis that the filler is mentally reactivated at this point during parsing. However, in L2 groups, an interaction between extraction and phrase type at the region involving '*had angered'* was not observed. Marinis et al. interpreted the results as evidence that L2 learners did not mentally reactivate the filler prior to processing of the subcategorizing verb in either of the two extraction conditions, but they try to establish a direct link between the filler and its subcategorizer in both extraction conditions instead. Furthermore, having overt wh-movement in the L1 appears to have no influence in the processing of wh-dependencies in the L2. Overall, the authors have proposed that unlike English native speakers, L2 learners employ a lexically-driven strategy in which they form long filler-gap dependencies using direct lexical association rather than structure-based gap-filling, regardless of their L1 background.

In the same vein, in a cross-modal picture priming study, Felser and Roberts (2007) examined whether or not filler integration is facilitated by syntactic gaps in L2 processing and obtained results similar to the findings of Marinis et al. (2005). The authors tested advanced Greek learners of English on sentences as in (42) (Felser & Roberts, 2007, p. 17).

42. John saw [the peacock]$_i$ to which the small penguin gave the nice birthday present é$_i$ in the garden last weekend.

They predicted that advanced L2 learners should show antecedent priming effect at the position of the indirect object gap following the NP '*present*' if they use a filler-driven strategy. Results indicated that L2 learners did not show any structurally determined antecedent priming effect at the object gap position following '*present*', but adult native speakers of English did. This confirmed the hypothesis that L2 learners do not postulate intermediate gaps when processing long-distance wh-dependencies in their L2.

Taken together, the limited number of studies reviewed above does not provide a conclusive answer to the question of whether L2 speakers process wh-dependencies in their L2 in the same way as L1 speakers do. A number of issues are still unclear such as: whether L2 learners rely upon the same or different processing mechanisms in wh-dependencies in the L2; whether they make use of syntactic information or lexical, semantic, pragmatic information in the formation of wh-dependencies. Furthermore, little is known about the specific strategies (e.g., a gap-based or gapless strategy) they adopt to integrate a fronted wh-filler with its subcategorizer; and to what extent the L1 background of L2 learners constrains their performance on processing wh-dependencies in the L2.

More importantly, changes to theoretical accounts of constraints on movement within the generative grammar theory have affected the reasoning used in theories of L2 acquisition. Initially, island constraints were unified under the Subjacency Principle, according to which an element can cross one bounding node at a time. Many studies in L2 research were carried out based on this account. However, in the syntactic literature, it has been shown that there are parametric differences in bounding nodes, with some languages having S' (or CP) as a bounding node instead of S (or IP) (e.g., Rizzi, 1982). For example, extractions from wh-islands are acceptable in languages like Italian, Spanish and French. Alternatively, an extra landing site was proposed by Reinhart (1981). Based on research on wh-in-situ languages, Huang, (1982) has proposed a condition on extraction site (CED), which unified islands as the argument condition and adjunct condition on the basis that arguments are not governed by lexical categories like arguments. In a recent version of the CED (e.g., Nunes & Uriagereka, 2000; Muller, 2007) noncomplements are cross linguistically barriers to movement, and therefore subject and adjuncts are universally islands. However, recent research in languages Turkish and Uyghur with rich case morphology has challenged the idea that all adjuncts are islands for extractions (e.g., Öztürk, in press; Ikizoğlu, 2007; Aygen, 2002; Meral, 2010). For example, in Turkish, case-marked adjuncts can be extracted while bare adjuncts cannot. Based on these findings, they propose that island effect is more related to the nature of extracted elements rather than the position from which they are extracted. All in all, these changes have important consequences for L2 research findings and their implications for the L2 studies which have examined the difference between native and nonnative speakers in ultimate attainment in SLA, testing adult L2 speakers with typologically different L1s on the processing of wh-dependencies in real time.

The present study will contribute to previous studies in L2 sentence processing by looking at online processing of wh-dependencies in L2 English by two adult L2 groups who have typologically different L1s - Turkish with a wh-in-situ L1 in which wh-phrases exhibit overt movement via scrambling, and Spanish which has real overt wh-movement. The focus of the study will be on the issue of whether the difference between the native and nonnative speakers in ultimate attainment is due to a processing problem (i.e., a subject and object in grammatical wh-extractions) or a deficit in L2 grammar.

In addition to these issues, WM resources are assumed to be another variable affecting sentence comprehension in L2. However, only a few studies have focused on the role of WM in L2 sentence processing. In the next section, I will discuss these briefly to build some grounds for the present study.

3.5.4. Working memory and L2 sentence processing

Large individual variations in on-line L2 performance have recently attracted L2 researchers to look at the role of WM resources in L2 processing. Only a few L2 researchers have examined whether or not WM resources play a significant role in L2 processing, particularly in processing of complex and ambiguous structures in L2 (e.g., Juffs, 2004; 2005; Roberts, Marinis, Felser, and Clahsen, 2007; Felser and Roberts, 2007; Havik et al. 2009).

The first attempt in this field has been made by Juffs (2004). Juffs initially measured Chinese, Japanese, and Spanish L2 learners' WM capacity using reading span and word span tests in the L1 and L2. The L2 reading test was developed by Harrington and Sawyer (1992) for L2 learners as a version of Daneman and Carpenter's (1980). L1 reading span tests included a Japanese reading span test developed by Osaka and Osaka, (1992), and two in-house tests developed for Chinese and Spanish. Then, he examined the variability in reading times in garden path sentence, particularly in their disambiguating region (i.e., *looked* in 43a), where the initial main clause analysis is turned out to be incorrect. Examples (43a-c) below were taken from Juffs (2004, p. 209).

43. a. ¿After the children cleaned the house looked very neat and tidy.

b. When the student arrived the professor asked her about her trip.

c. The doctor knew the nurses liked the man from England

Juffs first looked at the relationship among the word-span scores, L1 and L2 reading-span scores, the WM scores and the mean RTs of each participant. Results indicated a strong correlation between L1 and L2 reading spans, r=.61, p=0001. However, no relationship was observed between any of the WM scores and RTs each participant spent for the main verb in the garden path sentence. In addition, Juffs divided participants into three span groups (e. i., high-span, mid-span, and low-span) according to their reading span scores and compared them in terms of their reading performance on the main verb in the GP sentence. No significant results were obtained from this analysis, suggesting that both high- and low-span groups behaved similarly when reading the disambiguating region in the GP sentence. Similar grouping analysis was conducted with word-span scores. Juffs reported that the low-span participants spent longer RTs than the high-span participants on the main verb in GP and non-GP sentences, but no interaction was found between structure and span groups, suggesting that the processing load in the GP sentence did not incur more WM recourses than in the non-GP sentences.

Overall, Juffs interprets these results as weak evidence for the role of WM in an on-line processing task. Specifically, WM scores do not correlate with the mean scores at the point where the processing load is greatest in a GP sentence. The author concludes that these findings are consistent with the Waters and Caplan's (1996) view of separate WM resources for on-line syntactic processing. Another possible interpretation of these results is that if word span is related to vocabulary acquisition, learning an L2 consists of learning chunks that can be considered lexical constructs and this can explain the word span results. In other words, high word span learners can accumulate more chunks than low span learners. The more chunks a learner has, the more comparisons s/he can do for internal analysis. The more frequently chunk-internal analyses have been made, the easier it is to analyze new chunks on-line. The result is that high-span learners take less time to resolve ambiguity on-line.

In another self-paced L2 reading study, Juffs (2005) have reexamined the role of WM capacity in processing of syntactically complex sentences. More specifically, he has looked at the relationship between the WM capacity and the processing difficulty observed at the embedded main verb (i.e., *liked*) in subject extractions from long-distance finite clauses such as (*Who did the woman suggest liked the manager at the office*). The embedded verb was taken as the point where processing load was highest, and where individual differences in WM would most likely show an effect. However, the results were similar to those reported in Juffs (2004) in the sense that L1 and L2 reading span scores were strongly correlated with each other and with the word span scores, but none of the WM measures were correlated with the processing load at the verb '*liked*'.

Felser and Roberts (2007) have recently examined the role of WM capacity in processing of filler-gap dependencies by a group of advanced Greek learners of English in a cross-modal picture priming study. More specifically, they investigated whether there is a significant difference between high- and low-span L2 learners in retaining a wh-filler in WM and reactiving it at the identical gap position (i.e., [SQUIRREL] in (44a)) the during the filler-gap integration in indirect object relative clauses as in (44a-b) (Felser & Roberts, 2007, p. 20).

44. Fred chased the squirrel to which the nice monkey explained ….

a. *Identical, gap position:*

…..the game's difficult rules [SQUIRREL] in the class last Wednesday.

b. *Identical, pre-gap position*:

…..the game's [SQUIRREL] difficult rules in the class last Wednesday.

c. Unrelated, gap position:

……the game's difficult rules [TOOTHBRUSH] in the class last Wednesday.

The participants were told to listen carefully to the pre-recorded sentences over headphones and to watch the screen for pictures that would appear at some point during each sentence. Then, whenever a picture appeared on the screen, they had to decide as quickly as possible whether the animal or object in the picture was alive or not, by pressing a button of push-button box. The participants' response times were measured from the point at which the picture appeared on the screen to their pressing the response buttons. They assume that if the antecedent reactivation depends on the availability of sufficient WM resources, then L2 learners' performance should be influenced by individual WM difference. In other words, high-span L2 learners should spend shorter RTs to the picture of SQUIRREL at identical gap position in (44a) (known as antecedent priming effect), than to the picture of TOOTHBRUSH at the unrelated gap position in (44c). However, low-span L2 learners should not show an antecedent priming effect at the identical gap position in (44a) because they do not have enough WM resources to allocate holding an antecedent (or filler) in the WM until it integrates its trace or gaps. On the other hand, if syntactically defined gaps or traces are absent from the representations constructed during L2 wh-processing as suggested by Shallow Structure hypothesis (e.g., Marinis et al., 2005; see Clahsen & Felser, 2006 for more information), antecedent priming effect at the indirect gap should not be observed for any span groups.

Results showed that RTs to identical targets were shorter than RTs to unrelated targets at both the pre-gap and the post-gap position. However, there was no significant difference between the high- and low-span subjects in terms of RTs on identical targets, suggesting that individual WM differences do not affect the learners' RT pattern. To determine whether the L2 learners' performance pattern resembles that of either the high-span or the low-span adult native speakers, or of the high- or low-span children, the authors compared the L2 group with adult native speakers and children in different span groups taken from Roberts et al., (2007).

Results showed that L2 learners' RTs did not pattern either with the high-span or low-span adult native speakers. The L2 learners were different from the high-span native speakers in that only the high-span adult speakers showed antecedent reactivation at the identical gap site (64a). L2 learners' shorter RTs at the pre-gap site and the indirect object gap site suggest that they can identify pictures showing the referent of a wh-filler more easily than unrelated pictures but they cannot show antecedent reactivation effect at the indirect object gap. The L2 learners were also different from the low-span adult native speakers in that the L2 learners but not the low-span native speakers showed significantly shorter RTs to identical targets than unrelated targets. Results from the comparison of the L2 group with children revealed that the L2 learners' RTs did not pattern either with the high-span or the low-span children. High-span children showed antecedent reactivation affect at the identical gap site, indicated by their shorter RTs at this point. However, the L2 learners did not show this pattern. Low-span children spent longer RTs at the identical pre-gap, and the gap site, while the L2 group did the reverse. Felser and Roberts interpret these results as evidence that L2 learners can retain a wh-filler in the WM during processing of wh-dependencies, but they cannot retrieve it from the WM (i.e., reactive them) at structurally defined gap sites. Overall, these results show that L2 reading span did not affect L2 learners' performance in the cross-modal priming task. In other words, L2 learners' failure to posit gaps during on-line processing of wh-dependencies is not due to a shortage of WM resources.

The last research I will review here is the Havik, Roberts, Hout, Schreuder's (2009) study on the role of WM in on-line processing of relative clauses. Havik et al. tested German learners of Dutch to investigate the potential WM span effects on real time resolution of the subject-object ambiguity in Dutch relative clauses. The authors used both short- and long-distance subject and object relative clauses in Dutch as in (45a-d) (Havik et al., 2009, p. 82)

45. *a. subject relative-short*

Daar is de machinist die de conducteurs heft bevrijd uit het brandende treinstel.

that is the engine-driver who the guards has saved from the burning train-carriage

“That is the engine-driver who saved the guards from the burning train-carriage.”

b. Object relatives-short

Daar is de machinist die de conducteurs hebben bevrijd uit het brandende treinstel.

that is the engine-driver who the guards have saved from the burning train-carriage.

“That is the engine-driver who the guards have saved from the burning train-carriage.”

c. Subject relative-long

Daar is de machinist die de conductteurs nah et ongeluk met de trein heeft bevrijd uit het brandende treinstel.

that is the engine-driver who the guards after the accident with the train has saved from the burning train-carriage

That is the engine-diver who saved the guards after the accident with the train from the burning train-carriage.”

d. Object relative-long

Daar is de machinist die de conducteurs nah et ongeluk met de trein hebben bevrijd uit het brandende treinstel

That is the engine-driver who the guards after the accident with the train have saved from the burning train-carriage

"That is the engine-diver who the guards after the accident with the train have saved from the burning train-carriage."

In the subject and object relative clauses above, the NPs *de machinist* and *de conducteurs* are not marked for (nominative/accusative) case so they are ambiguous until a number agreement on the auxiliary *heft/hebben* determines their syntactic function. In the long-distance subject and object relative clauses, a prepositional phrase (i.e., *nah et ongeluk met de trein* 'after the accident with the train' in (45c and 45d) was added into the ambiguous region to increase processing load at this region. The assumption here is that the longer a reader is committed to an erroneous analysis, the more costly the reanalysis process becomes (e.g., Pickering & Traxler, 1998). The participants' WM scores were obtained from a version of Daneman and Carpenter, (1980) Reading Span test in their L1 German and L2 Dutch.

In their analyses, the authors first examined whether or not the processing of the experimental items in general was affected by the participants' WM span, by performing an ANOVA per group on their mean RTs collapsed across the four critical segments (i.e., *heeft/ hebben bevrijd uit het* 'has/have saved from'), sentence length and sentence type, and the participants WM span score as a covariate factor. Results showed an interaction between the WM span scores and sentence length and sentence type. Based on this interaction and high correlation between RTs to the ambiguous region and the WM scores in the L1 and L2, the authors divided both the L2 groups and the native speakers into a high-WM span group (they scored above the median on the memory tests in both the L1 and the L2) and low-span according to their WM span scores.

RTs for the auxiliary in four conditions showed that overall L2 learners were slower than native speakers. Nevertheless, both high-WM groups were slower than their low-WM counterparts. RTs to the verb following the auxiliary showed that the high-span L2 learners spent significantly longer RTs on the verb '*bevrijd*' in object relative clauses than in subject relative clauses in short clauses. There was no significant difference in mean RTs for any conditions for the other three groups. These findings suggest that the high-span L2 learners' longer RTs on the short object relative clauses versus the long object relative clauses may be due to their comparably higher accuracy for the former seen in the verification task. RTs to the preposition reflected that all native speakers regardless of their WM spans showed a processing preference overall for the short subject relatives over the short object relatives, whereas there was no such difference for the L2 learners. In addition, only the high-span native speakers were significantly faster on long distance subject relatives than long distance object relatives. The low-span native speakers were significantly faster on short subject relative clauses than short object relative clauses. None of the comparisons at this point were significant in the L2 groups. In a parallel accuracy analysis, the authors investigated whether the groups' WM span score affected their ability to correctly respond to the verification statements. The L2 group patterned with the low-span native speakers with much lower accuracy for long distance object relatives. This shows that experimental items involving long-distance object and subject relative clauses are rather difficult for both low-span native speakers and all L2 learners.

Manipulating these items, Havik et al., conducted a follow-up study with another group of German learners of Dutch. Results from RTs collapsed across the four segments in the ambiguous region were similar to those of the first experiment in that the L2 learners read all of the segments (*heeft/ hebben bevrijd uit het* ‘has/have saved from’) in the ambiguous region more slowly than the native speakers, and the high-span native speakers read the same region more slowly than the lowspan native speakers. The authors interpret these findings as evidence for the claim that those with more WM capacity are able to keep both interpretations (i.e., subject and object relative clause) of the ambiguous sentences activated as has been argued for in earlier monolingual studies (MacDonald, Just, & Carpenter, 1992; Pearlmutter & MacDonald, 1995). Furthermore, results from RTs for word-by-word reading of the ambiguous region showed that both native speakers and L2 groups spent longer RTs on disambiguating auxiliary in short sentences than in long sentences. The high-span native speakers showed a processing advantage for subject relatives over object relatives in the short condition, but this does not reach a significant level. This suggests that recovery from an erroneous analysis in both long and short sentences are equally difficult for native speakers. For the L2 learners, there was no processing advantage or disadvantage for any sentence type on any of the critical segments. The authors conclude that the results of the follow-up experiment suggest that WM capacity did not affect the L2 learners’ on-line processing.

Overall, the results of this study show that when a task demand is increased, the high-span L2 learners behave similarly to the low-span native speakers in subject-object ambiguity in short-sentences. In other words, they show an on-line preference for subject–resolved sentences in a short-sentence condition. However, when asked to read more for meaning, the L2 learners showed no such on-line RT advantage or disadvantage.

To sum up, most of the studies reviewed above have failed to show any clear evidence for a strong relationship or a significant interaction between WM capacity and L2 learners' on-line performance in processing syntactically complex sentences (Juffs, 2004, 2005), difficulty in long distance filler-gap integration (Marinis et al, 2005), or ambiguity resolution (Havik et al., 2009). Although, some studies have shown that high-span L2 learners behave similarly to low-span native speakers in processing either subject-object or antecedent reactivation in filler-gap dependencies (Havik et al, 2009; Felser & Roberts, 2007), more research is required to confirm these findings.

Alternatively, as Juffs (2004) pointed out, these results can be interpreted as evidence for the Separate Language Interpretation Resource theory (SLIR) of Waters and Caplan (1999), which suggests that verbal WM resources are divided into two separate pools of resources; one for linguistic processing, which involves automatic first-pass analysis of an input sentence (interpretive processes), and the other one is for the use of this extracted-meaning in accomplishing cognitive tasks such as reasoning, planning, and so on (post-interpretive processes). According to the SLIR model, no interaction should be observed between the reader's WM span and the interpretive processes (i.e., automatic first-past syntactic processing), this because of the fact that the WM demands brought onto this system by the sentences used in a linguistic task are within the limit of subject's verbal WM capacity which is specialized for the interpretive processes. For example, the RT measures used in the linguistics tasks are based on the sentences that subjects respond correctly and error rates are relatively low. However, there may be some interaction between the WM span and post-interpretive processes since post-interpretive processes rely on a general pool of WM resources (Caplan et al., 2007)

Another point is that there are still methodological concerns associated with WM resources in L1 sentence processing such as types and number of WM tasks and methods of applications (see Conway et al., 2005 and Roberts & Gibson, 2003 for a review). Similarly, these concerns are also relevant for L2 WM research. These factors might be responsible for the findings showing no relationship between the WM capacity and performance in linguistic task. To uncover this relationship, as Juffs (2005) noted, WM resources in L2 processing should be carefully investigated with more sophisticated statistical analysis.

To conclude, current studies are not able to show any relationship between WM resources and L2 sentence processing. Nevertheless, it is important to note that WM in L2 sentence processing is a relatively new construct and we need further research to gain a better understanding of its role within L2 sentence processing.

Within this context, this study aims to contribute to the field, examining the influence of WM capacity in processing wh-dependencies in L2 English by native Turkish and Spanish speakers. This will provide new empirical data to identify whether WM resources are related to sentence processing performance in the L2.

CHAPTER 4

METHODOLOGY

4.1 Introduction

This chapter reports on an experiment conducted to examine adult L2 learners' access to the Subjacency Principle and related island constraints in English. The experiment tested highly proficient Turkish- and Spanish-speaking learners of L2 English in an online grammaticality judgment task presented in the full sentence and the self-paced word-by-word reading conditions. In addition, to measure working memory (WM) capacity, the participants were given two types of online WM tasks in the L2 English; 1) automated reading span (ARSPAN) task; and 2) automated operation span (AOSPAN) task. Furthermore, a Turkish ARSPAN for Turkish participants; and a Spanish ARSPAN for Spanish participants were administered to find out whether or not WM span is language dependent. Before I discuss the details of the experiment, I would like to clarify the research questions in this study.

As mentioned earlier, this study mainly aims to investigate L2 processing of *wh*-movement (extractions) in English in order to examine the online processing of wh-dependencies in the L2 English by ultimate state adult L2 learners with a wh-in-situ L1 which has covert wh-movement at LF and overt movement via scrambling. More specifically, the study investigates whether or not adult Turkish- and Spanish-speaking learners of English can accept grammatical wh-extraction from both finite and nonfinite clauses and reject ungrammatical sentences which violate island constraints on wh-movement in the same way as native speakers of English. To test whether or not participants correctly accept grammatical *wh*-sentences in online processing, five types of grammatical sentences with long-distance *wh*-movement were used: 1) Subject extractions from finite clauses; 2) Object extraction from finite clauses; 3) Object extraction from finite clauses (with *that*); 4) Subject extraction from non-finite clauses; 5) object extraction from non-finite clauses. Also, to test whether or not participants reject ungrammatical sentences in real-time processing of *wh*-sentences, five types of long-distance wh-movement violations were used; (1) Complex Noun Phrase Island; (2) Relative Clause Island; (3) Adjunct Island; (4) Subject Island; and 5) *That*-trace sentences (see Appendix A for the list of grammatical and ungrammatical experimental items).

The assumption here is that if adult L2 learners' L1 have a positive influence on their L2, learners will be as successful as native English speakers in both accepting grammatical sentences and correctly rejecting ungrammatical sentences. That is, L2 learners should demonstrate native-like processing patterns in judging grammatical and ungrammatical sentences even if they might show slower reaction time (RT) in judging these sentences.

To examine the role of L1 in L2 sentence processing, speakers of Turkish, a language which allows overt wh-movement via scrambling are contrasted with speakers of Spanish, a language similar to the L2 English with respect to overt wh-movement. The prediction here is that coming from a L1 with overt wh-movement as in the L2 English; the L1 Spanish group will have more advantage than the L1 Turkish group in processing wh-movement (extractions) in the L2 English. Nevertheless, due to availability of overt-movement via scrambling with island constraints, L1 Turkish speakers, who are in the ultimate L2 state, are expected to behave like native English speakers and L1 Spanish speakers in their overall parsing performance and parsing strategies.

A further aim of this study is to explore the relationship (if any) between the native English speakers' and L2 learners' WM measures and their performance in processing wh-extraction from both finite and non-finite clauses. The assumption here is that if WM resources have an influence on the performance of native speakers and L2 learners in processing wh-extraction, there should be a significant correlation between the WM measures and the accuracy scores and RTs of the participants. More specifically, the accuracy scores and RTs for the sentence types which are difficult to process should strongly correlate with the WM capacity. Also, high-WM-span participants should be better than low-WM-span participants in accuracy and speed in processing wh-extraction, particularly in subject extractions from finite and nonfinite clauses, which are assumed to be more complex than other structures.

Another important aim of this study is to examine whether or not the subject/object asymmetry found in earlier studies can also be observed in native and nonnative English speakers' processing data.

The specific research questions in this study are formulated as follows:

Research questions:

1. Is there a subject/object asymmetry in processing *wh*-extraction from finite and non-finite grammatical sentences? If so, which wh-extraction type causes more difficulty for native and non-native groups?

2. Do L1 Turkish and L1 Spanish learners of L2 English pattern similarly to native English speakers in their acceptance of grammatical *wh*-extractions and rejection of ungrammatical *wh*-extractions with island constraint violation in English?

3. Does the L1 of the learner play a role in real-time processing of wh-extraction in L2 English? In other words, is a wh-in-situ L1, which allows overt movement in scrambling, and which exhibits certain island constraints, as effective as Spanish— a language similar to English in the way it exhibits overt movement and island constraints, in processing L2 English wh-dependencies?

4. Does the WM play a role in processing wh-dependencies in English?

    a) Is WM capacity language independent?

    b) Do differences in the WM capacity affect L2 learners' speed and accuracy in online processing of L2 wh-dependencies? More specifically, is there a relationship between L2 learners' WM capacity in the L1 and L2; and is there a relationship between L2 learners' WM capacity in the L2 and their L2 syntactic parsing performance in wh-extraction from both finite and non-finite grammatical sentences?

As for the predictions, I will first begin with the predictions about the subject-object asymmetry and explain, from a theoretical point of view, the potential processing problems predicted in subject and object extractions.

On this basis of Pritchett's Generalized Theta Attachment Theory (1992), I predict that subject extraction from both finite and non-finite clauses will be more difficult to process than object extractions from both finite and non-finite clauses. Accordingly, longer RTs on subject extractions than object extractions are expected. This is because subject extractions from both finite and non-finite clauses require more changes in theta roles and Case features as suggested by Pritchett (1992). For example, subject extractions from a finite clause (e.g. *Who did the nurse know _ saw the patient at the hospital*?) is assumed to be more difficult to process than object extractions from a finite clause (*Who did the nurse know the doctor saw_ at the hospital*?) under the *Active Filler Strategy* (Frazier, 1987; Frazier & Clifton, 1989; Frazier & d'Arcais, 1989), which assumes that in the first sentence, the parser initially identifies the wh-phrase (filler) *who*, and postulates a gap, which is the location of the trace of the moved wh-phrase. Since the subject position of the matrix clause is unavailable, it postulates a gap as the object of the matrix verb so that the wh-phrase can get a theta role and Case through A-bar chain from the trace that is governed by the matrix verb *know*. In this way, all principles of the grammar are satisfied locally as soon as possible. However, when the parser encounters the verb *saw*, the parse pauses and this is reflected as longer RTs at this region. The parser has to make a reanalysis because *saw* is a finite verb having two theta roles to discharge. The gap that is initially postulated as the object of the verb, *know* has to be changed as the subject gap of the embedded clause. Furthermore, previously assigned theta roles and Case have to change since the subcategorizing verb has changed.

In object extraction, on the other hand, there are fewer changes in theta roles and Case. In a sentence like *"Who did the nurse know the doctor saw_ at the hospital?",* as soon as the parser encounters the wh-phrase, it posits a gap to the first potential position of the moved wh-filler '*who*'. Since the subject position is not available, it postulates a gap as the object of the main verb '*know*'. However, when it encounters the subject NP '*the doctor'*, the parser cancels this initial analysis and reanalyzes the gap as the object of the embedded verb '*saw*'. The type of theta role and type of Case remain the same. It has therefore been claimed that subject extraction from a finite clause is more difficult for all speakers to process than object extraction from a finite clause. If learners have problems with extracted subjects, it should stem from a need to reanalyze of matrix object trace to embedded subject trace and the concomitant changes in theta and Case.

Also, subject extraction from a non-finite clause such as "*Who does the manager expect _ to meet the job applicants today*" is assumed to be the most difficult extraction to process, while object extraction from a non-finite clause such as *"Who does the manager expect to meet at work this morning?*" is the easiest of all. This is because the parser makes three reanalyses during subject extractions from non-finite clauses, which requires more reading times, whereas it only makes one analysis in object extractions from a non-finite clause.

Following Pritchett's Generalized Theta Attachment (GTA) theory, the prediction is that processing of a subject wh-phrase extracted from a nonfinite clause (e.g. *Who does the manager expect _ to meet the job applicants today*), is more complex than a subject extraction from a finite clause (e.g. *Who did the nurse know _ saw the patient at the hospital*?). In both types of clauses, the parser follows the same steps until the matrix verb '*expect*' as in (1).

1. "*Who$_i$ does the manager expect t$_i$* ?"

However, upon encountering '*to*', the parser makes the first reanalysis, which involves positing of a subject trace as in (2) because '*who*' requires Case and may get it from a subject trace that receives accusative Case from the Exceptional Case Marking (ECM) verb '*expect'* in the subject position of the embedded clause.

2. *Who*$_i$ *does the manager expect [*$_{IP}$ *t*$_i$ *to*]

However, as soon as '*meet'* is encountered, a second reanalysis takes place, because (*Who*$_i$ *does the manager expect to meet t*$_i$*?*) is a well-formed sentence. In this case, the parser will postulate a PRO in Spec IP of the embedded clause and automatically project a CP to block case assignment by the ECM verb. The parser posits an object trace in the embedded clause, all theta roles are assigned, and all chains are properly licensed as in (3)

3. *Who*$_i$ *does the manager*$_k$ *expect* [$_{CP}$ [$_{IP}$ PRO$_k$ *to meet t*$_i$*?*].

The third reanalysis takes place when the parser encounters the embedded object NP '*the job applicants*'. The parser has to backtrack and reanalyze the PRO as a subject trace of the embedded clause once again as in "*Who*$_i$ *does the manager expect* [$_{IP}$ t$_i$ *to meet the job applicants?*]".

In processing of object extraction from a nonfinite clause such as *"Who does the manager expect to meet at work this morning?"*, the initial steps (step1 and 2) are similar to those of subject extractions, but when parser encounters the embedded verb '*meet*', it will postulate a PRO in Spec IP of the embedded clause and automatically project a CP to block case assignment by the ECM verb. The parser will postulate a trace as the object of the embedded clause, all theta roles will be assigned, and all chains will be properly licensed as in (4) and this will be the correct analysis of the wh-filler '*who*'.

4. *Who*$_i$ *does the manager*$_k$ *expect* [$_{CP}$ [$_{IP}$ PRO$_k$ *to meet t*$_i$*?*

To sum up, three reanalyses take place in parsing a subject extraction from an embedded nonfinite clause: (1) from the matrix object trace to the embedded subject trace; (2) from the subject trace to PRO+object trace; (3) from PRO back to the embedded subject trace. Therefore, processing subject extraction from nonfinite clauses is expected to be the costliest of all wh-extraction types for both native and non-native groups. This cost might be reflected as longer RTs in all groups. Nevertheless, the L2 groups might be relatively slower than native speakers in executing this complex parsing in subject extraction from nonfinite clauses. As for accuracy, it is difficult to predict whether longer RTs may be translated into decreased accuracy in this sentence type across both the full-sentence and word-by-word reading conditions. In other words, participants might take longer to process a particular sentence but they may eventually arrive at a correct judgment. This is what we expect in all groups. If, on the other hand, one takes longer to read/process a sentence and still fails to make accurate judgment on them, this may imply a deeper grammatical representational problem.

With respect to the ungrammatical wh-extractions, I predict that both L1 Spanish and L1 Turkish groups will exhibit sensitivity to island constraints similar to native speakers because in each L1, there are certain movement constraints that are expected to make L2 learners much more conscientious to island violations. Thus, if L2 learners are found to be as sensitive as native speakers to island constraints, no difference should be found between native and non-native speakers in their accuracy on strong island (i.e., the relative clauses, subject, and adjunct islands) and weak island violations. (i.e., Complex NP island). However, all participants can be more accurate on rejecting strong islands than weak island (Complex NP island violation). Since extraction from weak islands is ungrammatical to a lesser extent than extraction from strong islands, participants may tend to incorrectly accept weak island violations. If participants are found to be less sensitive to violations in weak island than strong island, then their accuracy on complex NP/DP island should be lower than those on the strong islands such as relative clause island, adjunct island, and subject island

With respect to RTs, if L2 learners are sensitive to island constraints during the online first-pass reading, they will not expect to integrate a wh-filler with an illicit gap inside both strong and weak islands, which will be reflected short RTs for gap sites in islands.

In sum, RTs in ungrammatical constructions might be processed more slowly by the L2 groups than the control group. Nevertheless, I do not predict any differences between the two L2 groups with respect to processing speed in sentences with island violations in either the full-sentence or word-by-word reading conditions.

With respect of wh-extractions with *that*-trace violation (e.g. *Who do the police believe that attacked the man last night?*), we assume that both the Spanish and Turkish learners will be as accurate as native speakers because in the end-state L2 acquisition, we expect native-like judgments for *that*-trace violations in the L2. Nevertheless, the L2 groups might be less accurate than the native speakers in case they exhibit negative transfer from their L1s. Particularly, the Spanish group might have more problems with these constructions than the Turkish group because in the L1 Spanish, subject extraction from an embedded clause in the presence of an overt complementizer *that* is a grammatically well-formed structure. The pro-drop nature of L1 Spanish might also mislead the L1 Spanish group into incorrectly accepting ungrammatical wh-extractions with *that*-trace violation in the L2 English. The Turkish learners, on the other hand, may also fail to reject ungrammatical wh-extractions with *that*-trace violation due to the pro-drop feature of L1 Turkish. Nevertheless, since Turkish does not have a corresponding complementizer (except for the infrequently used *–ki*, which is, nevertheless irrelevant in this context), the L1 Turkish group is expected to do better than the L1 Spanish group in processing *that*-trace violation in the L2 English.

With respect to the role of WM in processing wh-dependencies in the L2 English, I, first of all, predict that L1 and L2 WM measures will correlate with each other. This will imply that we can talk about WM capacity independent of the language(s) involved (Osaka & Osaka, 1992). Furthermore, I predict that WM capacity will correlate with the sentence processing efficiency (i.e., the rate of accuracy and speed in processing) in the L2. More specifically, high-span participants will be more accurate and faster than low-span participants on grammatical and ungrammatical wh-extractions. High-span participants will experience less processing difficulty with the extraction of subject and object from finite and nonfinite clauses.

4.2 Participants

Two groups of adult L2 learners participated in this study: (1) 30 near-native Turkish-speakers of English; and (2) 30 near-native Spanish speakers of English. Similar to English, Spanish has overt wh-movement with an SVO word order. However, Turkish has overt wh-movement via scrambling, and has an SOV word order as its canonical word order. Therefore, including these syntactically different languages as the L1 of L2 learners is believed to contribute to the identification of L1 influence in processing grammatical wh-extractions as well as subjacency violations in English. In addition, 30 adult native speakers of English were tested as the control group in the study.

Table 1. L2 speakers' background information

Groups	Sex		Age			Length of stay in the USA or UK
	Male	Female	Mean age of first exposure to L2 English	Mean age at time of testing	Age range	
Turkish	21	9	11	36	30-54 (SD:5)	8.6
Spanish	12	13	11	33	20-69 (SD:9)	5.9

As can be seen on Table 1, Turkish and Spanish speakers were similar in terms of age, age of first exposure to L2 English, length of stay in an English-speaking country. All Turkish participants received a Ph.D. degree at a university in the USA or UK. 88% of the Spanish participants either obtained a Ph.D. degree or pursing to get it at the University of Essex in the UK or working at that university as faculty members.

The mean age in the English native speakers was 37 with a range of 19-58 (SD: 11). All of them were exposed to English as home language and took their primary, secondary, high school and university education in English. They were all either graduate students or faculty members at the University of Essex in the UK or graduates of various universities in the UK or the USA.

All participants had normal hearing, and normal vision, and were not informed of the ultimate purpose of the experiment.

4.3. Materials

Materials for the OGJT consisted of a total of 100 wh-questions (50 grammatical and 50 ungrammatical sentences). Five types of grammatical biclausal wh-extractions (1a-e); and five types of ungrammatical biclausal *wh*-extractions involving subjacency violations (1f-j). Judgments to sentences with grammatical wh-extraction will help us examine whether or not L2 learners correctly accept long distance wh-movement including subject and object extractions from finite and nonfinite clauses in the L2 English. Additionally, they help us identify whether there is an object preference over subject in processing of long-distance wh-extraction from both finite and non-finite clauses. The purpose of ungrammatical sentences, on the other hand, is to test whether or not the L2 learners correctly reject ungrammatical sentences, which violate island constraints, which will provide evidence for L2 learners' unconscious knowledge of restrictions on wh-movement in the L2 English.

(1)

Grammatical wh-sentences:

1. a. What does the woman think the plumber stole from the garage? (Object extraction/ finite clause)

   b. What does the inspector think that the boy stole from home? (Object extraction/ finite clause with *that*)

   c. Who does the manager expect to meet at work this morning? (Object extraction /nonfinite clause)

   d. Who does the woman think stole the bicycle in the garage? (Subject extraction/ finite clause)

   e. Who does the manager expect to meet the job applicants today? (Subject extraction/ nonfinite clause)

Ungrammatical wh-sentences:

f. *Who did Alison go to work after she took to school? (Adjunct Island)

g. *What does James believe the fact that Alison saw at work? (Complex Noun Phrase Island)

h. *What does Jane visit the architect who designed for her friend? (Relative Clause Island)

i. *Who does the teacher believe a story by amuses the children? (Subject Island)

k. *Who do the police believe that attacked the man last night? (*That*-trace)

There were 10 experimental sentences in each type of grammatical and ungrammatical wh-extraction. In addition, 80 filler sentences, which involved both interrogative forms of different grammatical structures such as relative clauses, were included as test items in order to prevent the participants from discovering experimental sentences and developing any response strategies.

4.4 Instrument

In this study, an online grammaticality judgment task was presented in two conditions; (1) the full- sentence condition to get an idea about RTs spent on each sentence (White & Juffs, 1998; Juffs & Harrington, 1995); and (2) the self-paced word-by-word reading with a moving window display (Just et al., 1992) to identify specific loci of processing difficulty. Additionally, two types of online WM tasks in the L2 English were administered to all participants. Turkish and Spanish participants were also given an Automated Reading Span task in their respective L1s.

(1*) Full- sentence condition*: in the first part of the experiment, participants read and judged a set of the sentences in the full-sentence presentation condition, in which the entire sentence appeared on the screen of a computer. The sentences were displayed one at a time. The participants were asked to read a sentence and press a (green) YES key if they find the sentence to be a grammatically correct in English or to press a (red) NO key on the keyboard if they find it to be grammatically incorrect. The letter q and p on the keyboard were covered in green and red, respectively. They were asked to do this as quickly as possible. The amount of time that participants spent reading each sentence and making grammaticality judgment was recorded as the time between the key-presses. After an incorrect response, the word ‘INCORRECT’ flashed briefly on the screen as a feedback to the participant. A similar feedback was also given for a correct response. The presentation of items is randomized for each participant. There was a practice session involving 10 samples of grammatical and ungrammatical wh-extractions before the real trial to familiarize the participants with the experiment.

(2) *The self-paced moving window reading technique* provided the collection of word-level readings to identify specific loci of processing difficulties. The participants first read and judged the sentences in the self-paced moving window condition (Just et al., 1982). In this technique, each sentence was presented on a computer screen one word at a time. The words appeared in a linear position in the sentence moving across the screen from left to right. Participants pressed the spacebar to reveal each word of the sentence. As a new word appeared, the preceding word disappeared from the screen. The amount of time the participant spent reading each word was recorded as the time between the key-presses. After the last word the sentence, a question asking whether the sentence was grammatically ‘correct’ or ‘incorrect in English appeared on the screen. Participants pressed one of two keys to respond a YES or NO to the question. The software collected

word-by word RTs and accuracy score for each sentence in the experiment. The presentation of items was randomized for each participant.

(3) Working Memory tasks

It has been claimed that although there are a number of measures of WM capacity such as reading span, counting span and operation span but it has been suggested that no single task alone is a perfect measure of the construct it ostensibly represents (Conway et al., 2005). For example, the Operation Span task measures WM capacity but, most likely also taps mathematical ability, motivation, and word knowledge among other things. (e.g., Turner & Engle, 1986). Similarly, the Reading Span Task measures WM capacity but it certainly also tests verbal ability. Therefore, despite being reliable and valid measures of WM capacity, WM span tasks are not perfect or process-pure. Therefore, it has been suggested that an optimal research strategy is to administer multiple WM span tasks and derive a latent variable from the average of scores obtained on all the tasks as one single overall measure of WM capacity.

In this study, I used two types of WM task in the L2 English; (1) Automated Reading Span (ARSPAN) (see Appendix B: Unsworth et al., 2005); and (2) Automated Operation Span (AOSPAN) (Unsworth et al., 2005). Then, I derived a latent variable from the common variance among reading span and operation span scores.

In addition, a Turkish ARSPAN task for Turkish participants and a Spanish ARSPAN for Spanish speakers were developed in line with the English version and administered to examine whether or not WM span is language-independent. Both the

ARSPAN and AOSPAN tasks were mouse-driven and required participants to click the mouse button.

First, I will describe AOSPAN task. The AOSPAN had three practice sessions. The first practice session involved a simple letter span, in which a letter appeared on the screen, and the participants were required to recall the letters in the same order, in which they were presented. In all experimental conditions, letters remained on screen for 800 msec. At recall, the participants saw a 4x3 matrix of letters (F, H, J, K, L, N, Q, R, S, T, and Y). Recall consisted of clicking the box next to the appropriate letters in the correct order. The recall phase was untimed. After the recall session, the computer provided feedback about the number of letters correctly recalled in the current set.

Next, the participants practiced the math portion of the task. They first saw a math operation (e.g.., (1*2) +1=?). The participants were instructed to solve the operation as quickly as possible and then click the mouse to move to the next screen. On the next screen a digit (e.g., 3) was presented and the participants were required to click either a 'true' or 'false' box, depending on their answer. After each operation, the participants were given accuracy feedback for their accuracy percentage. The math practice familiarized them with the math portion of the task and calculated how long it would take each person to solve the math operations. After the math practice, the program calculated each individual's mean time required to solve the equation. This time was then used as a time limit for the math portion of the experimental session for that individual. This practice session involved 15 math operations.

In the final practice session, the participants performed both the letter recall and math portions together, just as they would do in the real block of trails (see Figure 1) (Unsworth et al., 2005, p.500).

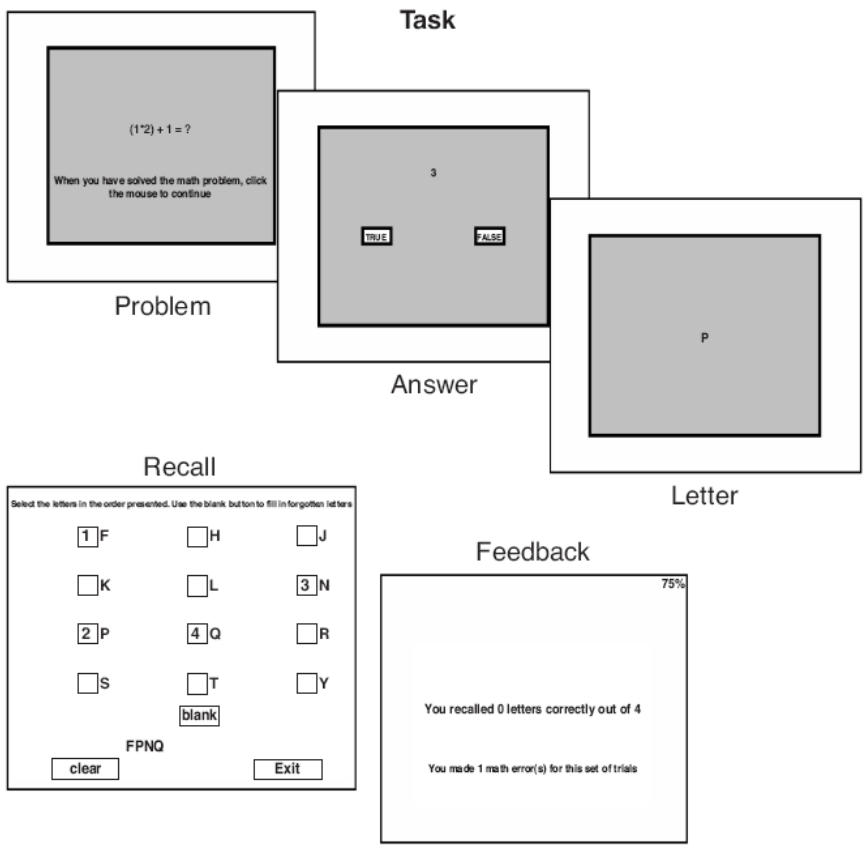


Figure 1. Illustration of the automated operation span task: In the task, first a math operation is presented. After it is solved, participants click the mouse and a digit is presented, which is to be judged as either correct or incorrect as answer to that math operation. This is followed by a letter for 800 msec. These steps are repeated for a certain number of times until the recall session begins. For recall, participants are asked to remember and click on the letters they saw in the exact order of their individual appearance on the screen. After recall, feedback is presented for 2.000 msec.

The participants first saw the math operation, and after they clicked the mouse button indicating that they had solved it, they saw the letter to be recalled. If the participant took more time to solve the math operation than their average time plus 2.5 SD, the program automatically moved on and counted that trail as a speed error. Participants completed three practice trails, each of set size 2. After participants completed all of the practice sessions, the program moved on the real trials, which consisted of three sets of each set size, with the set sizes ranging from 3-7. This made for a total of 75 letters and 75 math problems. The order of set sizes was random for each participant.

At the end of the real trials, the program reported five scores to the experimenter: Ospan score, total number correct, math errors, speed errors, and accuracy errors. The first Ospan score was the sum of all perfectly recalled sets. For example, if an individual correctly recall 3 letters in a set size of 3, 4 letters in set size of 4 , and 3 letters of in a set size of 5, his or her Ospan score would be 7 (3+4+0). The ‘total number of correct’ was the total number of letters recalled in the correct position. Three types of errors were reported: ‘Math errors’ were the total number of task errors, which was broken down into ‘speed errors,’ in which the participant run out of time in attempting to solve a given math operation, and ‘accuracy errors’ in which the participant solved the math operation incorrectly. The task took approximately 20-25 min to complete.

The ARSPAN task had the same procedure as the AOSPAN task. There were three practice sessions: letter practice, sentence practice and letter and sentence practice before real trials. The letter practice session was the same as the one in Ospan above. In the sentence practice session, the participants were required to read a sentence e.g., as "*The prosecutor's dish was lost because it was not based on fact*" as quickly as possible and then click the mouse to move the next screen. On the next screen, they determined whether the sentence makes sense or not by clicking a 'true' or 'false' box, depending on their answer. Nonsensical sentences were made by simply changing one word (e.g., "dish" from "case") from an otherwise normal sentence. Participants read 15 sentences in this session. This session served to familiarize participants with the sentence problems of the task. Also, during this practice session, each participant's mean RT to read a sentence is automatically computed by the software as in the second practice of Ospan task.

In the third practice session, participants first read a sentence, and decide whether made sense or not, Then, they saw a letter to be recalled as in the third practice of Ospan. Participants completed three practice trials each of set size 2. After participants completed all of the practice sessions, the program moved on to the real trials, which consisted of three sets of each set size, with the set sizes ranging from 3-7. There were a total of 75 letters and 75 sentence problems. The order of set sizes was random for each participant. At the end of the real trials, the program reported five scores to the experimenter: Rspan score, total number correct, sentence errors: speed errors, and accuracy errors).

The Turkish and Spanish versions of English ARSPAN were administered in the same way as the English ARSPAN task was administered (see Appendix C and D).

4.5 Procedure

All experimental tasks were individually administered on a laptop using E-prime 2.0 (Schneider, Eschman, & Zuccolotto, 2002) in two sessions, each of which took 1.5 hours. In the first session, the native speakers of English took two online tasks; (1) ARSPAN task in English, and (2) the GJT in the full-sentence condition. In their first session, the L2 learners took three tasks; (1) ARSPAN task in English, (2) the AOSPAN task in English, and (3) the GJT in the full-sentence condition. One week later, in the second session, the native speakers took the AOSPAN task and the GJT in the self-paced word-by-word reading fashion, while the L2 learners took ARSPAN in their L1 and the GJT in the self-paced word-by-word reading fashion.

CHAPTER 5

RESULTS

5.1 Introduction

In this section, I will first discuss the results from the online grammaticality judgment task (OGJT) which was presented in two conditions: (1) the full-sentence condition; and (2) the self-paced word-by-word reading condition. Recall that in both conditions, participants were asked to read and give judgments on a total of 100 grammatical and ungrammatical sentences, each of which included five types of grammatical and five types of ungrammatical wh-sentences as listed below:

Grammatical wh-sentences:

1. What does the woman think the plumber stole from the garage? (Object extraction/ finite clause)
2. What does the inspector think that the boy stole from home? (Object extraction/ finite clause with *that*)
3. Who does the manager expect to meet at work this morning? (Object extraction/nonfinite clause)
4. Who does the woman think stole the bicycle in the garage? (Subject extraction/ finite clause)
5. Who does the manager expect to meet the job applicants today? (Subject extraction/ nonfinite clause)

Ungrammatical wh-sentences:

1. *Who did Alison go to work after she took to school? (Adjunct Island)

2. *What does James believe the fact that Alison saw at work? ? (Complex noun phrase Island)

3. *What does Jane visit the architect who designed for her friend? (Relative Clause Island)

4. *Who does the teacher believe a story by amuses the children? (Subject Island)

5. *Who do the police believe that attacked the man last night? (*That*-trace)

In the full-sentence condition, the dependent variables are the accuracy in grammaticality judgments and reading times (RTs) (i.e., the total amount of time participants take to read the whole sentence). As mentioned earlier, a 'Yes' response for a grammatical sentence and a 'No' response for an ungrammatical sentence is considered to be a correct response.

In the moving-sentence condition, as in the full-sentence condition, the dependent variables are participants' accuracy and word-by-word reading times for grammatical and ungrammatical wh-sentences.

In the last section of this chapter, I will discuss the results of the working memory (WM) tasks. Recall that two online WM tasks were given in L2 English; (1) Automated Reading Span (ARSPAN); and (2) Automated Operation Span (AOSPAN). In addition, a Turkish ARSPAN task for Turkish participants and a Spanish ARSPAN task for Spanish speakers were administered to find out whether or not WM span plays a significant role in L2 learners' online processing of long-distance grammatical and ungrammatical wh-extractions, independent of language. Recall also that, in the ARSPAN task, participants were required to read sentences while trying to remember a set of unrelated letters (B, F, H, J, M, Q, R, and X). For this task, participants read a sentence and determined whether it made sense or not (e.g., "The prosecutors' dish was lost because it was not based on fact"). Then, they clicked on either a "True" or a "False" button on the keyboard, depending on their judgment. On the following screen they saw a letter to be remembered. There were three sets of each set size, with the set size ranging from 3 to 7, which made for a total of 75 letters and 75 sentences. At the end of the task, the software reported the following scores to the experimenter: (1) ARSPAN absolute score, which was the sum of all perfectly recalled sets; (2) ARSPAN total number correct, which was the total number of letters recalled in the correct order; (3) two types of sentence errors: a) speed errors in which the participant run out of time while judging a given sentence; b) accuracy error in which the participant judge the sentence incorrectly.

In the AOSPAN task, participants were required to solve simple math operations while trying to remember a set of unrelated letters as in the ARSPAN task. They first saw a math operation (e.g., ((1*2) +1=?). The participants were required to solve the problem as quickly as possible and click the mouse to move to the next screen. On the next screen, a digit (e.g., 3) was presented and the participants were required to click either a "True" or "False" button, depending on their answer. There were three sets of each set size, with the set size ranging from 3 to 7, which made for a total of 75 letters and 75 math operations. The same scoring procedure was used as in the ARSPAN task.

The Spanish and Turkish ARSPAN tasks were the translated versions of the English ARSPAN task **(**Unsworth, Heitz, Schrock, & Engle, (2005**).** The results of these tasks will be discussed in the final section of this chapter.

In what follows, I will discuss the results of grammatical wh-extractions.

5.2. Results from the full-sentence condition

In this section, I will first present an accuracy and RT analysis over correct responses in the grammatical wh-extractions. I will then discuss accuracy and RTs elicited in ungrammatical wh-extractions.

5.2.1. Accuracy judgments on grammatical wh-extraction types

Before conducting any statistical analyses on the accuracy responses to the five grammatical *wh*-extraction types in the full-sentence condition, the accuracy data were first screened for normal distribution, outliers, and missing values. It was found that the normality assumption was not sustained in the distribution of accuracy scores in grammatical wh-grammatical types. This was assumed to be due to the limited number of sentences (n=10) in each type. However, since there were roughly equal numbers of participants (Spanish: 25, Turkish: 31, and English: 31) in each language group and the variances of 5 types were roughly equal, parametric tests were used in the analysis of accuracy data. Only two participants (one Turkish and one Spanish), who had an overall accuracy below 60% in the grammaticality judgment task, were excluded from the analysis. Mean accuracy scores for the five grammatical wh-extraction types are reported as decimals in Table 1. Participants received a score of 1 for each correct response and a score of 0 for each incorrect response: the numbers in Table 1 are means out of 10 wh-sentences, calculated for each participant's scores in each group.

Table 2. Mean accuracy scores for five grammatical wh-extraction types

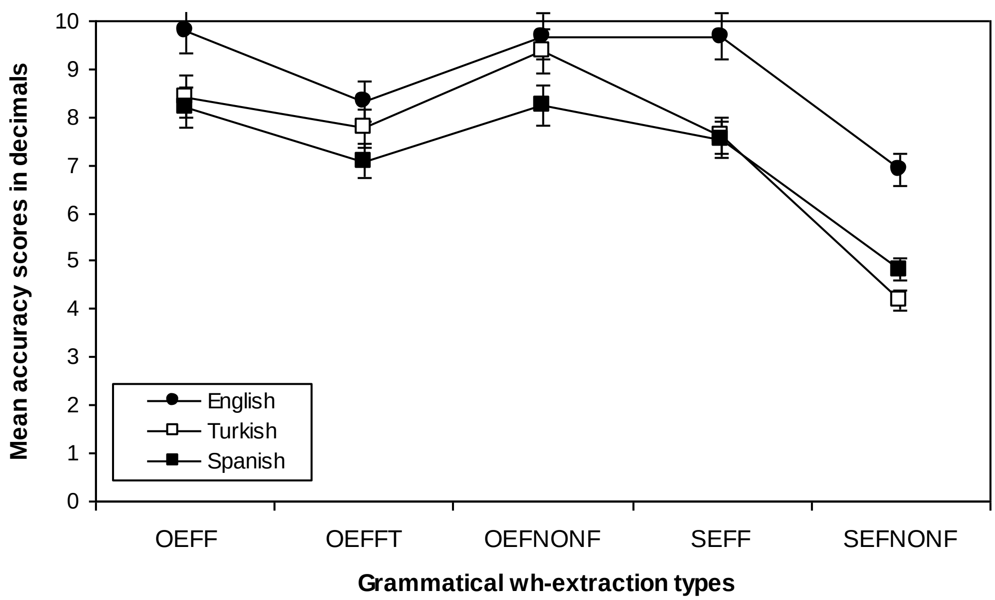
Structure	Spanish (n=24)		Turkish (n=31)		English (n=31)	
	M	SD	M	SD	M	SD
OEFF (n=10)	8.21	1.44	8.43	1.50	9.81	.402
OEFFT (n=10)	7.08	1.92	7.77	1.57	8.32	1.72
OEFNONF (n=10)	8.25	1.65	9.37	.890	9.68	.541
SEFF (n=10)	7.54	1.77	7.60	2.42	9.68	.702
SEFNONF (n=10)	4.83	1.99	4.20	2.58	6.90	1.89
Total (n=50)	7.18	1.75	7.47	1.62	8.88	1.05

OEFF (Object Extraction From Finite clause), OEFFT (Object Extraction From Finite clause with *that*), OEFNONF (Object Extraction From Nonfinite clause), SEFF (Subject Extraction From Finite clause), SEFNONF (Subject Extraction From Nonfinite Clause)

As can be seen from the mean RTs in Table 2, overall accuracy for five types of grammatical wh-extractions in each group is over 70%, which is a level above chance. English native speakers seem more accurate (M=8.88) than the Turkish learners (M=7.47) and the Spanish learners (M=7.18). However, when accuracy scores are examined on a type-by-type basis, all groups reflect a similar accuracy profile. The English native speakers are more accurate on object extraction from finite (OEFF) and nonfinite clauses (OEFNONF), compared to subject extraction from nonfinite clauses (SEFNONF). They are also quite accurate on subject extraction from finite clauses (SEFF), but not as accurate as they were on the OEFF clauses. The large difference between SEFNONF and OEFNONF clauses indicate an asymmetry between subject and object extraction from nonfinite clauses, while the small difference between SEFF and OEFF clauses provides relatively weak evidence for such asymmetry in finite clauses. The two L2 groups, on the other hand, behaved similarly to the native speakers in that they were least accurate on SEFNONF clauses and quite accurate on OEFNONF clauses. Like the native speakers, the L2 learners show a subject-object asymmetry in wh-extractions from nonfinite clauses and finite clauses.

Figure 2 shows accuracy judgment performance of all groups on five types of grammatical wh-extractions from finite and nonfinite clauses.

Figure 2. Mean accuracy scores for five grammatical wh-extraction types.



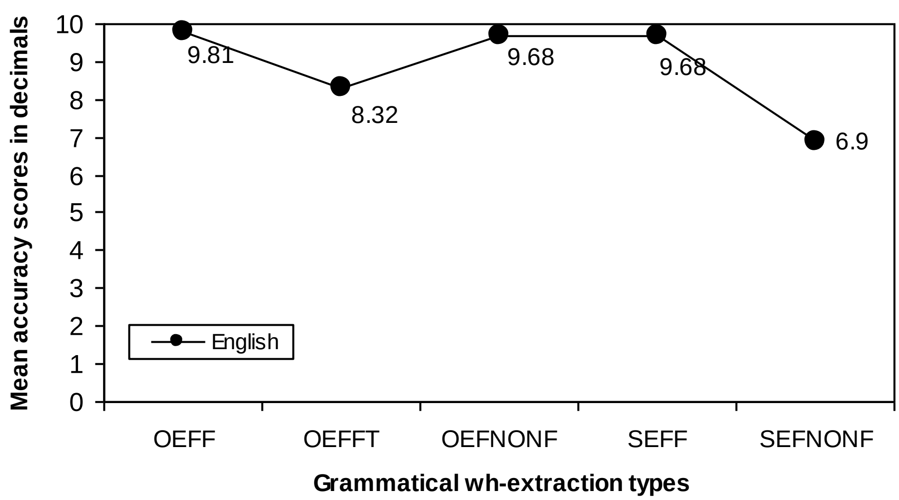
To examine whether the difference observed in the mean accuracy scores of five types in three groups are statistically significant, we conducted a 3x2 repeated measures ANOVA, with language (Spanish, English and Turkish) as the between-subjects factor and type (OEFF, OEFFT, OEFNONF, SEFF, SEFNONF) as the repeated within-subjects factor. The results revealed an overall significant effect for language (F (2, 82) =23.43; p<.01; *MSe*=118.507); a main effect for sentence type (F (4, 328) = 90.18; p<.01 *MSe* = 191.439); and significant interaction for language and type (F (8, 328) = 3.81; p<.01;

MSe =8.08), which suggests that there are significant differences among the language groups in terms of overall accuracy for the five grammatical types. The post-hoc analysis (Tukey HSD, p<.01) of language groups revealed that English native speakers were significantly more accurate on five types of grammatical wh-extractions than the two L2 speakers, but the L2 groups were not significantly different from each other in their judgments. However, the pooling of data hides the differences between the groups across the five structure types (see Table 1). A one-way ANOVA conducted per type indicated that the native speakers and the Turkish L2 learners were not significantly different from each other in accuracy judgment on OEFFT (p=.426) and OEFNONF (p=.497), but they were significantly different from each other in accuracy on SEFF, SEFNONF and OEFF clauses. The Spanish L2 learners, on the other hand, were significantly less accurate than the native speakers on five types, but they were not significantly different from the Turkish L2 learners, except for OEFNONF clauses. The Turkish L2 group was significantly more accurate than the Spanish group in this category.

Pairwise comparisons (Bonferroni, p<.05) of structure types following the 3x2 ANOVA above, indicated that SENONF clauses were significantly more difficult than the four other structure types including OEFNONF clauses while SEFF clauses were significantly more difficult than OEFF clauses (p=.037), confirming previous findings of Schachter and Yip (1990), White and Juffs (1998), Juffs and Harrington, (1995), which have shown an asymmetry between subject and object extractions from both finite and nonfinite clauses. Furthermore, object extraction from a finite clause with complementizer *that* was significantly more difficult compared to object extraction from finite and nonfinite clauses.

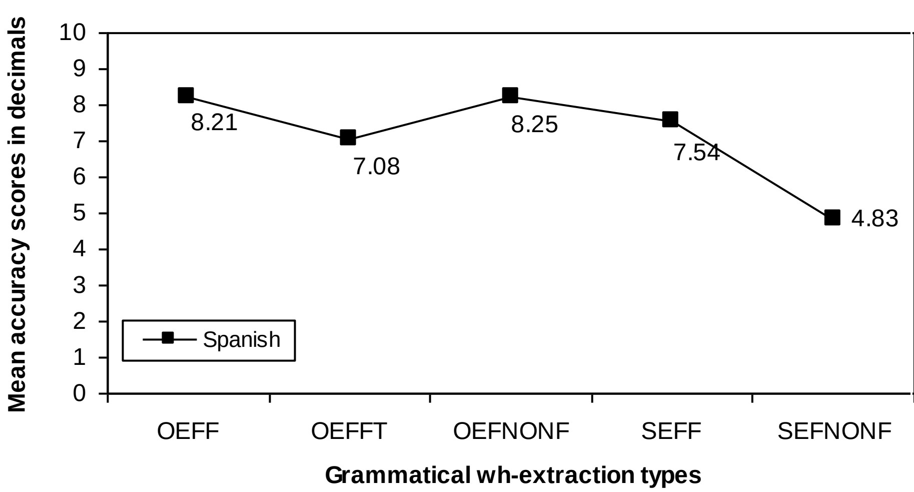
To examine significant differences in language groups by type, an ANOVA was conducted for each language group with type as repeated within-subjects. The first ANOVA conducted with the mean accuracy scores of English native speakers showed a significant main effect for type (F (4, 120) =34.98; p<.01; MSe=49.200). Post-hoc comparisons showed that, among all wh-extraction types, English native speakers were significantly least accurate in judging SEFNONF clauses. However, they did not show any significant difference between object and subject extraction from a finite clause. They had 97% accuracy in subject extractions, and 98% in object extraction from finite clauses. Among the three types of object extractions, they were found to be less accurate in object extractions from a finite clause with complementizer *that.* Figure 3 shows English speakers' mean accuracy scores for five grammatical wh-extraction types.

Figure 3. English speakers' mean accuracy scores on five grammatical wh-extraction types



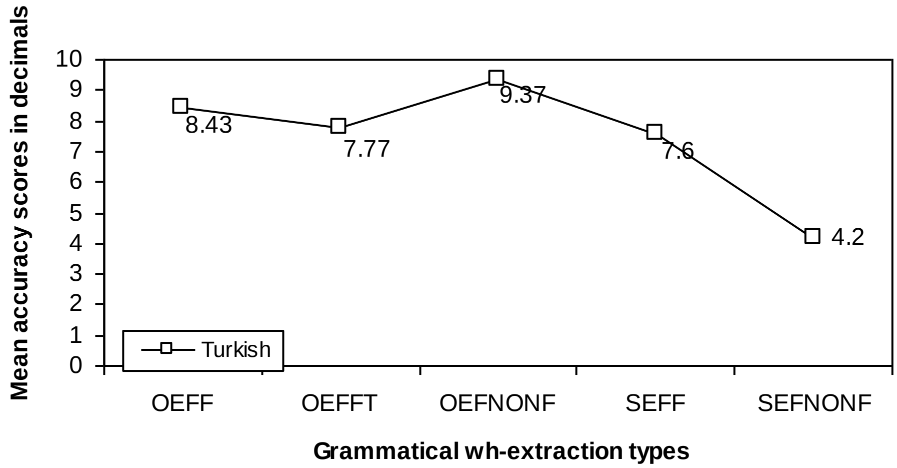
The second ANOVA carried out with the mean accuracy scores of Spanish speakers showed a significant main effect for type (*F* (4, 92) =20.83; p<.01; *MSe*=47.096). Pairwise comparisons of 5 sentence types (Bonferroni, p<.05) revealed that like English native speakers, Spanish speakers had most difficulty in judging SEFNONF clauses among the five sentence types. Although, they were also less accurate in extracting subject from finite clauses (M=7.54) than extracting object from finite clauses (M=8.21), the difference was not statistically significant (p=.650). This suggests that like native English speakers, Spanish speakers demonstrate a strong preference for object rather than subject extractions in nonfinite clauses. However, there was a less strong preference for object extractions from finite clauses, compared to nonfinite clauses. Object extractions from finite clauses with the complementizer *that* were significantly more difficult than object extractions from finite clauses without *that*. Figure 4 provides Spanish speakers' mean accuracy scores for five grammatical wh-extraction types.

Figure 4. Spanish speakers' mean accuracy scores on five grammatical wh-extraction types



The third ANOVA conducted with Turkish speakers' mean accuracy scores on the five grammatical wh-extraction types showed almost the same result. In other words, a significant main effect for type, ((F (4, 116) =41.72; $p<.01$); MSe=114.92) was found Pairwise comparisons (Bonferroni, $p<.05$) of the five types indicated that like English and Spanish speakers, Turkish speakers were significantly least accurate in judging SEFNONF clauses, compared to other types. They were also worse in judging subject extractions from finite clauses (M=7.60) than object extraction from finite clauses (M=8.43), however, this was not a statistically significant difference (p=.483).These results suggest that Turkish speakers show a clear asymmetry in wh-extractions from nonfinite clauses, indicating an object preference over subject from nonfinite clauses, but this preference is not so strong in finite clauses. In addition, unlike Spanish and English native speakers, Turkish speakers did not show a significant difference in judging object extractions from finite clauses with or without complementizer *that.* Figure 5 shows Turkish speakers' mean accuracy scores for five grammatical wh-extraction types.

Figure 5. Turkish speakers' mean accuracy scores on five grammatical wh-extraction types



To sum up, the results of the accuracy analysis on grammatical items showed that although with respect to overall scores, the two L2 groups appeared significantly different from the English native speakers, there were also some similar processing patterns observed across all groups. More specifically, results from the type-by-type analysis revealed that the two L2 groups were found to be similar to the native English speakers in the way they process grammatical wh-extractions across five types (see Figure 2 for a clear picture of this pattern). Furthermore, the two L2 groups, less accurate on subject extraction from long-distance nonfinite and finite clauses, compared to object extractions from finite and nonfinite clauses. The processing difficulty in SEFNONF clauses was significantly higher than the difficulty in SEFF clauses, particularly for the two L2 groups. This suggests an overall preference (i.e., more accurate judgments) for object extractions rather than subject extractions. In addition, the English native speakers and the Turkish learners were equally accurate on object extraction from nonfinite and finite clauses with complement *that*.

These results provide evidence for previous findings of Yips and Schachter (1990), White and Juffs (1998) and Juffs and Harrington (1995), and suggest that all three groups, in particular the L2 groups, have a clear processing difficulty in subject extraction from nonfinite clauses, providing support for subject and object asymmetry in wh-extractions from nonfinite clauses. The same subject-object asymmetry is also found in wh-extractions from finite clauses in conformity with previous findings (e.g., Juffs, 2005). Nevertheless, unlike non-finite clauses, the difference between the subject and object extraction from finite clauses did not reach a statistically significant level.

I will now discuss RT scores of the three groups on the five types of the grammatical wh-extractions to determine how fast participants read these sentences.

5.2.2 Reading times on grammatical wh-extraction types

In the RT analysis of the grammatical wh-extractions, first of all, the RT data obtained from the correct responses of each participant to the five types of the grammatical wh-extractions were screened for normal distribution and outliers. Then, each participant's mean RT for each type was calculated, excluding RTs beyond 2.5 standard deviations from her/his mean RTs. Then, mean RTs for the five types of the grammatical wh-extractions in each language group were checked for normal distribution, and outliers. The results indicated that the RT data were normally distributed for each type in each language group. There were a few outliers and extreme values. Only extreme cases[1] were excluded from the analysis in order not to cause data loss. There were six missing cases in the Turkish group and 1 in the Spanish group due to the incorrect responses. One Spanish participant and one Turkish participant who did not achieve overall accuracy above 60% were excluded from the analysis. My aim is two-fold here: first, I would like to examine whether there is a significant difference between the L2 and the native speakers in terms of RTs spent on five grammatical wh-extractions: secondly, I would like to identify whether RTs on sentence the types which were found difficult to judge are longer than those which were found easy to judge. In other words, I would like to investigate whether or not the sentences that triggered most accurate responses took less RTs. Mean RTs to five grammatical wh-extractions are presented in Table 2 in milliseconds with standard deviations.

[1]Extreme cases refer to cases with values more than 3 box lengths from upper or lower edge of the box. The box length is interquartile range.

Table 3. Mean RTs for five grammatical wh-extraction types

Structure	Spanish (n=23)		Turkish (n=24)		English (n=31)	
	M	SD	M	SD	M	SD
OEFF (n=10)	4459	1154	4059	1038	2940	770
OEFFT (n=10)	4403	1095	4350	995	3354	952
OEFNONF (n=10)	4210	1190	3668	961	2831	722
SEFF (n=10)	4298	1065	4226	1254	2827	774
SEFNONF (n=10)	4530	1169	4650	1041	3362	997
Total (n=50)	4364	1114	4233	1097	3063	843

OEFF (Object Extraction From Finite clause), OEFFT (Object Extraction From Finite clause with *that*), OEFNONF (Object Extraction From Nonfinite clause), SEFF (Subject Extraction From Finite clause), SEFNONF (Subject Extraction From Nonfinite Clause). Mean scores indicate RTs in milliseconds.

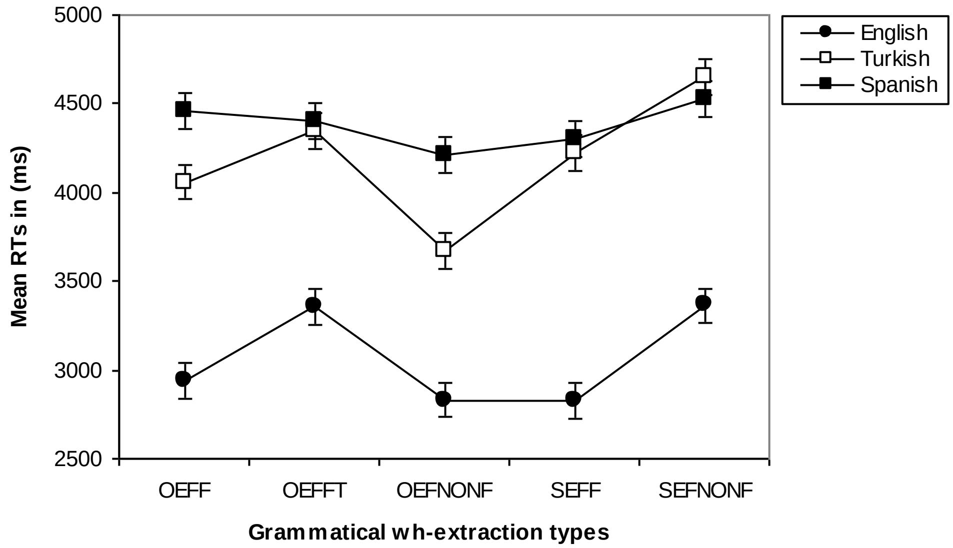
Total mean RTs in Table 3 indicate that the two L2 groups appear slower than the native speakers, but they do not look differ from each other in the overall RTs across five types of grammatical *wh*-extractions. Also, for all groups, SEFNONF clauses triggered the slowest RTs, among all types. This confirms the claim that subject extraction from nonfinite clauses is the most difficult type to process, and that the more difficult the structure is, the longer it takes to process.

In the case of finite clauses, the English native speakers and the Spanish learners appear relatively slower in OEFF clauses than SEFF clauses. The reverse is true for the Turkish learners. Recall from the previous section that all participants including the English native speakers were less accurate on SEFF clauses than OEFF clauses. This means that, unlike the Turkish speakers, the Spanish and English native speaker groups seem to take a long time to read OEFF sentences in which they achieve low accuracy in judgments. However, for the Turkish group, there is a consistency in accuracy and RT scores in the sense that, OEFF is the second most accurate and the second fastest category they are found to be.

The OEFFT clauses received longer RTs for the English native speakers and the Turkish learners than OEFF, OEFNONF and SEFF clauses. To a large extent, we see the same pattern in the Spanish group. Nevertheless, OEFFT clauses took slightly shorter to judge (M=4403msec) than OEFF clauses (M=4459).

Figure 6 illustrates the mean RTs to five grammatical wh-extraction types in milliseconds.

Figure 6. Mean RTs for the five grammatical wh-extraction types by three language groups



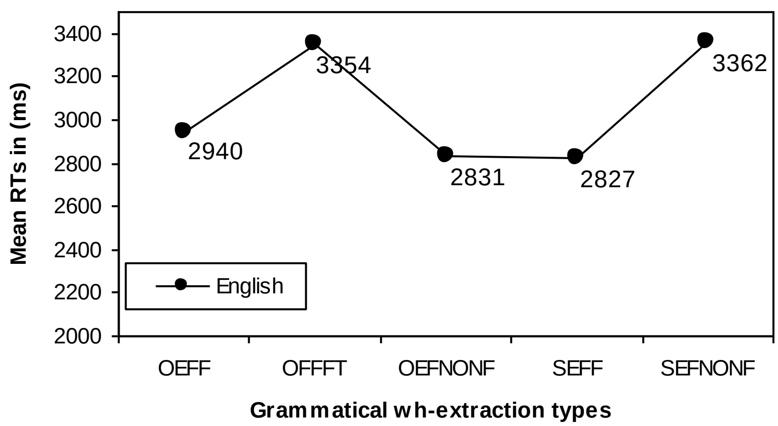
To examine whether variations in language groups by type are statistically significant, a two- way ANOVA was conducted with language as the between-subjects factor and type as the repeated within-subjects factor. The results revealed an overall significant effect for language (F (2, 75) =16.66; p<.01); and a main affect for type (F (4, 300) =20.67; p<.01); and a significant interaction for language and type (F (8, 300)

=3.31; p<.05), which means that differences in language group by type are statistically significant. The post-hoc analysis (Tukey HSD, p<.05) of language groups revealed that Spanish and Turkish speakers of L2 English were significantly slower than the native English speakers in reading five grammatical wh-extractions types, but they were not different from each other.

Pairwise comparisons (Bonferroni, p<.05) of sentence types revealed that the difference between subject extraction from a nonfinite clause and other four types was statistically significant, except for object extraction from a finite clause with complementizer *that*. This shows that all participants spent significantly longer RTs on subject extraction from a nonfinite clause, compared to the other three types but judging *that*-clauses were as slower as sentences involving subject extraction from nonfinite clauses. While we see an RT difference between subject and object extraction from nonfinite clauses, the difference between subject and object extraction from finite clauses was not statistically significant. These findings replicate accuracy results in the sense that there is a RT difference subject and object extractions from nonfinite clauses, but no such difference is found in finite clauses.

To examine significant interaction effects found for type and language, three ANOVAs were conducted; one for each language group, with type as the repeated within-subjects factor. The results from the first ANOVA conducted on English native speakers' mean RTs for five grammatical wh-types indicated a significant main effect for type (F (4, 120)) =19.73; p<.01). Pairwise comparisons of types revealed that English native speakers spent significantly longer RTs for SEFNONF clauses than OEFNONF clauses. No significant difference was found in RTs for subject and object extractions from finite clauses. Furthermore, the second longest RT was observed in object extraction from finite clauses with complementizer *that*. This suggests that native speakers have processing problems with object extraction from finite clause that contain the complementizer *that*. Recall from accuracy results that this category also elicited smaller number of correct responses from native English speakers. Figure 7 below presents English native speakers' mean RTs on five grammatical wh-extraction types.

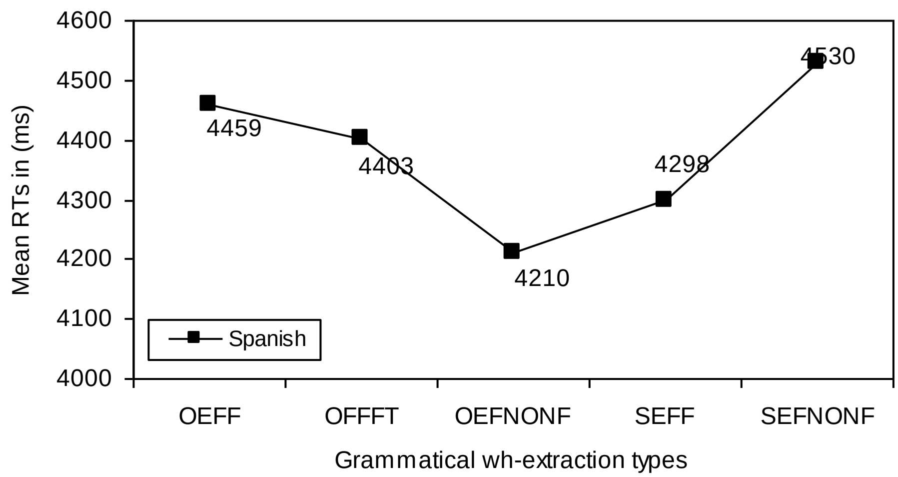
Figure 7. English native speakers' mean RTs on five grammatical wh-extraction types



Analysis of Spanish speakers' mean RTs for five types did not reveal a significant main effect for type ($F (4, 88) = 1.60; p>.05$), suggesting that the differences in Spanish speakers' RTs for five grammatical wh-extraction types were not statistically significant from each other. This implies that no particular sentence type triggered significantly longer RTs. Nevertheless, it is important to note that the Spanish group's RT for the subject extractions from nonfinite sentences was longer than the RTs for other sentence types. Recall that this category triggered more errors in the grammaticality judgment task.

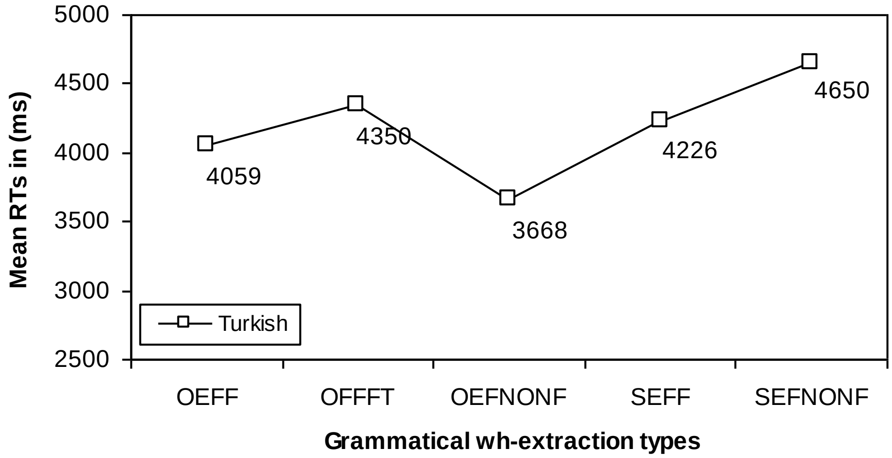
Also, the category on which the Spanish group was most accurate was object extractions from nonfinite clauses. Similarly, they were found to be quicker in reading these sentences than any other sentence types. Figure 8 shows Spanish speakers' mean RTs on five grammatical wh-extraction types.

Figure 8. Spanish speakers' mean RTs for five grammatical wh-extraction types



The third repeated measure ANOVA carried out with Turkish speakers' mean RTs indicated a significant main effect for type ($F(4, 92) = 10.98$ $p<.01$), which means that the differences between grammatical wh-extraction types were statistically significant. Pairwise comparisons of types revealed that like the two other groups, Turkish speakers spent significantly longer RTs on judging SEFNONF clauses than OEFNONF clauses. OEFNONF was the category, which was processed in significantly the shortest time. No significant difference was found in RTs of subject and object extraction from finite clauses. These results support previously reported asymmetry between subject and object wh-extractions from nonfinite clauses. Figure 9 shows Turkish speakers' mean RTs on five grammatical wh-extraction types.

Figure 9. Turkish speakers' mean RTs on five grammatical wh-extraction types



To sum up, the overall results from RTs spent on five types of grammatical wh-extractions revealed that although the L2 groups were significantly slower than the native speaker group, they were not different from each other.

It is important to note that similar to the accuracy results reported earlier, the RT results suggest that all groups seem to have difficulty in processing subject extractions from nonfinite clauses. However, none of the groups show a significant RT difference between subject and object extractions from finite clauses. This suggests that SEFNONF clauses incur more processing load, reflected in the participants' longer RTs, compared to the other types including SEFF clauses, which are also found to be difficult to correctly judge. This confirms the subject-object asymmetry in RTs for nonfinite clauses. Furthermore, the quickest RT in the L2 groups was elicited from object extractions from nonfinite clauses. Recall that the category on which the two L2 groups were most accurate was again this category. Thus, RT results are in line with accuracy judgment scores because low accuracy seems to mean longer RTs for grammatical wh-extractions. In other words, when participants are not sure about the accuracy of a grammatical item, they take longer to process/judge this sentence but longer time does not lead to correct judgment in the end. However, when they take less time to process a sentence, their grammaticality judgment also seems to be correct.

In conclusion, these results confirm the findings of Juffs and Harrington (1995), Juffs and Harrington (1996), White and Juffs (1998), and Juffs (2005), which show that L2 speakers do have the knowledge of constraints on wh-movement as revealed by their judgments on grammatical sentences with wh-movement in English. Although their accuracy on certain structures is lower than the native English speakers, they show the same pattern of sentence processing as the native speakers. In other words, the most accurately and quickly judged categories (or least accurately and slowly) are more or less the same in all groups. The most striking example of this was the sentences involving subject extraction from nonfinite clauses, which are found to be the most problematic category for all groups in terms of accuracy and RT. What is also important to note is that irrespective of the L1 syntactic features, both Spanish and Turkish groups revealed similar results across all categories.

In the following section I will present the results of the ungrammatical wh-sentences with wh-movement violation.

5.2.3 Accuracy judgments on ungrammatical wh- extraction types

For the accuracy analysis of five types of ungrammatical wh-sentences involving island violations, accuracy data were screened for outliers, missing values and normal distribution. As in grammatical wh-sentences, the accuracy data were not normally distributed within five types. However, the number of participants in each language group and variance among five sentence types were roughly equal. Therefore parametric tests were used for analyzing accuracy data from the ungrammatical wh-sentences. There were two missing values, one in Turkish and one in Spanish. Two participants (one from the Spanish group, and one from the Turkish group), who did not achieve an overall accuracy above 60% were excluded from the analysis. Mean accuracy scores for five types of ungrammatical wh-extraction types are presented in Table 3 in decimals.

Table 4. Mean accuracy scores for ungrammatical wh-extraction types

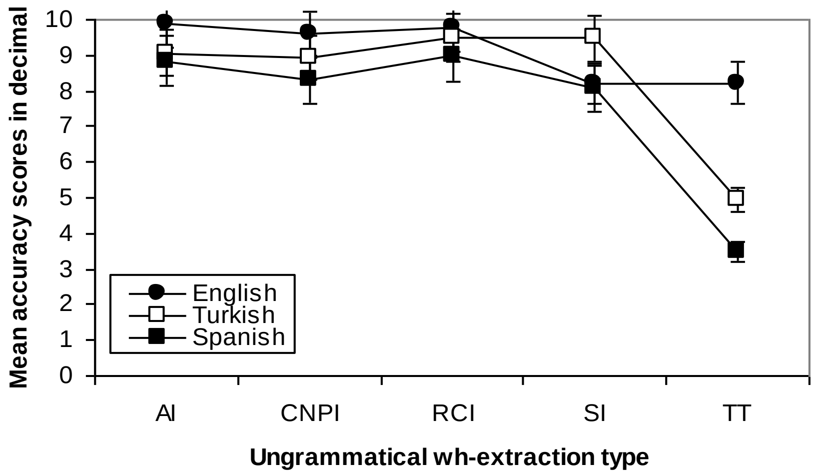
Structure	Spanish (n=24)		Turkish (n=30)		English (n=31)	
	M	SD	M	SD	M	SD
AI (n=10)	8.83	1.17	9.07	1.34	9.90	.301
CNPI (n=10)	8.33	1.79	8.93	1.41	9.58	.672
RCI (n=10)	9.00	1.33	9.50	.900	9.77	.560
SI (n=10)	8.08	1.50	9.47	.900	8.19	1.68
TT (n=10)	3.50	1.59	4.93	2.49	8.23	1.86
Total (n=50)	7.55	1.47	8.38	1.41	9.13	1.01

AI (Adjunct Island), CNPI (Complex Noun Phrase Island), RCI (Relative Clause Island), SI (Subject Island), TT (*That*-trace violation)

Overall results indicate that all groups were more accurate at correctly rejecting ungrammatical wh-extractions with island violation than accepting grammatical wh-extractions (compare Table 1 to Table 4 above). Also, the difference between the English native speakers and the two L2 groups with respect to accuracy on ungrammatical wh-extractions seems to be lower than that in the grammatical wh-extractions. That means that both L2 groups were more successful in judging ungrammatical items than judging grammatical items. Furthermore, the Turkish group appears almost as accurate as the English native speakers overall, except for wh-extraction with *that*-trace violation. Indeed, overall mean accuracy increases to 9.24 in the Turkish group and to 8.56 in the Spanish group in ungrammatical items if *that*-trace sentences are excluded in the analysis. These findings strongly suggest that the two L2 groups have demonstrated knowledge of constraints on wh-movement by correctly rejecting ungrammatical wh-extractions in L2 English. Recall that Spanish and English are both overt wh-movement languages with similar island constraints. They differ from each other only with respect to the bounding nodes in two languages (see Chapter 3). Turkish, on the other hand, is a wh-in-situ language, but has covert movement at LF and overt-movement through scrambling and exhibits certain constraints on wh-movement as in English and Spanish.

Mean accuracy scores for each type in each language group in Table 3 also show that rejecting wh-extractions with *that*-trace violation was very difficult for all groups. This was particularly obvious in the Spanish group. This suggests a local L1 influence for the Spanish learners whose L1 allows a subject trace in the embedded clause following the complementizer *that*. Figure 10 shows the mean accuracy scores for five types of ungrammatical sentences.

Figure 10. Mean accuracy scores for five ungrammatical types by three groups



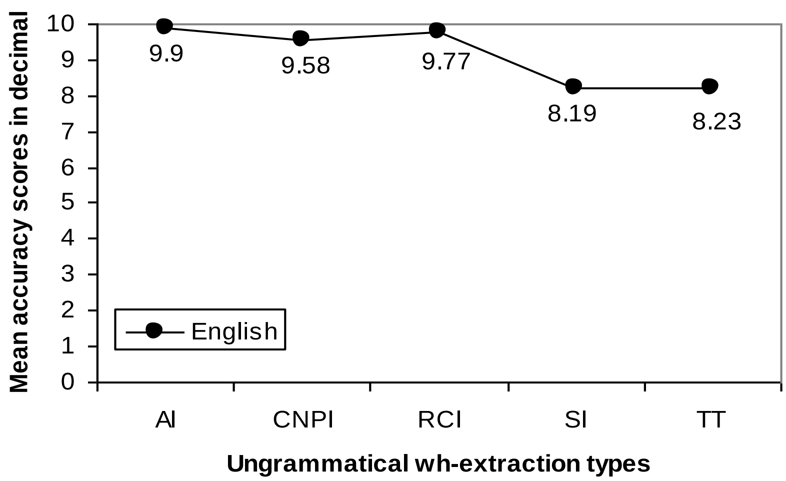
To examine whether the differences in language groups by types are statistically significant, a two-way ANOVA was carried out with language as the between-subjects factor and type as the repeated within-subjects factor. The results indicated an overall significant effect for language (F (2, 82) =30.22; p<.01); a significant main effect for type (F (4, 328) =122.27; p<.01; *MSe*=214.65) and a significant interaction for language and type (F (8, 328) =17.28; p<.01; *MSe*=25.87). Post-hoc analysis (Tukey HSD, p<.05) of language groups revealed that L2 speakers were significantly less accurate than the native speakers of English in judging these five types of ungrammatical wh-extractions. Also, Spanish speakers were significantly less accurate than the Turkish speakers. A further two-way ANOVA was conducted with the mean accuracy scores only across four ungrammatical types excluding the type with *that*-trace violation where the two L2 groups had dramatically the lowest accuracy. Results indicated that the native English speakers and the Turkish L2 learners were equally accurate on four types of ungrammatical wh-extractions (p=.806). However, the Spanish group was significantly less accurate on four ungrammatical wh-extraction types than both the English native speakers ($p=.001$) and the Turkish learners ($p=.004$). That means that even when we exclude the results of *that*-trace violation, where the Spanish group had the least accuracy, we see that the Spanish group was still significantly different from the Turkish and the native English groups.

These findings suggest that despite a wh-in-situ the L1 background, the Turkish group was as accurate as the native speakers in rejecting ungrammatical wh-extractions with island constraints except for *that*-trace violation, where they obtained a mean score of 4.93 in accuracy. The Spanish group, on the other hand, was less accurate than both the native and the Turkish speakers on all five ungrammatical types.

Pairwise comparisons (Bonferroni, $p<.05$) of overall accuracy scores for five ungrammatical wh-extraction types revealed that incorrect rejection of wh-sentences with *that*-trace (TT) violations was significantly the most prevalent error. This was followed by wh-sentences with subject-island (SI) violation, and the complex NP island (CNPI) violation. The most correctly rejected wh-sentences were those with the relative clause island (RCI) violations and adjunct island (AI) violation. These results confirm that wh-extractions with *that*-trace violation is the most problematic type while wh-extractions with the RCI violation is the most successful type.

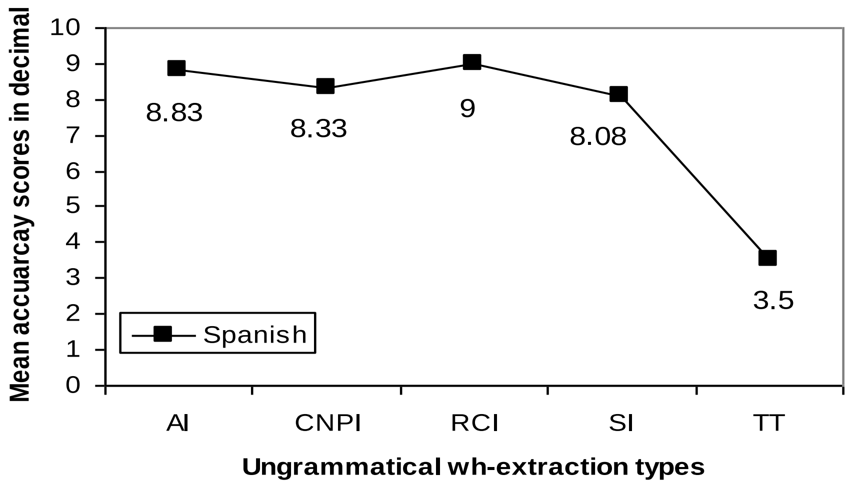
To explore the significant interaction for type and language, a repeated measure ANOVA per group conducted. The first ANOVA carried out with English native speakers' mean accuracy scores across five types of wh-extractions revealed a significant main effect for type ($F(4, 120) = 15.50$; $p<.01$; $MSe=22.55$), showing significant differences between the five types. Pairwise comparisons of the five types revealed that the ungrammatical category on which the control group was least accurate was wh-sentences involving subject island violations. Equally less accurate scores came from judgments on wh-sentences with *that*-trace violations. Figure 11 illustrates the distribution of English native speakers' mean accuracy scores for five ungrammatical wh-extraction types.

Figure 11. English native speakers' mean accuracy scores for ungrammatical wh-extraction types



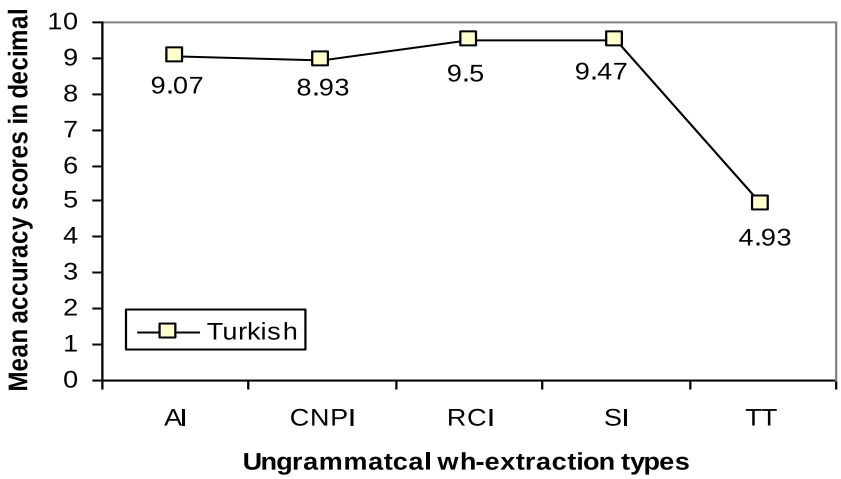
The second ANOVA carried out with the mean accuracy scores of Spanish speakers for five types of ungrammatical wh-extractions showed a significant main effect for type (F (4, 92) = 64.41; p<.01; *MSe*= 126.30). This means that the differences between 5 types were statistically significant. Pairwise comparisons (Bonferroni, p<.05) of types showed that the ungrammatical category on which the Spanish group was least accurate was sentences with *that*-trace violations. No significant difference was observed among four other sentence types. Figure 12 shows Spanish speakers' mean accuracy scores for the five ungrammatical wh-extraction types with wh-movement violation.

Figure 12 Spanish speakers' mean accuracy scores for ungrammatical wh-extraction types



The third ANOVA conducted on Turkish speakers' mean accuracy scores showed a significant main effect for type, (F (4,116) =59.46; p<.01; *MSe*=113.19), which means that the differences between five types of ungrammatical wh-extractions were statistically significant. Pairwise comparisons (Bonferroni, p<.05) of types revealed that like native English and Spanish speakers, Turkish speakers were significantly less accurate at detecting *that*-trace violations, compared to other four types. It should be noted that, an independent- sample t-test conducted with the Spanish and Turkish speakers' mean accuracy scores for the type with *that*-trace violation showed that Spanish speakers were significantly less accurate than Turkish speakers at judging ungrammatical wh-sentences with *that*-trace (t (52)=-2.45; p<.05). Figure 13 shows the distribution of Turkish speakers' mean accuracy scores across five types of ungrammatical wh-extractions.

Figure 13. Turkish speakers' mean accuracy scores for ungrammatical wh-extraction types



To sum up, the overall results of accuracy responses to five types of ungrammatical wh-extractions indicate that the Turkish L2 learners are as successful as the English native speakers at correctly rejecting ungrammatical wh-sentences, except for sentences with *that*-trace violation. However, although the Spanish learners are better at judgment of ungrammatical wh-extractions than that of grammatical wh-extractions, they are significantly less accurate than both the English native speakers and the Turkish learner in all categories including those involving *that*-trace violation. These findings suggest that the two L2 groups are sensitive to constraints on wh-movement in L2 English, but the Turkish learners are as sensitive as the native speakers to those constraints.

The finding that the two L2 groups, particularly the Spanish speakers are less accurate than English native speakers in judging ungrammatical wh-sentences with *that*-trace can be accounted for by a local influence of the L1 Spanish, which allows extracting a subject from an embedded clause leaving a trace following an overt complementizer *that*. In case of Turkish, L1 cannot account for incorrectly accepting wh-extractions with *that*-trace violations because Turkish does not have an overt complementizer *that* in embedded clauses. Nevertheless, the role of pro-drop L1 may have something to do with the Turkish and Spanish speakers' low accuracy in this type. In some earlier L2 studies, the pro-drop properties were associated with the absence of *that*-trace in Spanish (White, 1989).

Besides sentences with *that*-trace (TT) violations, both English and Spanish native speakers were found to be significantly less accurate at rejecting ungrammatical wh-sentences with subject island (SI) violation. However, other three types (adjunct island (AI), complex NP island (CNPI), relative clause island (RCI) were not as difficult for them. This might be related with some sort of processing difficulty associated with SI constructions. Nevertheless, subject islands like CNPI, AI, and RCI, are strong islands which block extractions (e.g., Saito, 1992; Szabolcsi & den Dikken, 1999); therefore one would not except more difficulty in SI items. In other words, it is difficult account for this lowest accuracy in the SI category. The results of self-paced word-by-word reading experiment may shed light on this issue. For the Turkish group, SI items did not pose particular problems. The most difficult type for the Turkish group was That-trace effect.

The following section presents the result of RTs to five types of ungrammatical wh-extractions.

5.2.4. Reading times on the ungrammatical wh-sentences

As in the analysis of RTs for the grammatical wh-sentences, only RTs spent on correct responses across five types of ungrammatical wh-extractions were included in the analysis. RTs beyond 2,5 SD of each participants' mean RTs were not included in the analysis. The mean RTs from five types of ungrammatical wh-sentences in each language group were checked for normal distribution, outliers and missing values. The results indicated that the data from three language groups for five types of ungrammatical wh-extraction were normally distributed. There were two missing values in the *that*-trace category: one from the Turkish group and one from the Spanish group due to incorrect answers to 10 sentences involving *that*-trace violations. One participant from each L2 group was eliminated from the analysis because their overall low accuracy score remained below 60%. The ungrammatical RT data were analyzed using parametric tests. Mean RTs for five ungrammatical wh-extraction types are presented in milliseconds in Table 4.

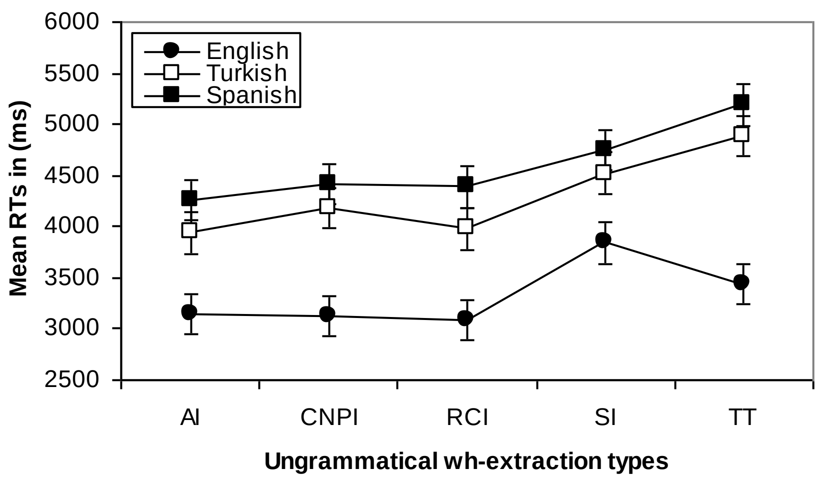
Table 5. Mean RTs for five ungrammatical wh-extraction types

Structure	Spanish (n=24)		Turkish (n=30)		English (n=31)	
	M	SD	M	SD	M	SD
AI (n=10)	4255	1078	3938	917	3142	933
CNPI (n=10)	4417	1168	4186	973	3130	884
RCI (n=10)	4387	1220	3980	1057	3084	920
SI (n=10)	4746	1223	4522	1185	3843	1145
TT (n=10)	5190	1245	4885	1084	3441	1155
TOTAL Average (50)	4599		4302		3328	

AI (Adjunct Island), CNPI (Complex Noun Phrase Island), RCI (Relative Clause Island), SI (Subject Island), TT (*That*-trace)

As can be seen from Table5, the English native speakers were faster than the two L2 speakers; and the Turkish speakers were faster that the Spanish speakers in correctly rejecting all ungrammatical wh-extractions. Nevertheless, the ranking of categories in terms of their mean RTs was the same for the Spanish and the Turkish groups. For example, both groups spent longer RTs for the wh-sentences with *that*-trace violation and shortest RTs for the adjunct island sentences. On the other hand, English native speakers demonstrated the longest RT for the wh-sentences involving subject island violations. The category on which native English speakers spent least time was sentences with relative clause islands. Figure 14 shows the mean RTs of English control and L2 groups for five ungrammatical wh-extraction types

Figure 14. Mean RTs of language groups across five ungrammatical wh-extraction types

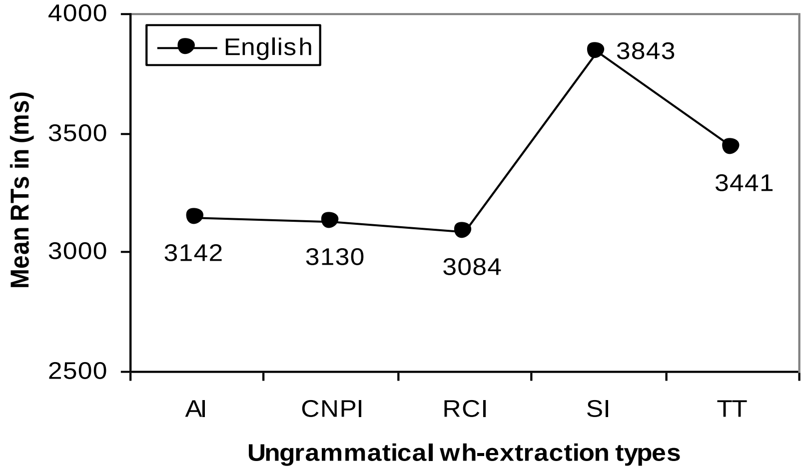


To investigate variations in language groups and types, a two-way ANOVA was carried out with language as the between-subjects factor and type as the repeated within-subjects factors. This analysis revealed an overall significant effect for language ($F(2, 76) = 14.80$; $p<.01$); a significant main effect for type ($F(4, 304) = 35.03$; $p<.01$) and a significant interaction for language and type (F (8, 304) =3, 98; p<.05). A post-hoc analysis (Tukey HSD, p<.05) of language groups showed that English native speakers were significantly faster than the two L2 groups in rejecting ungrammatical wh-extractions. However, L2 groups were not significantly different from each other. Pairwise comparisons (Bonferroni, p<.05) of five types of ungrammatical wh-extractions, on the other hand, showed that RTs spent on the wh-sentences with *that*-trace violations and with subject-island violations were significantly longer, compared to other types, but the difference between these two types was not statistically significant.

To examine significant interaction for type and language, a separate Repeated Measures ANOVA was conducted for each language group with type as the repeated within-subjects factor.

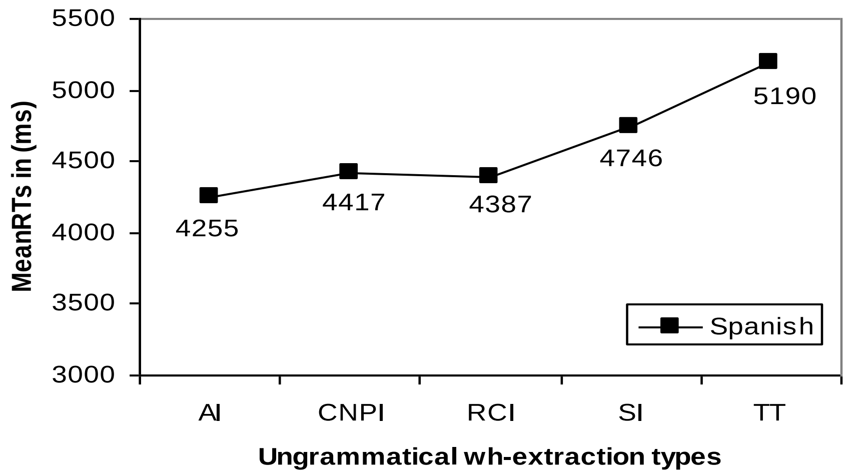
The first ANOVA conducted on English native speakers' mean RTs indicated a significant main effect for type (F (4, 120) =20.18; p<.01). That is, a significant difference in RTs was found among five sentence types. In the pairwise comparisons, it was found that English native speakers spent significantly longer time reading wh-sentences with subject island violations than reading other sentence types. The second longest RT was found in sentences involving *that*-trace violations. Recall that ungrammatical sentences in these two categories triggered more errors than other three categories in native English speakers' data. Figure 15 shows English native speakers' mean RTs for five ungrammatical wh-extractions.

Figure 15. English native speakers' mean RTs for five ungrammatical wh-extraction types



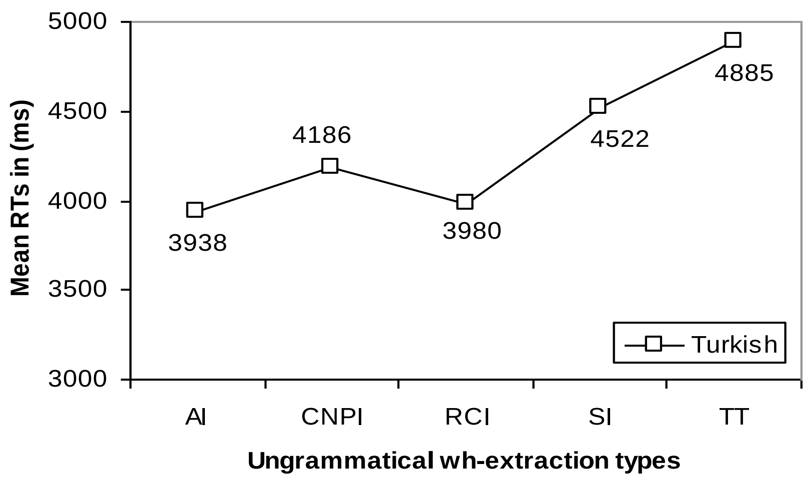
Another ANOVA conducted on mean RTs of Spanish speakers on five ungrammatical wh-sentence types indicated a significant main effect for type (F (4, 80)= 5.53; p<.01), which means that there were significant differences in mean RTs across five types. Pairwise comparisons of types indicated that Spanish speakers were significantly slower at judging the ungrammatical wh-sentences with *that*-trace and subject island violations. It is important to note that the control group was also found to spent longer time reading these two types of ungrammatical sentences. The category on which the Spanish group spent least amount of time was ungrammatical sentences involving adjunct islands. Recall that this was one of the categories in which the Spanish group was found to be the least accurate. This suggests that the slow RT implies less accuracy. In other words, the sentence type in which the participants are found to be slowest is also the least accurate category. This pattern is also observed in the native control group (see results above). Figure 16 shows the distribution of Spanish speakers' mean RTs across five wh-types.

Figure 16. Spanish speakers' mean RTs for five ungrammatical wh-extract types



The third ANOVA carried out on Turkish speakers' mean RTs on five wh-extraction types also revealed a significant main effect for type (F (4, 104) =22.52; p<.01), which means that there were significant differences between RTs spent on five types of ungrammatical wh-extractions. Pairwise comparisons of types showed that like Spanish speakers, Turkish speakers spent significantly longer time for reading the wh-sentences with both *that*-trace and subject island violations and shorter time for reading sentences with adjunct islands. Figure 17 provides Turkish speakers' mean RTs for five ungrammatical wh-extraction types. It is interesting to note that detecting subject island violations required relatively longer time for the Turkish group. However, the Turkish participants were accurate in their final judgments (compare Figure 13 to Figure 17). Nevertheless, for other categories we can still observe the pattern where lower accuracy is equated with slower RT (see for example, CNPI and TT results).

Figure 17. Turkish speakers' mean RTs on five ungrammatical wh-extraction types



To sum up, results from overall RTs spent on five ungrammatical wh-extraction types reveal a difference between the L2 groups and the control group. This difference was particularly obvious in the Spanish, who obtained the lowest mean RT score across categories. Despite this proceeding time difference, we observe striking similarities among three groups in terms of the complexity hierarchy. In other words, what took longer to read for native speakers was almost the same in L2 groups. Similarly what took shorter time to read was also the same in all groups to a large extent. For example, for all groups, ungrammatical wh-sentences with *that*-trace and sentences with subject islands appeared to take longer to read. Sentences with adjunct islands, complex NP islands, and with relative clause islands, however, did not require a long time to process. Furthermore, it also appeared that there is a reverse relationship between accuracy and processing time; when processing time increases, accuracy drops.

5.2.5. Summary of the accuracy and RT analysis for the grammatical and ungrammatical wh-extractions

The accuracy and RT analyses of grammatical and ungrammatical wh-extractions can be summarized as follows: First, the adult L2 learners' accuracy is similar to that of the English native speakers in judging grammatical as well as ungrammatical wh-extractions. Like the native speakers, they are more accurate on ungrammatical wh-extractions than grammatical wh-extractions. Moreover, the Turkish L2 learners are as accurate as the English native speakers on some grammatical wh-extractions (e.g., object extraction from a nonfinite and finite clause with *that)* and all ungrammatical wh-extractions except for wh-extraction with *that*-trace violation. These results suggest that end-state L2 learners can ultimately have knowledge of constraints on wh-movement in the L2.

Second, the two L2 groups show almost an identical order of difficulty, which is also similar to that of the English native speakers in processing grammatical wh-extractions. More specifically, the two L2 groups demonstrate more errors and longer RTs in processing subject extraction from nonfinite clauses and finite clauses. This confirms previous findings of White and Juffs (1998), and Juffs and Harrington (1995), where an asymmetry between subject and object extractions from finite and nonfinite clauses was found.

Third, for all groups, object extractions from finite clauses with complementizer *that* trigger more errors and longer RTs than object extractions from finite clauses without the complementizer *that.*

Fourth, among five types of ungrammatical wh-extractions, wh-extractions with *that*-trace violations and wh-sentences with subject island violations seem to be more difficult to process for all groups. The two L2 groups, particularly the Spanish participants are worse in rejecting ungrammatical wh-extractions with *that*-trace violation, suggesting L1 influence of Spanish, which allows subject extractions from embedded clauses in the presence of the complementizer *that*. Failure in rejection of subject island violation, on the other hand, suggests an illicit gap inside the subject island associated with wh-filler. The results of the self-paced word-by-word reading experiment will provide a clearer picture about the loci of the processing problems in both grammatical and ungrammatical wh-extractions.

In the following section, I will discuss the accuracy and RT analysis for grammatical and ungrammatical wh-sentences in the moving window condition.

5.3. Results from the moving window condition

Recall that in the self-paced word–by-word reading task, participants were asked to read a sentence on a computer screen in a word-by-word fashion. At the end of each sentence, participants judged the grammaticality of the sentence by pressing the "Yes" or "No" button on the keyboard. In the meantime, the computer recorded RTs spent on each word of the sentence and collected participants' accuracy responses for each sentence. As in the full-sentence condition, in the moving window condition, participants read and judged a total of 100 sentences (50 grammatical and 50 ungrammatical). The same grammatical and ungrammatical wh-extractions were used in the moving window condition as in the full sentence condition because word-by word reading of these extractions allow us to find out the locus of the processing difficulty the participants might have experienced during processing of grammatical and ungrammatical wh-extractions.

In this section, I will first present the results of accuracy analyses conducted for five types of grammatical wh-extractions from both finite and nonfinite clauses. Following the accuracy scores, word-by-word RTs for each sentence type will be discussed. I will then discuss the accuracy and RT analyses of five types of ungrammatical wh-sentences. The main aim in these analyses is to examine the RTs spent on each word of a sentence and to identify the critical region in each sentence types that creates more processing load for participants. Therefore, the results of the moving window condition will add to our understanding of which regions of wh-sentences cause more processing problems for native speakers and L2 learners. Thus, the moving window condition can be regarded as complementary to the full-sentence condition.

5.3.1. Accuracy judgments on the grammatical wh-extractions

Accuracy judgment data for five types of grammatical wh-extractions were first checked for normal distribution, missing values, and outliers. As in the full-sentence condition, the accuracy data were not normally distributed due to the nature of the OGJT, in which the accuracy scores ranged between 0-10 for each sentence type. However, the number of participants and the variance of five wh-extraction types were roughly equal. Therefore, I used parametric tests for the analysis. There were no missing values. Outliers and extreme values were kept to avoid data loss, which is as important as normal distribution. Three participants from the Spanish group were excluded from the analysis due to the low accuracy score they obtained in grammaticality judgments. Table 5 shows mean accuracy scores for five types of grammatical wh-extractions in the moving window condition (MWC).

Table 6. Mean accuracy scores for five grammatical wh-extraction types in the MWC

Structure	Spanish (n=21)		Turkish (n=31)		English (n=31)	
	M	SD	M	SD	M	SD
OEFF (n=10)	8.71	1.10	8.84	1.50	9.71	.588
OEFFT (n=10)	8.24	1.55	8.13	1.45	7.81	1.89
OEFNONF (n=10)	8.81	1.47	9.26	.893	9.55	.850
SEFF (n=10)	8.57	1.80	8.71	1.32	9.58	.564
SENONF (n=10)	6.38	2.33	5.16	2.45	7.52	1.69
Total (n=50)	8.14	1.65	8.02	1.52	8.83	1.12

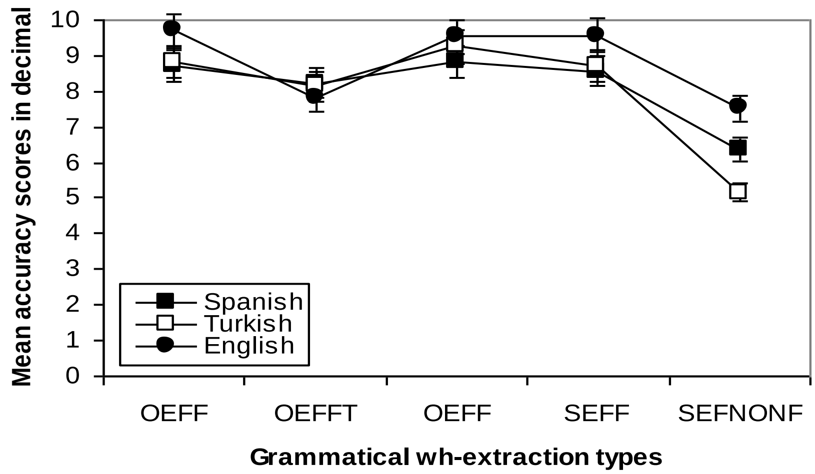
OEFF (Object Extraction From Finite clause), OEFFT (Object Extraction From Finite clause with *that*), OEFNONF (Object Extraction From Nonfinite clause), SEFF (Subject Extraction From Finite clause), SEFNONF (Subject Extraction From Nonfinite Clause)

As in the full-sentence condition, mean accuracy scores (out of 10) on Table 6 show that overall, the two L2 groups performed better in the moving window condition than in the full-sentence condition, but the English native speakers' accuracy scores remained the same. Overall mean accuracy across five types show that the difference between the English native speakers and the two L2 groups was smaller in the moving window condition than that in the full-sentence condition. Furthermore, subject extraction from nonfinite clauses again appears to be the main reason of the difference between the English native speakers and the two L2 groups, particularly the Turkish L2 learners. A comparison of Table 2 and Table 6 shows that both the full-sentence condition and the moving-window condition elicited almost identical accuracy order for grammatical items across subjects.

The accuracy results of the moving-window condition show that the two L2 groups pattern similarly to the native speakers in processing grammatical wh-construction in the sense that for all three groups, the least accurate judgments came from subject extraction from nonfinite clauses. The Spanish and the Turkish groups were less accurate than the native speakers, but they were not different from each other either in terms of mean accuracy or in terms of accuracy ranking of grammatical sentence types. The mean accuracy scores on finite clauses show that both the English native speakers and the two L2 groups were relatively less accurate on subject extractions compared to object extractions, but the difference is not as large as it is in nonfinite clauses.

Figure 18 shows mean accuracy scores of three language groups over five types of grammatical wh-extractions.

Figure 18. Mean accuracy scores for five types of grammatical wh-extractions in the MWC



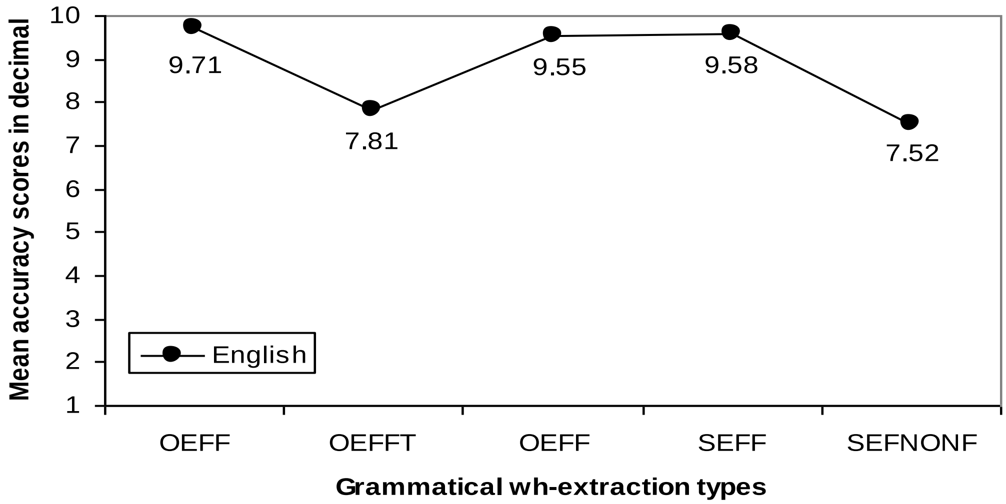
Results from the two-way ANOVA conducted on mean accuracy scores of five grammatical wh-extraction types, with language as the between-subjects factor and type as the repeated within-subjects factor indicated similar significant effects to those in the full sentence condition; an overall significant effect for language (*F* (2, 80) =9.36; p<.01); and a significant main effect for type (*F* (4,320) =54.91; p<.01) *MSe* = 114.62) and a significant interaction for language and type (*F* (8.320) =4.29; p<.01; MSe= 8.96). Both L2 groups were significantly less accurate than native speakers, but they were not different from each other. However, when the mean accuracy scores for subject extraction from nonfinite clauses are excluded from the analysis, the difference between the English native speakers and the Turkish learners becomes insignificant (p=.126); and the difference between the Spanish learners and the English native speakers becomes almost insignificant (p=.49).

As for types, for all groups, subject extraction from a nonfinite clause was significantly more difficult than the other types. Object extractions from nonfinite clauses, on the other hand, was the category on which all three groups were quite accurate. This suggests that both native speakers and L2 learners have object preference in wh-extractions from nonfinite clauses. Nevertheless, the difference between subject and object extraction from a finite clause was not statistically significant, suggesting that no such a strong asymmetry holds in finite clauses. Again, similar to what was observed in the full sentence condition, the second most problematic category was object extraction from a finite clause with complementizer *that*.

Individual analysis of English native speakers' data confirmed a subject–object asymmetry in nonfinite clauses but not in finite clauses. In other words, English native speakers were significantly less accurate in subject extractions from nonfinite clauses, compared to object extraction from nonfinite clauses. However, the difference between subject and object extractions from a finite clause was not significant. While most of the errors appeared in subject extractions from nonfinite clauses, the second most problematic category was object extraction from finite clauses with complementizer *that*. Most accurate judgments were obtained for object and subject extraction from finite clauses. These results are similar to those we found in the full-sentence condition.

Figure 19 displays English native speakers mean accuracy scores across five types of wh-extractions.

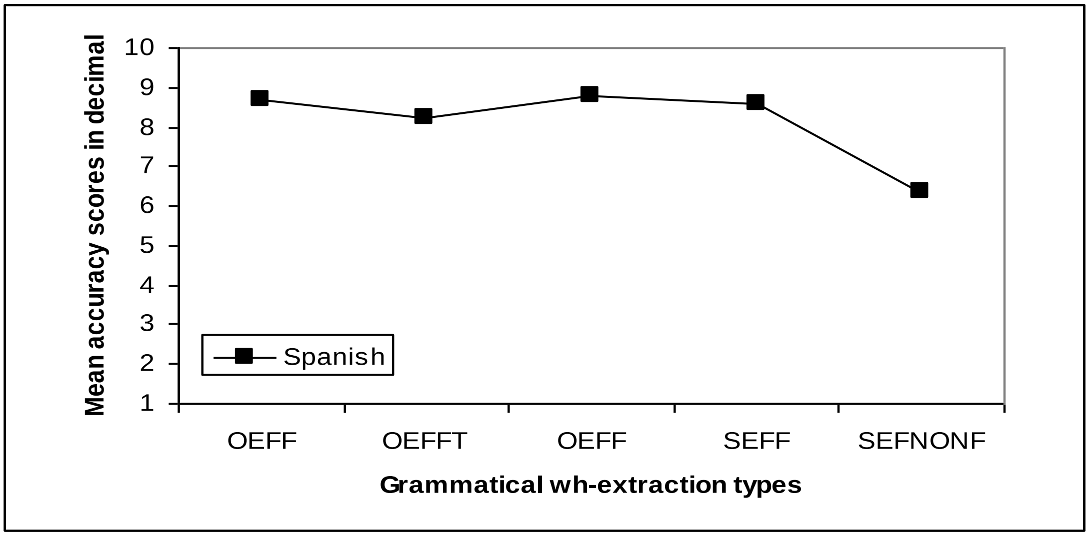
Figure 19. English native speakers' mean accuracy scores for grammatical wh-Extractions in the MWC



Results from the analysis of Spanish speakers' mean accuracy scores over five types of wh-extractions were similar to those of the English native speakers in that the category on which Spanish speakers were significantly less accurate was the subject extraction from nonfinite clauses. Judgments for object extractions from nonfinite clauses, on the other hand, were quite accurate, suggesting that the Spanish group, like the English native speakers, had object preference over subjects in wh-extractions from nonfinite clauses. Also, as in the English control group, the difference between subject and object extractions from finite clauses was not statistically significant. This suggests that subject extraction from finite clauses is not as difficult as subject extraction form nonfinite clauses for the two groups. As we saw earlier in the full sentence condition, the Spanish group also had low accuracy on object extraction from finite clauses with *that*. It is important to note that, the Spanish group demonstrated a difficulty hierarchy similar to the one observed in native English speakers in parsing grammatical wh-constructions.

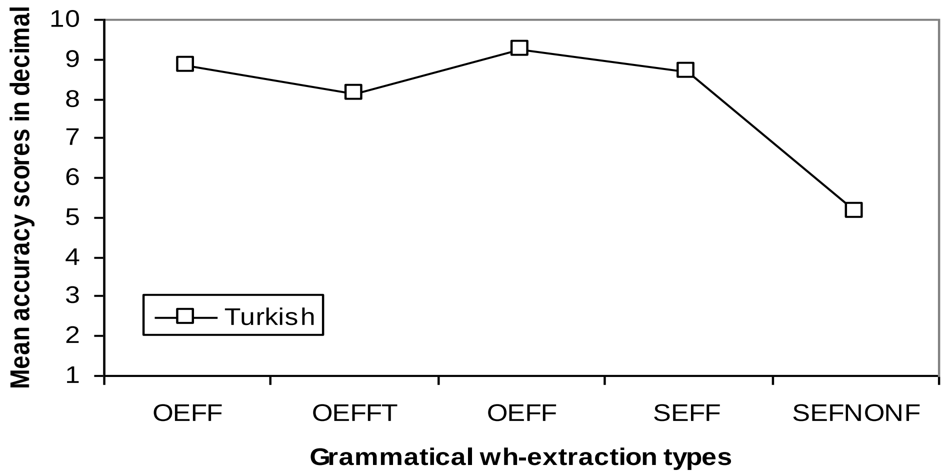
Figure 20 shows the Spanish speakers' mean accuracy scores for five types of grammatical wh-extractions.

Figure 20. Spanish learners' mean accuracy scores for grammatical wh-extractions in the MWC



The analysis of Turkish speakers' mean accuracy scores over five types of wh-extractions revealed an accuracy order identical to the one obtained from the Spanish group. Indeed, the Turkish speakers were not different from native English speakers either; in nonfinite clauses, they did significantly worse in subject extractions but did quite well in object extractions, revealing again an asymmetry between subjects and objects in nonfinite clauses. However, neither subject nor object extraction from finite clauses was problematic. Figure 21 shows Turkish speakers' mean accuracy scores on five types of wh-extractions.

Figure 21. Turkish speakers' mean accuracy scores grammatical wh-extractions in the MWC



To summarize, results from accuracy judgments in grammatical wh-sentences in the moving window condition appear to be similar to the results obtained in the full-sentence condition for all groups. L2 speakers, specifically the Turkish speakers, were not significantly different from the English native speakers in their judgments to grammatical wh-extractions, except for subject extraction from nonfinite clauses. The two L2 groups were significantly less accurate than English native speakers in subject extractions from nonfinite clauses, but not different from each other.

It is, however, important to note that despite slight quantitative differences between the L2 learners and the native speakers, all groups displayed the same accuracy order across grammatical sentence types. For example, all groups demonstrated a subject-object asymmetry in wh-extraction from a nonfinite clause in the sense that while the most difficult wh-extraction type for all participants was subject extractions from nonfinite clauses, no such difficulty was observed in object extraction from nonfinite clauses. Also, the subject-object extraction difference in finite clauses was not significant for any of the groups. Furthermore, for all groups, object extractions from a finite clause with complementizer *that* was more difficult to judge than object extractions from finite clauses that do not include *that.* That suggests that all groups have difficulty in accepting object extractions such as '*What does the inspector think that the boy stole from home?'* English native speakers were slightly less accurate in judging this type than the Spanish and Turkish groups. This can be accounted for by the low frequency of the complementizer *that* with verbs like *believe* and *think* in object extractions from finite clauses in English.

As a final note, it is important to mention that a similar accuracy hierarchy has been found in both full-sentence and the moving window conditions. Nevertheless, in terms of overall mean accuracy, only English native speakers performed identically in both conditions (89% accuracy). However, L2 speakers were more accurate in the self-paced word-by-word reading in the moving window condition than the full-sentence condition (compare Table 1 with Table 5). This might be due to the fact that the full sentence condition, which was presented prior to the moving window condition, might have had facilitative effects on their judgments. Thus, participants might have benefited from the second presentation of similar sentences.

In the next section, I will discuss the results of RTs spent on five types of wh-extractions in the self-paced word-by-word reading in moving window condition.

5.3.2. Word-by-word reading times for grammatical wh-extractions

This section reports on the analyses of self-paced word-by-word reading times (RTs) to only correct responses involving grammatical wh-sentences. I will start discussing word-by-word RTs spent on subject and object extractions from finite clauses. Then, I will discuss RTs for subject and object extractions from nonfinite clauses. Finally, I will present RTs spent on object extraction from finite clauses with complementizer *that*, comparing with object extraction from finite clauses without complementizer *that*.

5.3.2.1. Word-by-word reading times for object and subject extractions from finite clauses

For the statistical analysis, RT data were screened, by language group, for each word in each sentence type, for outliers of RTs that are beyond ±2.5 *SD* from the mean of that language group. As in the previous analyses, only extreme cases were eliminated in order not to cause data loss. This elimination caused only 6.5% loss in the Turkish RT data from subject and object extractions. RT data were found normally distributed (Kolmogorov, Smirnov, $p<.05$) in each language group, therefore parametric tests were used to analyze RTs for subject and object extractions from finite clauses. Also, three participants from the Spanish group were eliminated from all analyses due to their total accuracy scores, which were below 60%. To note, the first and the last words in each sentence were excluded from all statistical analyses because the first words were read at approximately the same speed by all participants, while last words took longer RTs which was assumed due to the “wrap up effect” at sentence final position.

Recall that previous studies (Yip and Schachter, 1990; White and Juffs, 1998; Juffs and Harrington, 1995; 1996; Juffs 2005) have reported a subject/object asymmetry in wh-extractions from finite clauses. Furthermore, Juffs and his colleagues have found longer RTs in the location of subject gap in the embedded clause as in "*Who*$_i$ *did the police believe* t_i *shot the lawyer in the street*?" compared to the object gap in object extractions as in "*Who*$_i$ *did the police believe the lawyer shot* t_i *in the street*?". They have suggested that the difficulty in subject extractions was due to the processing load in the reanalysis of the matrix object trace to embedded subject trace and the relevant changes in theta role and Case (see Juffs & Harrington, 1995; and Juffs 2005 for further details).

Within the framework of Pritchett's Generalized Theta theory (1992), Juffs and his colleagues hypothesize that on encountering the wh-phrase 'who", as in "*Who did the police believe shot the lawyer in the street*?" the parser constructs a CP because the grammar contains the information that wh-phrases can occur in the clause initial Spec CP position. (The CP is the locus of the [+strong] wh-feature in English.) As soon as the wh-phrase is identified, the parser seeks a possible gap, which is the location of the trace of the moved wh-phrase. Since the subject position of the matrix clause is not quickly available, the parser posits a gap in the object position so that the wh-phrase is licensed by receiving Case and a theta role through an A-bar chain from the trace that is governed by the verb '*believe*' as in (*Who*$_i$ *did the police believe* t_i ?). However, on encountering the verb '*shot*', the parse of the gap as a matrix object is no longer viable because '*shot*' is a finite verb with two theta roles to be discharged. In order to satisfy the theta requirements of both '*believe*' and '*shot*', the alternative analysis (the first reanalysis) is to posit a CP complement of '*believe*' and establish the gap in the subject position of the embedded clause (Who$_i$ did the police believe [$_{IP}$ t$_i$ shot]?). Then, the NP '*the lawyer*' is encountered and receives the internal theta role from '*shot*', and accusative Case. Thus, all principles are satisfied. In subject extraction, a matrix object trace must be reanalyzed as a subject trace in an embedded clause, requiring both a change in theta role assigner (*believe* → *shot*), a change in the theta role (internal → external), a change in Case assignment/assigner (accusative →nominative). However, in object extraction as in "*Who did the police believe the lawyer shot in the street?",* the initial parsing is the same until the embedded verb '*believe*' (Who$_i$ did the police believe t$_i$?). Then, a lexical NP appears '*the lawyer'* in the Spec IP in the embedded clause, which simply requires reanalysis of the trace as an embedded object as in (*Who$_i$ did the police believe the lawyer shot t$_i$* ?). The type of theta role and type of Case remain the same. Therefore, they claim that processing subject extraction from a finite clause should be more difficult for all speakers than processing object extraction from a finite clause. If learners have problems with subject extractions, this should be due to the reanalysis of matrix object trace to embedded subject trace and the concomitant changes in theta role and Case.

In this study, recall that, although all three groups tended to be more accurate and faster in judging object extractions than subject extractions from finite clauses, the difference between the two wh-extraction types was not statistically significant. Nevertheless, it is still important to find out whether the locus of the subject gap in the embedded clauses triggers longer RTs than the locus of object gap, a finding that was reported in Juffs (2005). Mean RTs were presented on the Table 6a for each word of clauses involving subject (e.g., *Who did the police believe shot the editor in the street?*"), and object extractions (e.g., *Who did the police believe the lawyer shot in the street*) from finite clauses. Mean RTs for the first and the last words of the sentence were excluded from the Table 7a.

Table 7a. Word-by-word RTs for subject and object extraction from finite clauses in milliseconds (ms.)

L1 Groups		Word 2	Word 3	Word 4	Word 5	Word 6	Word 7	Word 8	Word 9	Word 10
Spanish (n=21)										
SEFF	(M)	407	384	468	563	718	530	501	436	389
	(SD)	(118)	(107)	(179)	(251)	(381)	(146)	(148)	(86)	(80)
OEFF	(M)	419	400	470	525	502	636	585	513	397
	(SD)	(144)	(135)	(145)	(187)	(160)	(237)	(213)	(167)	(66)
Turkish (n=29)										
SEFF	(M)	391	385	433	482	572	496	478	444	364
	(SD)	(145)	(117)	(218)	(234)	(317)	(130)	(227)	(197)	(71)
OEFF	(M)	349	358	396	454	443	467	484	447	401
	(SD)	(91)	(98)	(127)	(153)	(114)	(197)	(194)	(153)	(70)
English (n=30)										
SEFF	(M)	385	389	420	442	465	433	445	423	414
	(SD)	(117)	(126)	(147)	(157)	(152)	(161)	(175)	(123)	(130)
OEFF	(M)	386	385	421	443	452	453	472	431	402
	(SD)	(119)	(122)	(182)	(114)	(158)	(152)	(156)	(120)	(108)

SEFF: Subject extraction from finite clauses. OEFF: Object extraction from finite clauses

For the sake of clarity, the above examples of subject and object extractions are divided into numbered words as in Table 7b.

Table 7b. Subject and object extractions from finite clauses

Types	1	2	3	4	5	6	7	8	9	10	11
SEFF	Who	did	the	police	believe	shot	the	editor	in	the	street
OEFF	Who	did	the	police	believe	the	lawyer	shot	in	the	street

Mean RTs on Table 7a display that RTs on the matrix verb *believe* (Word 5) appear longer than those on the subject NP (Word 4) in both subject and object extractions. This suggests that all participants attempt to associate wh-filler with the matrix object trace (or subcategorizing verb) as in (Who$_i$ did the police believe t$_i$?). RTs on the embedded verb '*shot*' (Word 6) in subject extractions sharply increase, compared to the same region involving the determiner '*the*' (Word 6) in object extractions. This suggests that in subject extractions, all participants, particularly the L2 learners have more difficulty in processing the embedded finite verb '*shot*' following the main verb '*believe*' but they did not get such difficulty when they encountered the determiner '*the*" after the main verb '*believ*e' in object extractions. These results suggest that the locus of the difficulty in subject extractions is the finite verb in the embedded clause as suggested by Juffs and colleagues (Juffs and Harrington, 1995; 1996; Juffs, 2005).

These results can be clearly seen in Figure 22 and Figure 23 below, which display word-by-word RTs for subject and object extractions from finite clauses, respectively.

Figure 22.Word-by-word RTs for subject extraction from finite clauses

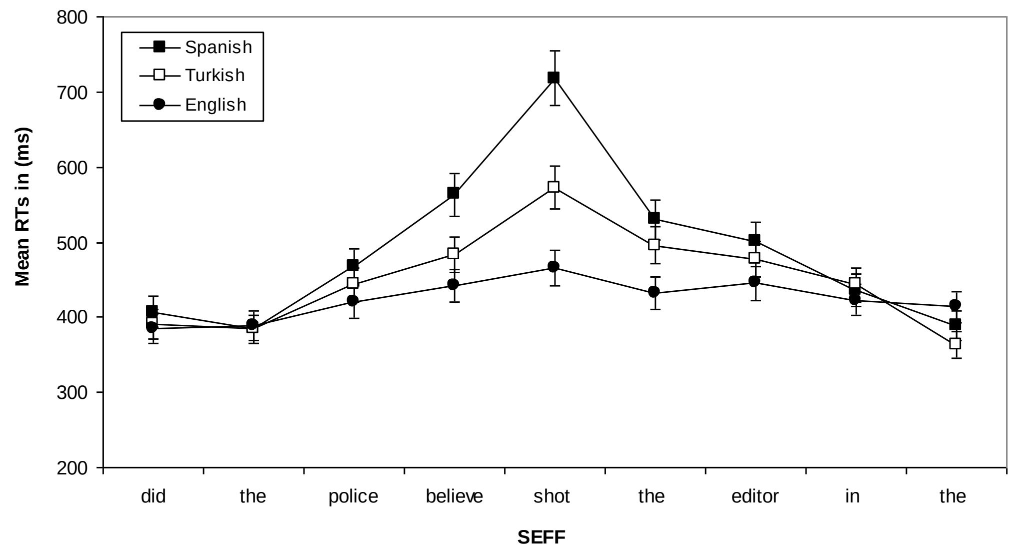
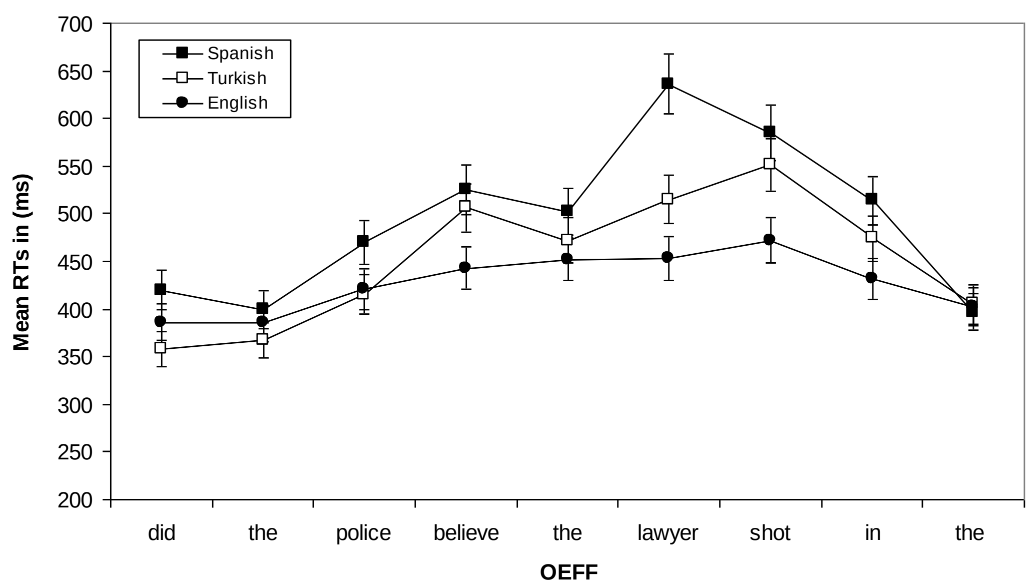


Figure23.Word-by-word RTs for object extraction from finite clauses



The spike in Figure 22 confirms longer RTs to the embedded verb *"shot*" in the subject extractions, compared to RTs to the determiner '*the*' in the object extractions in Figure 23. Although, the RT patterns were similar in three language groups, there were differences among them in terms of the reading speed.

To examine differences in language groups by type and region, a three-way repeated measures ANOVA (3x2x2) was conducted with language (English, Spanish and Turkish) as the between-subjects factor, type (subject and object extractions) and region (word 5 (*believe/believe*) and word 6 (*shot/the*)) as the repeated within-subject factors. The results of the three-way repeated ANOVA indicated a marginal effect for language ($F\ (2, 79) = 3.09$; $p=.051$), and a significant main effect for type ($F\ (1, 79) =5.37$; $p<.05$) and a significant main effect for region ($F(1, 79) =6.69$; $p<.05$), a significant interaction for type by region ($F(1, 79) =19.54$; $p<.01$), and a significant triple interaction for language by type by region ($F(2, 79)=4.14$; p>,05). However, the other interactions such as type by language group and region by language group were not statistically significant. These results show that three language groups were marginally different from one another in terms of overall RTs for the critical region 'shot' in the subject and 'the' object extractions. However, post hoc analysis revealed that only the Spanish learners were significantly slower ($p=.039$) than the English native speakers and the Turkish learners. The other two groups were not significantly different from each other.

The significant main effect for type suggests that RTs for subject extractions were significantly different from those for object extractions. Post hoc analysis of type revealed that subject extraction incurred longer RTs than object extraction. The significant main effect for region showed that overall RTs to the critical region involving the word 6 (*shot/the*) were significantly longer than RTs to the main verb (*believe*) preceding them in both types. However, the significant interaction for type by region suggests that the two types were significantly different from each other in relation to RTs to region (the word 5 and 6). Also, the triple significant interaction for language by type by region suggests that that there are significant differences among language groups by type and by region.

To examine this, a two-way repeated measure ANOVA per language group was conducted, with type and region as the repeated within-subjects factors. The ANOVA conducted on the English native speakers' mean RTs indicated only a marginal main effect for region (F (1, 30) =3.48; p=.072), suggesting that overall RTs to subject and object extraction were not significantly different but RTs to the region (the word 5 and 6) were significantly different. Post hoc analysis of regions showed overall RTs to the word 6 (*shot/the*) were significantly longer than RTs to the word 5 (*believe/believe*) (p=.72). Further analysis of region per type revealed a significant main effect for region in subject extraction, (F (1, 30) =8.81; p<.05), but not for region in object extraction (F (1, 30) =.470; p>.05). These results suggest that English native speakers are significantly slower in processing the embedded verb *shot* than the main verb *believe* in subject extractions, but they are not significantly slower in processing the determiner '*the*' following the main verb '*believe'* in object extraction.

The results of Spanish group's RT analysis indicated a significant main effect for type (F (1, 20) =11.84; p<.05), and a significant interaction between the type and region, (F (1, 20) =6.46; p<.05), but not a significant main effect for region (F (1, 20)= 2.19; p>.05). Pairwise comparison of types revealed that the Spanish speakers spent significantly longer RTs on subject extraction from finite clauses (p=.003) than object extraction from finite clauses. Further analysis of region per type indicated only a marginal main effect for region in subject extractions (F (1, 20) =4.24; p=.053), but not a significant main effect for region in object extractions. Pairwise analysis of region in subject extractions showed that, the Spanish learners spent significantly longer RTs on the embedded verb *shot* in than the main verb *believe*. This suggests that like English native speakers, the Spanish speakers had most difficulty in processing the embedded verb '*shot*' in subject extraction than the subject NP in object extraction.

Results from Turkish group's RTs indicated a significant interaction for type and region (F (1.29) =10.29; p<.01). No significant difference was observed between the two types in terms of overall RTs (p=.524). However, further analysis of region in each type indicated a significant main effect for region in subject extractions (F (1, 29) =14.71; p<.01). Like the other two groups, the Turkish group did not show a significant difference in RTs spent on Region 5 and Region 6 in object extraction (F (1.30) =1.06; p>.05). These results suggest that the Turkish learners had processing difficulty in the embedded verb *shot* following the main verb *believe* in subject extractions.

In addition to this analysis, a 3x2x2 ANOVA was conducted with the mean RTs on the main verb '*believe*' and the subject NP preceding it, to examine whether all groups initially analyzed the wh-filler '*who*' as the object of the main verb '*believe'* in both subject and object extractions. Results indicated that all participants' RTs significantly increased at the main verb *believe* in both types, suggesting that all groups initially associated the wh-filler with its subcategorizing verb or the gap at this point.

Also, to examine whether or not mean RTs significantly increased at the embedded verb *shot* in object extractions due to filler-gap integration at this point, another 3x2x2 ANOVA was conducted with mean RTs to the embedded verb *shot* and the subject noun *lawyer* preceding it in object extractions and the object noun '*editor*' and the determiner '*the*' in the subject extractions. Results indicated that only English native speakers showed significantly increased RTs on the embedded verb *shot*, where wh-filler is associated with its object gap-gap/subcategorizering verb.

To sum up, overall results from the self-paced word-by-word reading times indicated that both the English native speakers and the two L2 groups showed a similar processing profile in that they spent longer RTs on the embedded verb '*shot*' (Word 6) in the subject extractions, compared to the determiner '*the*' (Word 6) in the object extraction. This suggests that the locus of the difficulty in subject extraction is the embedded finite verb '*shot*' in subject extraction as suggested by Juffs and Harrington (1995; 1996) and Juffs (2005). Also, both the English native speakers and the two L2 groups spent significantly longer RTs on the main verb '*believe*' (Word 5), which suggests that they attempted to associate wh-filler with its object trace (or subcategorizer) at this point. The only difference among the three groups is that Turkish and English native speakers were significantly faster than Spanish speakers in processing Word 5 and Word 6 in subject and object extractions. The Spanish learners were also slower in processing subject extractions than object extractions. All in all, these results converge with the finding of Juffs and Harrington, 1995; 1996; and Juffs, 2005 in term of the locus of processing difficulty that subject extraction caused. However, they diverge from them with the finding that not only L2 learners but also the native speakers experienced the same processing difficulty at the critical region in subject extraction. More importantly, they show that both the Turkish and Spanish learners process the subject and object wh-extractions from finite clauses in the same way as the English native speakers.

5.3.2.2. Word-by-word reading times for subject and object extraction from nonfinite clauses

RT data from subject and object extractions from nonfinite clauses were checked for accuracy using the same procedure applied in the analysis of subject and object extractions from finite clauses. It was found that RTs for both object and subject extractions in each language group were normally distributed. There were a few outliers and extreme values. Only extreme values were excluded from the analyses and this caused 6.5% loss in Turkish participants' data from subject extractions; and 3.2 % loss in English native participants' data. In object extraction, there were no extreme values, which were 2000 ms beyond the mean, so all cases were included except for three Spanish participants, who failed to obtain 60% accuracy in judging wh-sentences.

Recall that previous accuracy and RT analyses above indicated that for all three groups, subject extraction from a nonfinite clause was more difficult than object extraction from a nonfinite clause. This result is in line with the finding of previous research studies (Schachter, 1989; White and Juffs, 1998; Juffs and Harrington, 1995; 1996). Juffs and Harrington (1995) demonstrated that the locus of processing difficulty in subject extraction from a nonfinite clause was the critical region involving the embedded object NP (i.e., t*he job applicant* as in "*Who does the manager expect to meet the job applicants*?" Following Pritchett's Generalized Theta Attachment (GTA) theory, they suggested that processing of a subject wh-phrase extracted from a nonfinite clause is more complex than a subject extraction from a finite clause. In the subject extraction from a nonfinite clause, the parser follows the same steps as those found in subject extractions from finite clauses until the matrix verb '*expect*' as in "*Who*$_i$ *did the manager expect* t_i?". However, upon encountering '*to*', the parser makes a reanalysis, which involves positing of a subject trace as in (*Who*$_i$ *did the manager expect [*$_{IP}$ t_i *to*]), because '*Who*' requires Case and may get it from a subject trace that receives accusative Case from the Exceptional Case Marking (ECM) verb '*expect'* in the subject position of the embedded clause.

However, as soon as '*meet'* is encountered, a second reanalysis takes place, because (*Who*$_i$ *did the manager expect to meet* t_i*?*) is a well-formed sentence. In this case, the parser will posit a PRO in Spec IP of the embedded clause and automatically project a CP to block case assignment by the ECM verb. The parser posits an object trace in the embedded clause, all theta roles are assigned, and all chains are properly licensed as in (*Who*$_i$ *does the manager*$_k$ *expect* [$_{CP}$ [$_{IP}$ PRO$_k$ *to meet* t_i*?*].

A third reanalysis takes place when the parser encounters the embedded object NP "*the job applicants*". The parser has to backtrack and reanalyze the PRO as a subject trace of the embedded clause once again as in (*Who*$_i$ *does the manager expect* [$_{IP}$ t_i *to meet the job applicants?*]).

To sum up, three reanalyses take place in parsing a subject extraction from an embedded nonfinite clause: (1) from matrix object trace to embedded subject trace; (2) from subject trace to PRO+ object trace; (3) from PRO back to embedded subject trace). Therefore, the processing of subject extraction from finite and nonfinite clauses is characterized by a reanalysis of the structural position and type of theta role and Case of variables, whereas the parsing of object extraction involves reanalysis of the structural position and theta/case assigner.

Table 8 presents mean RTs and standard deviations for each word of the subject extractions from a nonfinite clause as in (*Who does the manager expect to meet the job applicants today?*), and object extractions from a nonfinite clause as in (*Who does the manager expect to meet at work this morning?*), excluding mean RTs for the first and the last words of the sentences.

Table 8a. Word-by-word RTs for subject and object extraction from nonfinite clauses

L1 Groups		Word 2	Word 3	Word 4	Word 5	Word 6	Word 7	Word 8	Word 9	Word 10
Spanish (n=21)										
SEFNONF	(M)	425	367	490	593	437	478	538	596	465
	(SD)	(146)	(83)	(186)	(221)	(96)	(153)	(213)	(244)	(153)
OEFNONF	(M)	393	393	477	536	421	478	461	438	463
	(SD)	(110)	(110)	(181)	(235)	(76)	(143)	(126)	(100)	(129)
Turkish (n=30)										
SEFNONF	(M)	389	371	443	469	421	458	557	521	458
	(SD)	(144)	(108)	(127)	(167)	(101)	(176)	(311)	(223)	(145)
OEFNONF	(M)	372	354	399	440	421	429	437	385	405
	(SD)	(99)	(82)	(107)	(135)	(95)	(126)	(147)	(85)	(110)
English (n=31)										
SEFNONF	(M)	405	407	445	491	431	444	535	509	477
	(SD)	(126)	(141)	(160)	(192)	(113)	(166)	(186)	(167)	(123)
OEFNONF	(M)	378	383	410	443	413	419	433	434	433
	(SD)	(120)	(124)	(142)	(151)	(121)	(142)	(136)	(136)	(117)

SEFNONF: Subject extraction from a nonfinite clause. OEFNONF: Object extraction from a nonfinite clause

Again for ease of explanation, the above examples of subject and object extractions are divided into numbered words:

Table 8b. Regions for subject and object extractions from nonfinite clauses

Types	1	2	3	4	5	6	7	8	9	10	11
SEFNONF	Who	does	the	manager	expect	to	meet	the	job	applicants	today
OEFNONF	Who	does	the	manager	expect	to	meet	at	work	this	morning

Mean RTs for each word of wh-sentences with subject extraction on Table 8a show that in three language groups, RTs increased on the matrix verb '*expect*' (Word 5), suggesting that the parser initially integrated wh-filler "who' with its object gap (or subcategorizering verb).

Mean RTs decreased when the parser encountered "*to*" (Word 6), suggesting that the parser did not get surprised at this point. However, RTs started to increase when the parser encounters the embedded verb '*mee*t' (Word 7). This shows that the parser canceled the initial analysis of matrix object NP trace and posited a PRO+ an embedded object NP trace as in ((Who_i *does the* $manager_k$ *expect* [$_{CP}$ [$_{IP}$ PRO_k *to meet* t_i*?*]).

Mean RTs on the embedded determiner '*the*' (Word 8) appear higher than the embedded verb '*mee*t' (Word 7), which suggests that the parser experienced a filled gap effect at the site of the overt embedded object NP (*the job applicants*), where an embedded object NP trace is expected. At this point, the parser made another reanalysis to integrate wh-filler '*who*' with its subject trace in the embedded clauses as in (Who_i *does the manager expect* [$_{IP}$ t_i *to meet the job applicants?*]). These findings are in line with Juffs and Harrington, (1995), which suggests that the locus of the difficulty is the embedded object NP ‘*the job applicants’*, where all participants experienced a strong filled-gap effect.

To note that, relatively short RTs on ‘*to*’ do not support the claim that upon encountering ‘*to*’, the parser makes a first reanalysis of the first parse of matrix object trace and changes it with a subject trace as in (*Who_i did the manager expect [$_{IP}$ t_i to*]), but show instead that, the parser waits until it meets the embedded verb “*meet*’ as in (*Who_i did the manager expect [$_{IP}$ t_i to meet*]).

Mean RTs for each word of wh-sentences with object extractions on Table 8a, on the other hand, show only an increase at the matrix verb ‘expect’ (Word 5) where the wh-filler is associated with its matrix object gap (or subcategorizer). Unlike subject extraction, there seems no increase in RTs of three groups for Word 8.

Figure 24 and 25 present word level RTs for subject and object extractions in ms. respectively. Mean RTs for the first and the last words of these sentences were not displayed in the figures.

Figure 24. Word-by-word reading times for subject extraction from nonfinite clauses

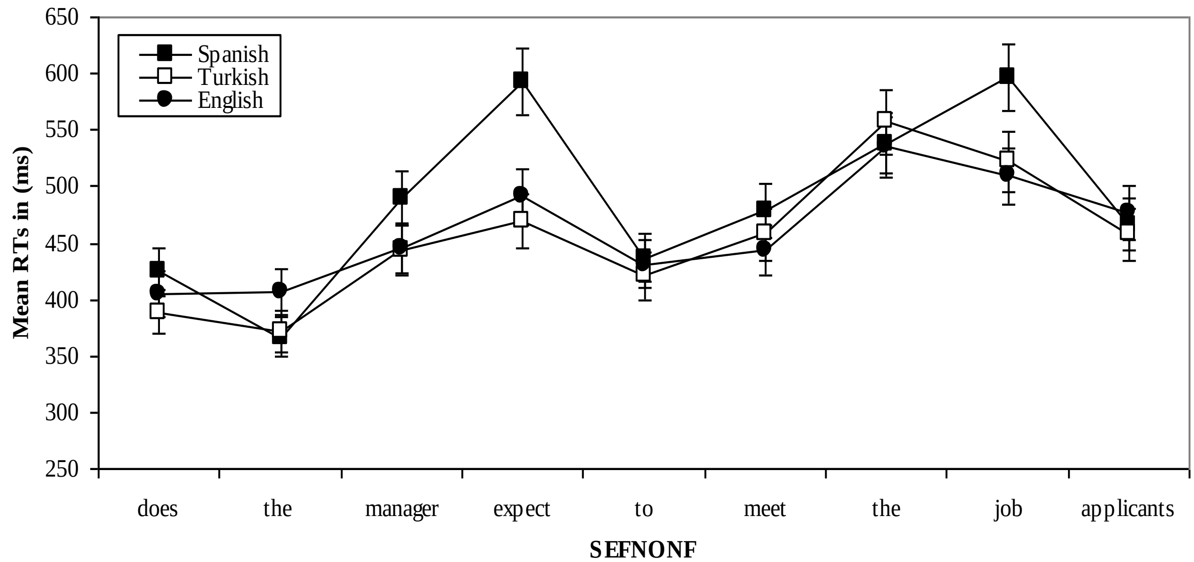
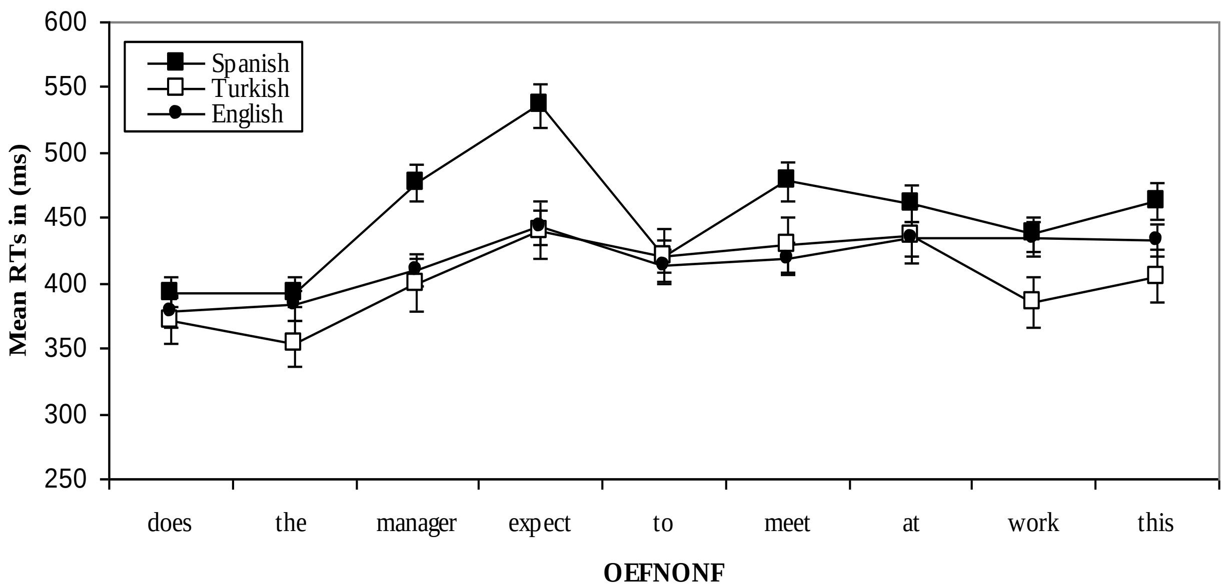


Figure 25. Word-by-word reading times for object extractions from nonfinite clauses



As can be seen in Figure 24, the spikes on the matrix verb '*expec*t', the embedded verb *'meet'* and the determiner '*the*' of the embedded object NP in subject extraction show that all participants, particularly the Spanish learners spent longer RTs at these points, which suggests there are at least two reanalyses taking place as described above during processing of subject extractions: (from matrix object to PRO+ embedded object trace; from PRO+ embedded object trace to embedded subject trace). The locus of the processing difficulty in subject extraction from a nonfinite clause appears to be the determiner 'the' in the embedded object NP, which took the longest RTs during the processing.

Figure 25 shows that in object extraction, the spike is on the main verb '*expect*' as in the initial parsing profile of subject extractions, where wh-filler is associated with its gap. However, no peaked RTs appear on the words in the embedded clause. Only Spanish speakers had increased RTs at the embedded verb '*meet*', which is also a possible gap site for the wh-filler. Although both native and L2 speakers demonstrated the same processing pattern for subject and object extractions from nonfinite clauses, they differ in reading speed. To examine whether or not the differences in mean RTs for the main verb 'expect' (Word 5), the embedded verb 'meet' (Word 7) and the determiner 'the' (Word 8) in subject extraction are statistically significant, a 3x2x2 ANOVA per region involving Word 5, Word 7, and Word 6 in two types (subject and object extraction) was conducted. The first 3x2x2 ANOVA was run on mean RTs for the main verb 'expect' (Word 5) and for the preceding noun '*manager'* (Word 4), with language (English, Spanish, and Turkish) as the between-subjects factors, type (subject and object extractions) and region (Word 4 and Word 5) as the repeated within-subject factors. The results revealed a significant main effect for type (F (1, 78) =15.52; p<.01), and a significant main effect for region (F (1, 78) =34.86; p<.01). The difference between language groups was approaching a significant level (p=.095), and the interaction between region and language was also approaching a significant level (p=.093). A post hoc analysis (Tukey, p<.05) of language groups showed that the difference in overall RTs to subject and object extractions was not statistically significant. A pairwise comparison (Bonferroni, p<.05) of type revealed that subject extraction received significantly longer RTs than object extractions in nonfinite clauses. A pairwise comparison of region indicated that RTs significantly increased on the main verb '*expect*' (Word 5), compared to the subject noun '*manager*' (Word 4) preceding it. These results suggest that all participants initially interpreted wh-filler as the antecedent of the matrix object trace and thus posited an object trace to associate it with its object trace (or subcategorizing verb).

The second 3x2x2 ANOVA was conducted for RTs on the embedded verb '*mee*t' (Word 7) and the word '*to*' (Word 6) preceding it, with language (English, Spanish, and Turkish) as the between-subjects factors, type (subject and object extractions) and region (Word 6 and Word 7) as the repeated within-subject factors. The aim was to explore whether all groups spent longer RTs at the embedded verb to reanalyze the initial analysis. Results indicated only a significant main affect for region (F (1, 79) =9.42; p<.05. A pairwise comparison of regions revealed that all participants spent significantly longer RTs on the embedded verb 'meet' than the words preceding it in subject and object extractions. This suggests that all groups reanalyzed the initial matrix object trace and posited a PRO + embedded object trace at this point in both subject and object extractions from finite clauses.

The third 3x2x2 ANOVA was conducted, with language (English, Spanish, and Turkish) as the between-subjects factors, type (subject and object extractions) and region (Word 7and Word 8) as the repeated within-subject factors. Results showed a significant main effect for type ($F(1, 78) = 30.01; p < .01$), a significant main effect for region ($F(1, 78) = 10.21; p < .01$) and a significant interaction between region and type ($F(1, 78) = 11.58; p < .05$). A pairwise comparison of types revealed that all groups spent longer RTs to Word 7 and Word 8 in subject extraction than those in object extraction. Also, a pair-wise comparison of regions showed that Word 8 incurred longer RTs than Word 7. However, the significant interaction between region and type suggests that there are differences between types in terms of RTs for Word 7 and Word 8. A further analysis of type by region revealed that the difference in RTs for the embedded verb '*meet*' (Word 7) and the following preposition '*at*' (Word 8) in object extraction was not statistically significantly ($F(1, 80) = .064; p > .05$). However, the difference in RTs for the embedded verb '*meet*' (Word 7) and the determiner '*the*' (Word 8) were statistically significant, which suggests that all groups experienced a filled-gap effect at the embedded object NP '*the job applicants'* and revised the previous embedded object trace as the embedded subject trace.

To sum up, overall results from word-by-word reading times for subject and object extractions from nonfinite clauses indicated the same patterns of reading for all three groups. Both the native and L2 speakers first attempted to associate a wh-filler with a matrix object trace (or subcategorizing verb) at the matrix verb '*expect*', then they revised this analysis when they encountered the embedded verb '*meet*' and posited an embedded object trace for the wh-filler '*who*'. However, they experienced a filled-gap effect when they encountered the overt embedded NP '*the job applicants'*, and executed a last reanalysis for embedded subject trace. These results confirm the previous findings reported by Juffs and Harrington (1995), which suggest that there is a subject and object asymmetry in wh-extractions from nonfinite clauses, and that the locus of the difficulty in subject extractions from a nonfinite clause is the location of embedded object NP, where the parser expects an object trace. These findings are also consistent with Frazier's AFS and De Vincenzi's MCP and Pritchett's GTA. The common assumption of all these accounts is that as soon as a wh-filler is identified, the parser searches for a trace to associate it with a wh-filler, establishing minimal chains. In addition, these results can be accounted for by the Direct Association Hypothesis of Pickering and Barry (1992) since all traces in subject and object extractions from finite and nonfinite clauses are adjacent to the subcategorizing verb that assign theta roles and Case to wh-fillers.

To conclude, these results show that the Turkish and Spanish L2 learners of English use the same processing strategies as the English native speaker to process subject and object extractions from nonfinite clauses.

5.3.2.3. Word-by-word reading times for object extraction from finite clauses with complementizer that.

Recall that this category included wh-sentences with object extraction from finite clauses with (*Who did the police believe that the man killed in London?*) and without the complementizer '*that*' (*Who do the police believe the lawyer shot in the street?*). RTs for the word-by-word reading of wh-sentences with object extraction from finite clauses with complementizer *that* was checked for accuracy and normal distribution. The result of this analysis revealed that mean RTs for each word of wh-sentences in this type were normally distributed. Elimination of extreme cases caused 3.2% loss in the Turkish group's data and 6.5% loss in the native speakers' data. Three Spanish participants were excluded from this analysis due to their failure to reach an accuracy level above 60% in the GJT.

Overall results reveal that the locus of the difficulty appears to be the location of Word 6 in object extraction with complementizer *that*. The same region in object extraction without *that* does not trigger longer RTs. Table 9a displays mean RTs for these categories.

Table 9a**.** Mean RTs for word-by-word reading for object extractions from finite clauses with and without complementizer *that* in ms

Groups	Word 2	Word 3	Word 4	Word 5	Word 6	Word 7	Word 8	Word 9	Word 10
Spanish (n=21)									
OEFFT (M)	417	392	465	516	478	496	493	519	522
(SD)	(156)	(109)	(160)	(204)	(142)	(227)	(128)	(169)	(149)
OEFF (M)	419	400	470	525	502	636	585	513	397
(SD)	(144)	(135)	(145)	(187)	(160)	(237)	(213)	(167)	(66)
Turkish (n=29)									
OEFFT (M)	366	356	416	462	467	456	441	459	475
(SD)	(101)	(82)	(118)	(183)	(139)	(133)	(131)	(168)	(129)
OEFF (M)	349	358	396	454	443	467	484	447	401
(SD)	(91)	(98)	(127)	(153)	(114)	(197)	(194)	(153)	(66)
English (n=30)									
OEFFT (M)	374	372	401	433	456	437	429	439	473
(SD)	(99)	(101)	(113)	(132)	(157)	(113)	(129)	(137)	(141)
OEFF (M)	386	385	421	443	452	453	472	431	402
(SD)	(119)	(122)	(182)	(114)	(158)	(152)	(156)	(120)	(108)

OEFFT: Object extraction from a finite clause with complementizer *that*. OEFF: Object extraction from a finite without complementizer *that*

Again for ease of explanation, the above examples of are divided into numbered words:

Table 9b. Object extraction from finite clause with ‘*that*’

Types	1	2	3	4	5	6	7	8	9	10	11
OEFFT	Who	did	the	police	believe	that	the	man	killed	in	London
OEFF	Who	do	the	police	believe	the	lawyer	shot	in	the	street

Mean RTs on Table 9a show that the control group and the Turkish speakers spent slightly longer RTs on Word 6 in object extraction with *that* than the same region in object extraction without *that*. This result is consistent with the assumption that the locus of difficulty in the first type of sentences is the complementizer *that.* Apart from this, all language groups had almost the same reading profile for object extractions with or without *that*. Recall that in the full sentence condition, the results of accuracy judgments and of RT analyses for two types of object extractions from finite clauses revealed that object extraction from finite clauses with ‘*that*’ elicited longer RTs and lower accuracy for all participants. This can be seen particularly in the Spanish and the native speaker groups (see Table 2 and Table 3). Also, similar findings were observed for the two types in accuracy judgments in the moving window condition (see Table 6). However, in the moving window condition, Turkish and English native speakers obtained lower accuracy and slower RTs for object extraction from finite clauses with ‘*that*’, compared to those obtained in the full sentence condition.

Figures 26 and 27below present mean RTs for the each word of the wh-sentences involving object extractions with or without *that*.

Figure 26. Word-by-word reading times for object extractions from finite clauses with complementizer *that*

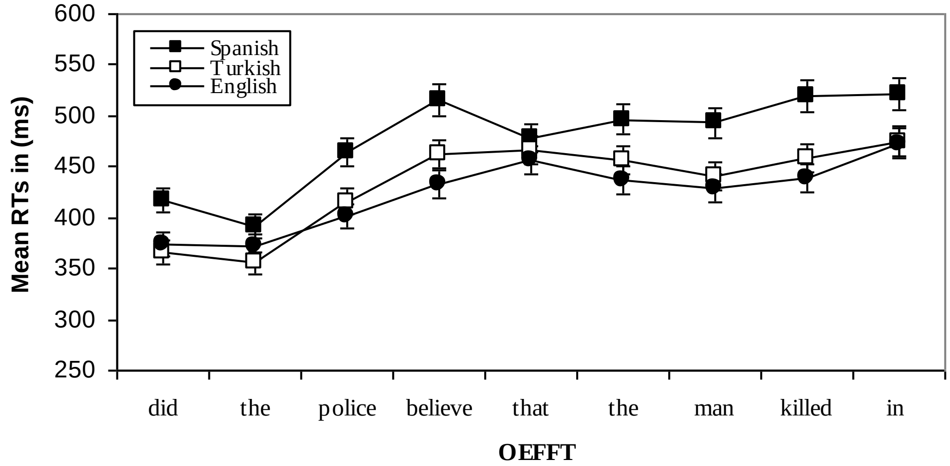
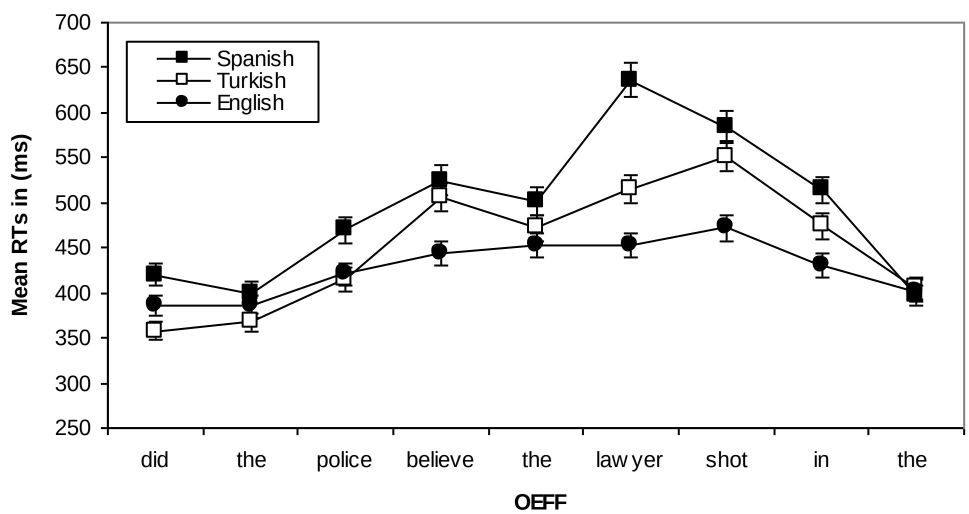


Figure 27. Word-by-word reading times for object extractions from finite clauses without complementizer *that*



As can be seen in Figures 26, Turkish and English native speakers' RTs continued to increase after the matrix verb '*believe*' until the complementizer *that* (Word 6) in object extraction from finite clauses with complementizer *that*. However, RTs started to decrease after the matrix verb '*believe*' until the determiner '*the*' (Word 6), in object extractions without *that* (see Figure 27). This suggests that English native speakers do not expect complementizer *that* after the main verb although it is grammatically correctly. This can be accounted for by the optional use of '*that*' in object extractions in English. Object extractions are more frequently used without the complementizer *that*. This account can also be true for Turkish learners' case. Another difference in RTs between these two types is seen on the embedded verbs *'shot'* and '*killed'*, where the wh-filler is integrated with its object gap (or subcategorizing verb). All participants appear slow at the embedded verb '*shot*' in object extraction without *that*, but they do not seem slow in reading the embedded verb '*killed*' in object extraction with *that*. Furthermore, RTs continued to increase after the embedded verb.

To examine the difference in RTs spent on the complementizer that (Word 6) in object extraction with *that*, and the determiner *'the'* (Word 6) in object extraction without *that* a three way ANOVA (3x2x2) was carried out with language as between subject-factors and type (object extractions with and without *that*), and region (Word 5 and Word 6). The results of this analysis indicated no significant main effect or interaction for language, type and region. This means that the difference between language groups, types and regions are not statistically significant. Another 3x2x2 ANOVA per language group was conducted with mean RTs on the embedded verbs and the preceding noun in both types. Results indicated a significant main effect for region (F (1. 30) =4.19; p=.050) in object extraction without *that* in the native speakers' group. This suggests that none of the participants slowed down at the embedded verb '*killed*' for the association of filler-gap dependency in object extraction with *that*. In the case of object extraction without *that*, only native speakers slowed down at the embedded verb '*shot'.*

To sum up, results from the RTs for word-by-wording showed that the complementizer *that* can be the locus of the processing difficulty in object extractions with the complementizer *that,* indicated by slightly longer RTs at this point. One explanation for this difficulty can be frequency effect in the use of object extractions with and without that in English. The use of complementizer *that* in object extraction is optional. Also, object extractions without that are more frequent used than those with that in English.

All in all, this section presented RT results from self paced word-by-word reading for five types of grammatical wh-extraction from finite and nonfinite clauses. Results indicated that the Turkish and the Spanish learners had a similar reading profile to that of the English native speakers in the processing of grammatical wh-extractions in L2 English. In the finite clauses, all participants had more difficulty in processing the embedded verb *'shot'* in subject extractions as in (*Who did the police believe shot the editor in the street*?), compared to the determiner '*the*' in object extractions as in (*Who did the police believe the lawyer shot in the street*?). This suggests a weak asymmetry between subject and object extractions in finite clauses.

In nonfinite clauses, in all groups the processing of subject and object extractions proceeds in a similar manner until the embedded verb '*meet*' as in (*Who does the manager expect to meet the job applicants?* vs. *Who does the manager expect to meet at work this morning?*). They had an attempt to associate the wh-filler with its object gap (or subcategorizing verb) at the main verb '*expect'*; and the first reanalysis for initial analysis of matrix object trace at the embedded verb '*meet*' and positing a PRO+ embedded object trace. Then, all participants experienced a severe filled-gap effect at the embedded object NP, where they expected an embedded object NP trace. These results confirm an asymmetry between subject and object extraction in nonfinite clauses and show that the locus of the difficult for all participants is the embedded object NP.

Also, RT results from object extractions with the complementizer *that* showed that the locus of the difficulty in this type can be the complementizer *that* which took longer to read by particularly the Turkish learners and English native speakers.

In the following section, I will discuss accuracy and RT results from five types of ungrammatical wh-extractions which violate island constraints in English.

5.3.3. Accuracy judgments on ungrammatical wh-extractions

In this analysis, accuracy responses for five types of the ungrammatical wh-extractions were checked for normal distribution and outliers in each language group. As in the full sentence condition, it was found that the data were not normally distributed. Since the number of subjects and variances among the five types were roughly equal, parametric tests were used in the analysis. Only three participants from the Spanish group, who failed to achieve 60% accuracy in judging both grammatical and ungrammatical wh-sentences, were excluded from the analysis. Table 10 provides mean accuracy scores for five types of ungrammatical wh-sentences with standard deviations.

Table 10**.** Mean accuracy scores for five ungrammatical wh-extraction types

Structure	Spanish (n=21)		Turkish (n=31)		English (n=31)	
	M	SD	M	SD	M	SD
AI (n=10)	9.52	.680	9.23	.884	9.55	.850
CNPI (n=10)	8.76	1.76	8.77	1.36	9.65	.661
RCI (n=10)	9.29	1.19	9.55	.961	9.77	.617
SI (n=10)	8.43	1.75	8.16	2.40	7.23	2.45
TT (n=10)	2.95	2.42	4.32	3.11	7.87	2.36
Total (n=50)	7.79	1.56	8.01	1.74	8.80	1.39

AI (Adjunct Island), CNPI (Complex Noun Phrase Island), RCI (Relative Clause Island), SI (Subject Island), TT (*That*-trace)

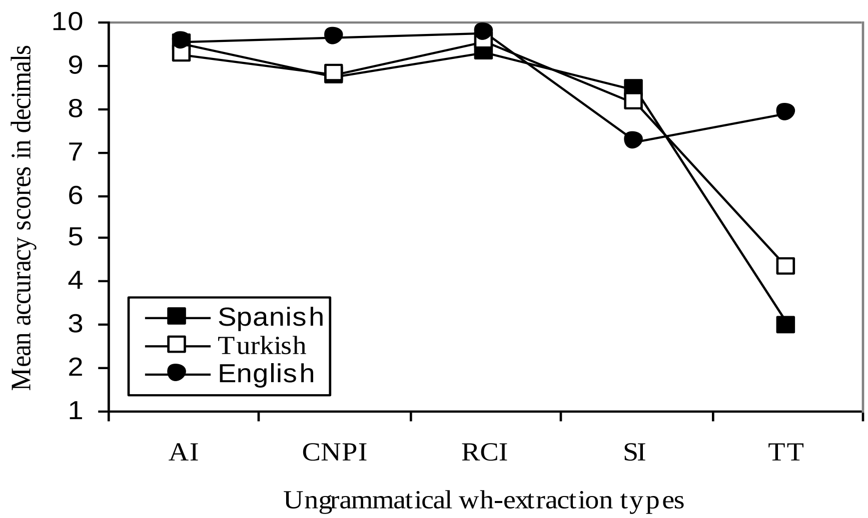
Mean accuracy scores on Table 10 indicate that all participants were slightly worse in correctly rejecting to ungrammatical wh-extractions in the moving window condition than in the full-sentence condition (compare Table 4 with Table 10). Recall that in the full-sentence condition, the English native speakers' total means were 9.13 out of 10; the Turkish learners' were 8.38, and the Spanish learners' was 7.55.

Mean accuracy scores on Table 10 also show that the two L2 groups' accuracy scores do not seem very different from those of the English native speakers, except for mean scores for the type with *that*-trace violation. The two L2 groups, particularly the Spanish speakers were less accurate than the English native speakers in correctly rejecting wh-extraction with *that*-trace violation.

In addition, it is important to note that the two L2 groups demonstrated almost the same accuracy order; while the most difficult sentence type was *that*-trace sentences, the easiest category involved either adjunct islands or relative clause islands. This pattern was very similar to what we saw in the native English speakers' data. The Spanish speakers were less accurate in correctly rejecting wh-sentences with *that*-trace, compared to other four types. Spanish speakers had lower accuracy in judging wh-sentence with *that*-trace in the moving window condition, compared to the full sentence condition. English native speakers were less accurate on wh-sentences with subject island and *that*-trace. They also had lower accuracy in rejecting these types in the moving window condition than the full sentence condition.

Figure 28 provides mean accuracy scores for five types of ungrammatical wh-extractions.

Figure 28. Mean accuracy scores for five types of ungrammatical wh-extractions



A two-way ANOVA was conducted with language as the between-subjects factor and type as the repeated within-subjects factor to examine variations in language groups and types. The results indicated an overall significant effect for language (F (2, 80) = 11.87; p<.01), a significant main effect for type (F (4, 320) =95.86; p<.01); and a significant interaction for language and type (F (8, 320) = 13. 25; p<.01), which shows that the difference between language groups was statistically significant; and that types were significantly different from each other. In pairwise comparisons of language groups, and types, it was found that the L2 groups were significantly different from the native speakers, but not different from each other. The native speakers were significantly better than the two L2 groups at judging ungrammatical wh-extractions.

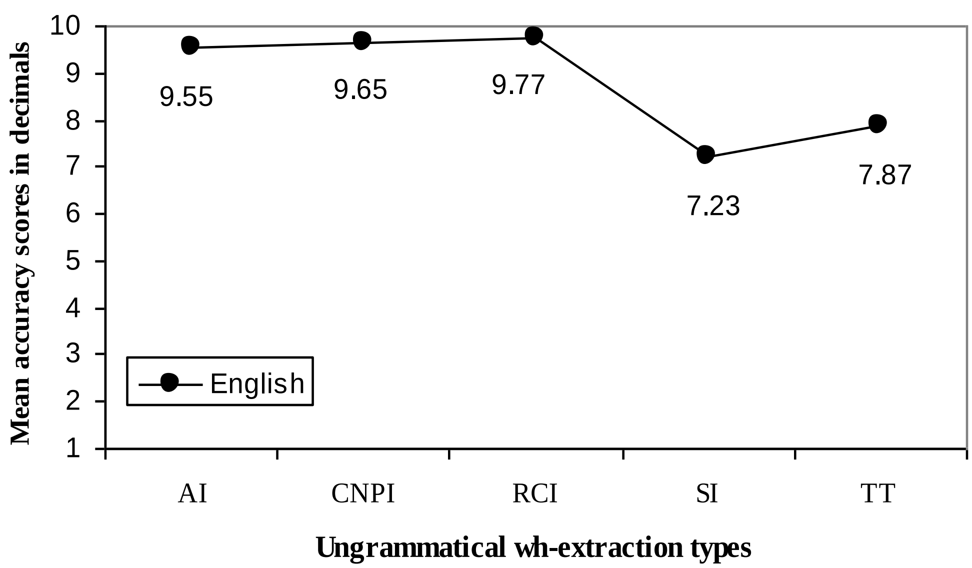
A further 3x2 ANOVA was conducted on mean accuracy scores for grammatical wh-extraction types excluding the type with *that*-trace violation where the two L2 groups were dramatically less accurate than the English native speakers. The aim was to see whether or not the difference between the two L2 learners and the English native speakers was due to the L2 learners' low accuracy in wh-extractions with *that*-trace violation. The results of the analysis did not indicate a significant effect for language ($F_{(2, 80)} = .207$; $p>.05$), showing that the L2 learners were not significantly different from either the English native speakers or from each other when *that*-trace violation was excluded from the analysis. This suggests that L2 learners are as successful as the English native speakers in correctly rejecting wh-extractions with other types of island violation.

Among the five ungrammatical types, participants were significantly least accurate at judging wh-sentences with *that*-trace violation, followed by wh-sentences with subject-island violation. They were most accurate on relative clause island (M=9.54), and adjunct island (M=9.43). Mean accuracy score (M=9.06) for the complex NP island fell between them To examine differences in language groups by type, three separate ANOVAs were conducted with type as the repeated within-subjects factor, one for each language group.

The first ANOVA carried out for the native speakers' mean accuracy scores revealed a significant main effect for type, (F (4, 120) =18.22; $p<.01$), which indicated that, English native speakers were significantly worse at both wh-extractions with subject-island and *that*-trace violations compared to the other sentence types. The difference between these two types was not statistically significant. No significant difference was observed among the other three types.

Figure 29 provides English native speakers' mean accuracy scores for five types of ungrammatical wh-extractions

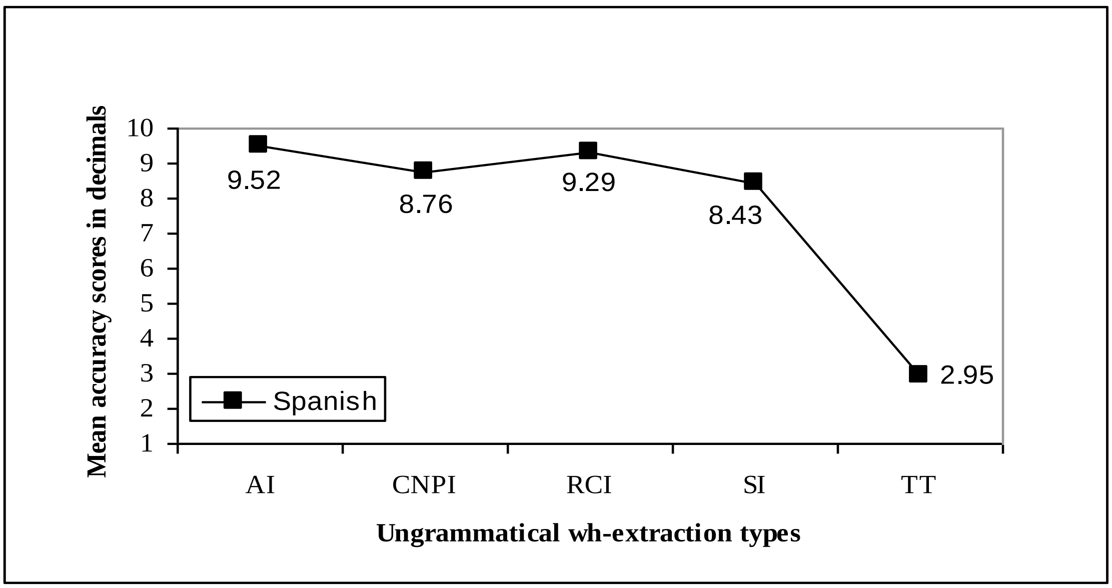
Figure 29. English native speakers' mean accuracy scores for five ungrammatical wh-extraction types



The second ANOVA carried out with mean accuracy scores of the Spanish speakers indicated a significant main effect for type (F (4, 80) = 59.42; $p<.01$), which means that among all sentence types, Spanish speakers were significantly least accurate at correctly judging wh-sentences with *that*-trace violation. Also, they had problems in correctly rejecting wh-sentences with subject-island violation. Like English native speakers, they were most accurate on adjunct and relative clause islands

Figure 30 presents Spanish speakers' mean accuracy scores for five types of ungrammatical wh-extractions.

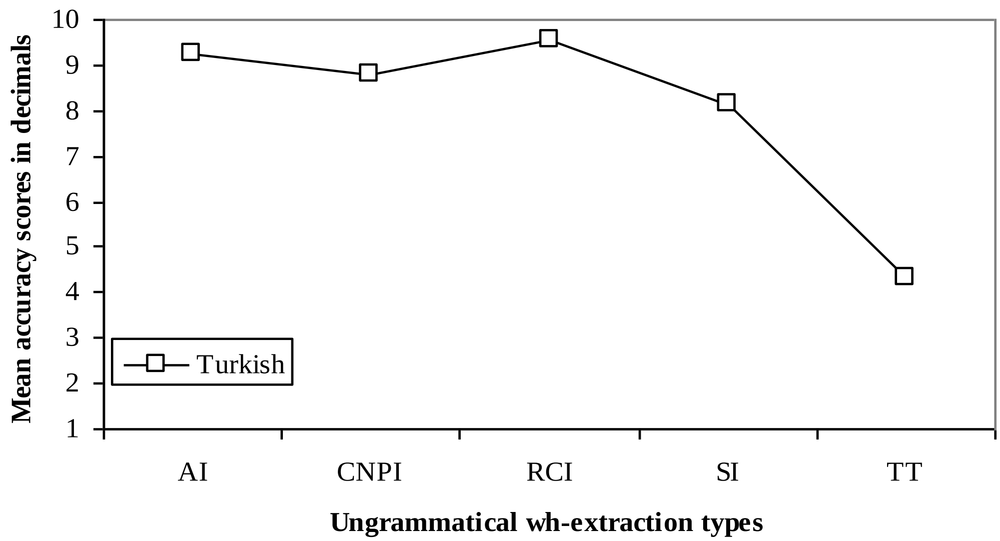
Figure 30. Spanish speakers' mean accuracy scores for five types of ungrammatical wh-extractions



The third ANOVA conducted for Turkish speakers' mean accuracy scores for five types of wh-extraction types revealed a significant main effect for type, too (F (4, 120) =37.73; p<.01). This means that, similar to the Spanish group, the Turkish group was significantly least accurate at judging wh-sentences with *that*-trace violation. They were also significantly less accurate on subject island, compared to the other three islands (AI, RCI and CNPI). In addition, they were significantly worse on the complex NP island than adjunct island.

Figure 31 illustrates Turkish speakers' mean accuracy scores for five types of ungrammatical wh-extractions.

Figure 31. Turkish speakers' mean accuracy scores for five types of wh-extractions



To summarize, these results are similar to those obtained in the full-sentence condition, and reveal that, except for the wh-sentences with *that*-trace violation, the Turkish and Spanish speakers were as successful as the native speakers at correctly rejecting ungrammatical wh-sentences with island violations in the moving window condition. These results suggest that they have knowledge of island constraints on wh-movement in English to correctly reject sentences with island violations.

As an account for the difference between the Spanish group and the other two groups, we can say that there might be L1 influence on processing of ungrammatical subject extractions from finite clauses with *that*-trace. Spanish subject NPs are allowed to move across the complementizer *that*. In Spanish, the complementizer *that* is used in many structures including subject and object extractions and it is obligatory. In Turkish, on the other hand, *that*-trace sequence is not relevant because Turkish does not have an overt complementizer similar to the one in Spanish or in English.[1] In English, the complementizer *that* is used in fewer structures and is often optional. With this background, we can say that the Spanish learners' inaccurate judgments of ungrammatical subject extractions across the complementizer *that* might be due to transfer from their L1 Spanish into the L2 English. It is important to note that we see an obvious problem with *that*-sentences across all groups. Therefore, the L1 transfer account may not be completely plausible. Both the Turkish and English native speakers display the difficulty in these sentences. Furthermore, an L1 transfer account is not tenable because Turkish does not have the complementizer in corresponding wh-sentences. The frequent use of the complementizer *that* with some verbs such as *believe*, and *think* might have mislead the reader to make incorrect decisions during online accuracy judgment. This account may also apply to the Spanish data. However, it is important to notice that the Spanish group is significantly more inaccurate than the Turkish and English participants. Therefore, we believe that the L1 transfer effect might have doubled the difficulty that the Spanish participants experienced in rejecting these sentences.

In the next section, I will discuss word-by-word reading times for five types of ungrammatical wh-extractions, beginning with the type involving *that*-trace violation.

[1] The only Turkish complementizer, which is syntactically similar to the one in English is the form –*ki* and it was borrowed from Persian. However, it is not used frequently.

5.3.4. Word-by-word reading times for the ungrammatical wh-extractions with island violations

For the word-by word RT analysis of correct responses to five ungrammatical wh-extractions, each participant's RT to each word of 10 sentences in each type were screened for normal distribution and outliers. The RTs which are ±2.5 *SD* beyond the means were excluded from the data of each participant. Then, each participant's mean RT for each word in each type was calculated. Later, mean RT data for 11 words in each type were checked for normal distribution and outliers. It was found that the RT data were normally distributed in each language group. Only extreme values were deleted from the analysis. Also, three participants from the Spanish group, who did not achieve an overall accuracy above 60%, were not included in further analyses. In the following section, I will first discuss RTs for wh-sentences with *that*-trace. This will be followed by the RT analysis for wh-sentences with subject island, relative clause island, adjunct island and complex noun phrase island.

5.3.4.1. Word-by-word reading times for the ungrammatical wh-sentences with that-trace violation

For the analysis of RTs spent on 11 words of wh-sentences with *that*-trace violations, the data were screened for normal distribution, and outliers. It was found that data was normally distributed for each word in each language group. There were five Spanish, and three Turkish participants, who failed to judge correctly all of the sentences with *that*-trace violations. Therefore, they were accepted as missing data, which caused 19% loss in the Spanish data, and 12.9% loss in the Turkish data and 3.2% in the English data. Furthermore, the participants who failed to achieve 60% overall accuracy were also removed from the analysis, which led to two more Spanish participants to be excluded from the analysis.

Recall that, in English unlike object extractions, subjects are not allowed to be extracted across overt complementizer as in "**Who do the police believe that attacked the man last night*?", because this violates the Empty Category Principle (ECP), which states that traces must be properly governed. Subject trace (*t*) in the [Spec, IP] position is not properly governed due to complementizer *that*, which prevents the trace in [Spec CP]. This was termed as *that*-trace *effect* by Chomsky and Lasnik (1977). In order to move a subject from the subject position of an embedded clause, the complementizer must be absent.

In this context, the parser should start searching a gap (trace) after reading '*Who*'. If the parser expects an object gap after the main verb '*believe*', which is the first and most plausible location for the construction of wh-filler-gap formation, longer RTs should be found after the word '*believe*'. On encountering the complementizer in this position, the parser should revise this initial analysis (since verb '*believe*' can take either a DP or CP complement, this revision can be very quick). If the parser is sensitive to the *that*-trace constraint, it should not expect a subject trace after the complementizer for the wh-filler '*who*' to form wh-dependency at this point. Therefore, having noted the implausibility of wh-filler as the subject of the embedded verb '*attacked*', the parser should get surprised by finding a finite verb requiring an embedded subject trace. Then, it should make a reanalysis at '*attacked*' to make sure that the wh-extraction stops being grammatical at this point, which will be reflected longer RTs to the embedded verb ‘*attacked*’. If parser is not sensitive to the *that*-trace constraint, it should expect a subject trace after the complementizer and should not get surprised by encountering the embedded verb ‘*attacked*’ following the complementizer, which means the parser will not spend longer RTs for ‘*attacked*’.

Table 11 shows RTs for each word of the ungrammatical subject extractions from finite clauses with *that*-trace, excluding mean RTs for the first and the last words of the sentences.

Table 11a. Mean RTs for subject extractions from finite clauses with *that-trace*

L1Groups	Word 2	Word 3	Word 4	Word 5	Word 6	Word 7	Word 8	Word 9	Word 10
Spanish (n= 16)									
TT (M)	405	376	435	593	508	699	669	464	351
(SD)	(113)	(89)	(121)	(265)	(237)	(409)	(515)	(239)	(101)
Turkish (n=26)									
TT (M)	376	363	427	431	462	651	527	501	417
(SD)	(124)	(101)	(204)	210)	(155)	(345)	(232)	(311)	(127)
English (n=30)									
TT (M)	395	398	421	445	446	601	495	394	413
(SD)	(137)	(130)	(170)	(214)	(153)	(243)	(187)	(125)	(152)

TT: Subject extraction with that-trace

The regions in this ungrammatical sentence type are given below:

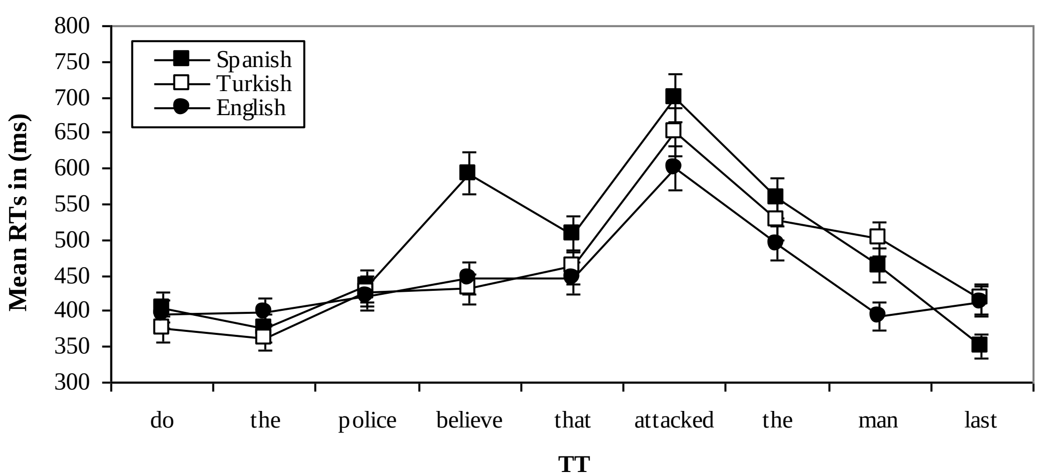
Table 11b**.** Regions in ungrammatical wh-extractions with *that*-trace violations

1	2	3	4	5	6	7	8	9	10	11
*Who	do	the	police	believe	that	attacked	the	man	last	night

Mean RTs on Table 11a show that among the three groups, only the Spanish speakers showed longer RTs for the main verb ‘*believe*’ (Word 6). This might be due to the fact that the native English and Turkish speakers might be faster than Spanish speakers in wh-filler-gap formation and reanalyzing it. Table 10a also shows that RTs for the embedded verb '*attacked*' (Word 7) were longer in all three groups, in particular the Spanish group. This result shows that both native and L2 speakers were surprised by finding a finite verb after the complementizer (i.e., they experienced a filled gap effect upon finding a finite verb at this region) and revised the initial analysis to be sure that the initial analysis of wh-filler '*who*' was correct.

Figure 32 presents mean RTs for each word of the ungrammatical subject extraction from finite clauses with *that*-trace, excluding mean RTs for the first and the last words of the sentences.

Figure 32. Word-by-word RTs for the ungrammatical subject extractions from finite clauses with *that*-trace



As can be seen in Figure 32, all three groups had similar reading pattern for the ungrammatical subject extractions with *that*-trace; all of them had the highest peak at the embedded verb '*attacked*', which is the locus of the ungrammaticality in this type. Unlike other groups, the Spanish group seems to spend longer RTs at the main verb '*believe*'. These results also show that although three groups had similar reading pattern, they were different in terms of reading speed. Recall that in the full sentence condition, wh-sentences with *that*-trace violation also triggered more errors and longer RTs in the two L2 groups, particularly in the Spanish group. Furthermore, similar mean RTs were found in the moving window condition: Spanish (M=2.95), Turkish (M=4.32), and English native speakers (M=7.23).

To examine the differences in RTs for the main verb '*believe*', a 3x2 ANOVA was conducted with language (English, Turkish, Spanish) as the between-subject factor and region (Words 4, '*police*' and Word 5, '*believe*') as the within-subject factor. The results of this analysis indicated a significant main effect for region ($F (1, 73) = 9.64$; $p<.05$) and a significant interaction between language and region ($F (2, 73) = 5.02$; $p<.05$). However, no significant difference was observed among the three language groups in overall RTs to Word 4 and Word 5 in this type, suggesting that three groups were not significantly different in overall speed at this region. A pairwise analysis of regions showed that '*believe*' (Word 5) incurred longer RTs than '*police*' (Word 4). However, significant interaction does not support this finding in three groups. To examine whether '*believe*' (Word 5) took longer RTs in three language groups, an ANOVA per group was carried out with region (Word 4 and Word 5) as the within-subject factor. Results revealed a significant main effect for region only in the Spanish group ($F (1, 16) = 7.21$; $p<.05$), which means that only the Spanish group spent significantly longer RTs on the main verb '*believe*' where wh-filler '*who*' is assumed to be initially associated with its matrix object trace (or the subcategorizer). The other groups were not as slow as the Spanish group in the formation of wh-dependencies at this point.

More importantly, to examine the differences in RTs obtained on the embedded verb '*attacked*' among the three language groups, a 3x2 repeated measure ANOVA was conducted with language as the between subject-factors and region ('*that*' (Words 6) and '*attacked*' (Word 7)) as the within subject-factors. The results of this analysis indicated a significant main effect for only region (F (1, 71) =24.97; p<.01), but not a significant effect for language or a significant interaction for language by region. Pairwise A pairwise comparison of region revealed that '*attacked*' incurred significantly longer RTs than '*that*' (p=000). These findings suggest that language groups were not significantly different from each other in total RTs for '*that*' and '*attacked*' and that total RTs for the embedded verb '*attacked*' were significantly higher than those for '*that*'.

Further analysis of each language group revealed that unlike Turkish and the native English speakers, the Spanish speakers did not show a significant difference in RTs for region (F (1, 16) =2.86; p>.05). This finding shows that the Spanish speakers were not as surprised as the native English and Turkish speakers at encountering an embedded finite verb after the complementizer *that*. This suggests that the Spanish group is not as sensitive to the *that*-trace constraint in L2 English as the English native speakers and the Turkish group. Thus, this finding confirms a local negative transfer from L1 Spanish (where complementizer *that* obligatorily exists in subject extractions from embedded clauses) to L2 English. This finding also accounts for the Spanish group's inaccurate judgments of ungrammatical wh-extractions with *that*-trace violation in the full sentence and in the self-paced word-by-word reading condition. We assume that most of the Spanish participants incorrectly accepted ungrammatical wh-sentences with *that*-trace violation under the influence of the L1 Spanish knowledge which dictates that these structures are grammatically correct in L1 Spanish.

To sum up, results from RTs for the ungrammatical wh-sentences with *that*-trace violation indicated that the locus of the difficulty is the embedded verb (here, *attacked*) following the complementizer *that*. All three groups had processing difficulty at this point, but only the native English and Turkish speakers showed significant increase at '*attacked*', suggesting that like the native English speakers, Turkish speakers are sensitive to *that*-trace constraint in the L2 English and therefore they did not expect a finite verb after '*that*' requiring a subject trace at the embedded clause. The Spanish speakers did not show significantly longer RTs at this region, suggesting that they were expecting a finite verb after '*that*' as in their L1.

In the next section I will discuss word-by-word RTs for ungrammatical wh-sentences with subject island violations.

5.3.4.2. Word-by-word reading times for ungrammatical wh-sentences with subject-island violation

As in the previous RT analysis, for the analysis of the word-by-word reading times for the ungrammatical wh-sentences with subject island violation, first, the mean RT obtained for each of the 10 sentences was calculated for each participant, excluding the RTs beyond ±2.5 standard deviations from each participant's mean score. Then, the mean RTs for each word were screened for outliers and normal distribution. The RT data were normally distributed for each word in each language group, but the mean RTs beyond 2000 (ms) were also excluded from the analysis. Only the three Spanish participants were excluded from the analysis due to low accuracy (below 60%) in the OGJT.

In English, wh-words cannot move out of a subject DP as in "*Who does the teacher believe a story by amuses the children*?" because subject DPs are islands, and therefore they do not allow an element to be moved out of the island. When an element is extracted out of subject island, it violates restrictions, or barriers (i.e., the Subjacency Principle, the Minimal Link Condition or Relativized Minimality Principle) on movement as in: "*[$_{CP}$ [Who$_i$ does [$_{TP}$ the teacher believe [$_{CPi}$ [$_{TP}$ [$_{DP}$ a story by t$_i$] amuses the children]]]]]?". The wh phrase '*Who'* jumps out of the DP to the matrix Spec CP in one long step at a time, crossing two bounding nodes (i.e., DP, TP)

In this context, the parser starts looking for a gap upon encountering the wh-filler, '*who*'. The first possible object gap site for wh-filler takes place after verb '*believe*'. To form wh-dependency, the parser should spend longer RTs for '*believe*'. This analysis will be revised as soon as the subject NP is encountered. If the parser expects a subject gap after '*by*', (here, the embedded verb '*amuses*') in the subject island for the wh-filler, or attempts to associate wh-filler with its subcategorizer '*amuses*' via a subject gap, it should spend longer RTs on the verb '*amuses*'. However, if the parser is sensitive to the island constraint information, it will not posit a subject gap inside the subject island, or associate the wh-filler with verb '*amuses*' and hence the parser will read the subject island as fast as the other parts of the sentence.

Table 12a presents mean RTs for each word of the ungrammatical sentences with subject island, excluding mean RTs for the first and the last words of these sentences.

Table 12a. Mean RTs for the ungrammatical *wh*-sentences with subject island violation

Nations	Word 2	Word 3	Word 4	Word 5	Word 6	Word 7	Word 8	Word 9	Word 10
Spanish (n=21)									
SI (M)	411	393	477	532	492	539	479	529	465
(SD)	(153)	(119)	(161)	(232)	(127)	(158)	(141)	(182)	(137)
Turkish (n=29)									
SI (M)	364	381	442	480	518	476	447	528	437
(SD)	(89)	(109)	(233)	(169)	(204)	(122)	(218)	(230)	(141)
English (n=30)									
SI (M)	397	386	476	480	471	466	470	535	523
(SD)	(125)	(118)	(204)	(203)	(142)	(155)	(221)	(252)	(227)

SI: Subject Island

The division of a subject island sentence is given below:

Table 12b. Regions for Subject Island

1	2	3	4	5	6	7	8	9	10	11
*Who	does	the	teacher	believe	a	story	by	amuses	the	children

Mean RTs on Table 12a, show an increase for the main verb '*believe*' (Word 5) in three groups. This suggests that all three groups posited an object gap for the wh-filler '*who*' and associated the wh-filler with its object gap (or subcategorizing verb). However, RTs to the determiner '*a*' (Word 6) following '*believe*' decreased in the native English group and the Spanish group. This suggests that they did not experience a filled gap effect at the determiner '*a*' where they expected an object gap (trace). Also, RTs for the embedded verb '*amuses*' (Word 9) incurred longer RTs in three groups. This shows that all three groups expected a subject gap in the subject island (or attempted to illicitly associate wh-filler '*who*' with the verb '*amuses*') and suggests that the parser is not sensitive subject-island constraint during the initial stage of processing. This finding can account for the low accuracy of groups in rejecting ungrammatical wh-extractions with subject island violation.

Figure 33 shows mean RTs for each word of the ungrammatical wh-sentences with subject island, excluding the first and the last words of the sentences.

Figure 33. Word-by-word reading for the ungrammatical *wh*-sentence with SI violation

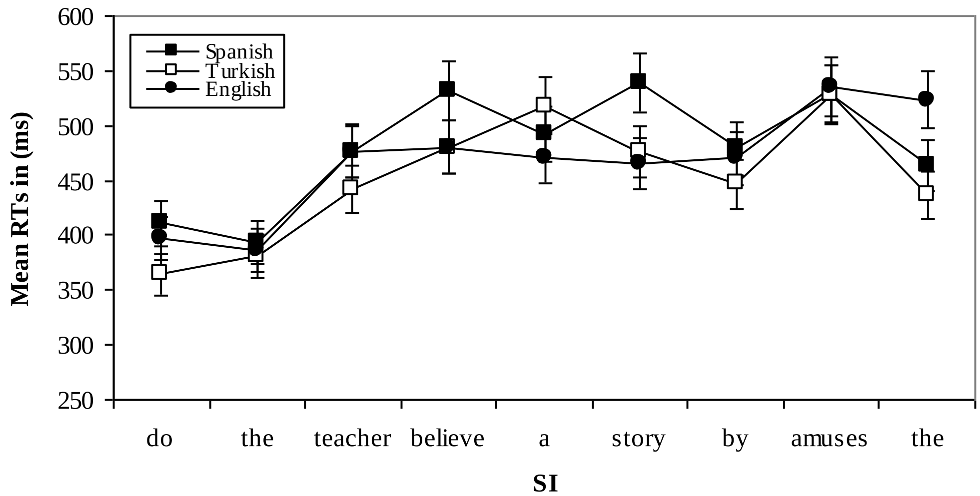


Figure 33 shows that the Spanish speakers had a higher spike on the main verb '*believe*' than Turkish and the native English speakers. Then, they had another spike on the word '*story*'. Turkish speakers, on the other hand, had the second spike on the word "*a*". Unlike the L2 groups, the native speakers did not have any spike until the embedded verb '*amuses*'. Indeed, all three groups had a spike on the embedded verb 'amuses', showing that they attempted to form wh-filler-gap (filler-verb) dependency. Although the native and L2 speakers demonstrated the same reading pattern for the ungrammatical subject island, they were different in reading speed.

To examine the differences in RTs among language groups, a two-way ANOVA per region was conducted with language as between subject-factors (Spanish, Turkish and English) and region as within subject-factors. The first ANOVA conducted with mean RTs for the main verb ' *believe*' (Word 5) did not show any significant effect for region or for language, which means that the difference in RTs at '*believe*' was not statistically significant. Similar results were obtained for the RTs to the determiner '*a*' (Word 6) which shows that none of the groups experienced a filled gap effect upon reading '*a*' after '*believe*'. However, the results of the third ANOVA conducted with RTs on the embedded verb '*amuses*' (Word 9) and the preceding preposition '*by*' (Word 8) revealed a significant main effect for region ($F(1, 80) = 13.30$; $p < .05$), but not a significant effect for language and a significant interaction for language and region. These findings suggest that the difference between the words 8 and 9 was statistically significant: the word 9 (*amuses*) took significantly longer RTs than the word 8 (*by*) in total. However, language groups were not significantly different from each other in terms of RTs for these words.

Further analysis of RTs for the words 8 and 9 in each language group indicated a significant main effect for region ($F(1, 30) = 9.74$; $p < .01$) in RTs of the native English speakers. A pairwise comparison of region showed that RTs for '*amuses*' (Word 9) were significantly longer than those for '*by*' (Word 8). This suggests that the native speakers attempted to integrate wh-filler '*who*' with its illicit subject gap inside the subject island. This result supports the view that the parser is not sensitive to subject island constraint for subject islands in English.

Results from Spanish speakers' RTs for '*by*' (Word 8) and '*amuses*' (Word 9) revealed no significant main effect for region ($F(1, 21) = 1.01$; $p > .05$), which means that the difference in RTs for the words 8 and 9 did not reach a significant level in the Spanish group. The results of RT analysis for the Words 8 and 9 by Turkish speakers revealed a significant main effect for region ($F (1, 30) = 7.88$; $p<.01$). Like English native speakers, Turkish speakers spent significantly longer RTs for than '*by*' (Word 8). This shows that like the native speakers, Turkish speakers attempted to form filler-gap dependency inside the subject island. This means that like English speakers, Turkish speakers are not sensitive to subject island constraint in wh-extractions.

To sum up, the word-by-word RTs on wh-sentences with subject island indicated that all three groups had the same reading pattern in processing wh-extractions with subject island violation. Both the native and nonnative speakers (though Spanish speakers' RTs could not reach a statistically significant level) appeared not sensitive to subject island constraint as reflected in longer RTs at the embedded verb '*amuses*', where they attempted to form filler-gap (or filler-verb) dependency. To note, this data do not provide a control (non-island) condition for the ungrammatical *wh*-sentences with relative clause-island constraints, therefore, these results should be interpreted cautiously.

In the following section, I will discuss word-by-word RTs for the ungrammatical *wh*-sentences with relative clause–island violation.

5.3.4.3. Word-by-word reading times for ungrammatical wh-sentences with relative clause (RC) island violation

This category involves ungrammatical wh-sentences with relative clause island as in "*[*Who$_i$ does [Jane trust the man [who hires t_i for the company*?]". In English, moving a wh-phrase outside the relative clause island is ungrammatical because relative clauses are islands and therefore, islands do not allow movement due to certain restrictions or barriers as mentioned above and previously discussed (see Chapter 3).

As in the ungrammatical wh-sentences given above, there are two potential gap sites for the wh-filler '*who*' in the wh-sentence with RC island in the above example: (1) a licit object gap after the main verb '*trust*' as in "*Who$_i$ does Jane trust t_i* ..), and (2) an illicit object gap after the embedded verb "hires" "*(*Who$_i$ does Jane trust the man who hires t_i for the company*?]". If the parser first associates the wh-filler with the object gap after "*trust*", a filled gap effect should be found at the NP 'the *man*'. Then, if it associates the wh-filler with the embedded verb '*hires*' via an object gap inside the RC island, it should spend longer RTs at the embedded verb '*hires*', showing that they associate the wh-filler with its object gap (or subcategorizing verb) inside the RC island. If the parser is sensitive to the RC island constraint, then it will not expect an object gap after '*hires*' inside the relative clause island or attempt to link wh-filler with its subcategorizer, hence it should read this point as quickly as the other words in the sentence.

For the word-by-word analysis of the ungrammatical wh-sentences with relative clause-island violation, mean RTs for each word by each participant were calculated, excluding RTs beyond ±2.5 standard deviations beyond the mean RTs of each participant. Then, RT data were checked for normal distribution and outliers. The data had normal distribution for each word in each language group. As in the previous analyses, only extreme values and participants were excluded from the analysis, which caused 3.2% loss in Turkish speakers' data; and 3.2 % loss in English native speakers' data. Mean RTs for each word of the ungrammatical wh-sentences with relative clause-island violation are presented in Table 13a, excluding the first and the last words of the sentences. as in *"*Who does Jane trust the man who hires for the company*?" .

Table 13a. Mean RTs for the ungrammatical *wh*-sentences with RCI violation

Groups	Word 2	Word 3	Word 4	Word 5	Word 6	Word 7	Word 8	Word 9	Word 10
Spanish (n=20)									
RCI (M)	400	438	558	566	565	501	424	374	357
(SD)	(133)	(150)	(292)	(203)	(169)	(197)	(99)	(80)	(80)
Turkish (n=29)									
RCI (M)	360	393	476	535	494	467	422	326	302
(SD)	(90)	(115)	(180)	(202)	(204)	(265)	(166)	(84)	(76)
English (n=30)									
RCI (M)	386	414	468	520	454	426	397	376	368
(SD)	(123)	(155)	(195)	(182)	(159)	(174)	(155)	(166)	(138)

RCI: Relative Clause Island

The division of a relevant sentence is given below:

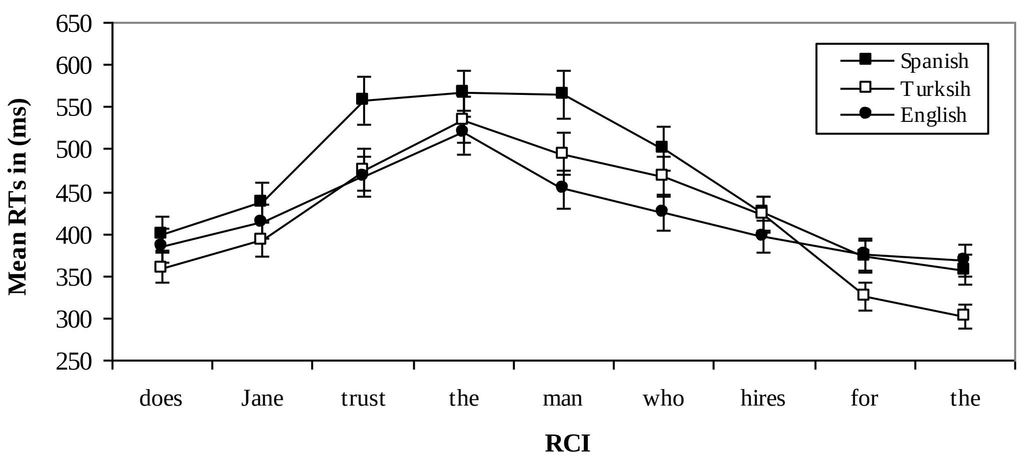
Table 13b. Regions for relative clause island

1	2	3	4	5	6	7	8	9	10	11
*Who	does	Jane	trust	the	man	who	hires	for	the	company

Mean RTs on Table 13a show that all three groups spent longer RTs for the main verb '*trust*' (Word 4) , where they associated the wh-filler "*wh*o" with the matrix object gap. Also, they had longer RTs for the determiner '*the*' (Word 5), "the" (the first word of object NP), which shows that they were surprised by encountering an overt object NP, where they expected an object gap for the wh-filler '*Who*'. There is no increase in RTs for the embedded verb '*hires*' (Word 8), which was the illicit object gap site for the wh-filler "*who*". This shows that the parser in three groups did not expect an object gap inside the relative clause island, which suggests that the L2 groups and the native speakers are sensitive to RC island constraint.

Figure 34 presents mean RTs for each word of the ungrammatical sentences with RC island.

Figure 34. Word-by-word RTs for the ungrammatical wh-sentences with RCI violation
.



As can be seen in Figure 34, both the native and L2 speakers had a peak on the determiner '*the*' (Word 5) (the first word of the embedded object NP), where they were assumed to experience a filled gap effect. This confirms the assumption that the parser posited an object gap for the wh-filler and associated it with its subcategorizing verb '*trust*', but when it encountered the overt object NP '*the*' (man), the parser got surprised and revised the initial object-gap analysis. RTs started to increase at the main verb '*trust*', where the wh-filler was linked with the object gap and peaked at the word '*the*'. Then, they decreased starting from '*man*', which confirms that they were not expecting an NP at this place. In the Spanish group, a decrease in RTs appears to start after reading the whole object NP, '*the man*'. RTs for the words in the rest of the sentence did not show any spikes, which suggests the participants did not posit an illicit gap inside the RC island.

A two-way ANOVA conducted on the RTs for the main verb '*trust*' (Word 4), and the following two words '*the*'(Word 5) in the sample sentence, with language as the between-subjects factor and region, as the repeated within-subjects factor. The results indicated only a significant effect for region ($F (1, 78) =5.22; p<.05$), which means that RTs for '*trust'* were significantly different to RTs for '*the*' in three groups. A pairwise comparison of regions revealed that '*the*' incurred significantly longer RTs than '*trust'*. This confirms the presence of a filled-gap effect experienced by both the native English speakers and the two L2 groups at the overt object NP in the sentence.

To look at the differences in RTs for the embedded verb '*hires*' (Word 8), by language, a two-way ANOVA was conducted with language as between subject factors and region (Words 7, 8 and 9) as within subject factors. The results showed that RTs kept decreasing from '*who*' (Word 7) to '*for*' and that there was no significant difference in RTs for language. This result suggests that both the native and L2 speakers are sensitive to the RC island constraint. Neither the native and nonnative speakers expect a gap inside the island for the wh-filler. However, recall that, as in the case of the ungrammatical wh-sentences with subject-island violation, this data do not provide a control (non-island) condition for the ungrammatical *wh*-sentences with relative clause-island constraints, therefore, these results should be interpreted with caution.

To summarize, results form RTs for the ungrammatical wh-extractions from inside the relative clause island revealed that all three groups had the same reading pattern. They did not expect an object gap inside the relative clause island after the embedded verb *'hires'.* This result suggests that the parser is sensitive to the RC island constraint.

In the next section, I will discuss word-by-word RTs for the ungrammatical wh-sentences with adjunct island violation.

5.3.4.4. Word-by-word reading times for ungrammatical wh-sentences with adjunct island (AI) violation

This category included wh-sentences with adjunct island such as "**What did Ann eat her dinner after she watched on TV?". Here,* moving a wh-phrase outside the adjunct island produces ungrammatical wh-sentences because adjuncts are also islands for movement. As in the previous ungrammatical wh-sentences with island constraints, we assume that after reading wh-phrase "*who*", the parser should initially integrate the wh-filler with the main verb with '*eat'* via positing an object gap, which will be reflected with long RTs at this point. Then, the parser will experience a filled-gap effect as soon as it encounters '*her*', where it expects an object gap, which will be indicated longer RTs at ‘*her*’. If the parser also integrates the wh-filler ‘*who*’ with the embedded verb ‘*watched*’ via an object gap inside the adjunct island, then longer RTs should be found at ‘*watched*’.

For the word-by-word analysis of the ungrammatical wh-sentences with adjunct island violation, only RTs to seven sentences out of 10 were included. The reason for excluding three sentences from the RT analysis was that they had a two-word NP preceding the main verb, where the gap site was expected as in **Who did* ***the police*** *arrest Sam because he attacked last night?,* whereas in the other seven sentences, there was a one-word NP as in (**What did Ann eat her dinner after she watched on TV?*). The former had 8 words before the illicit gap site after the embedded verb “attacked” inside the adjunct island, whereas the latter had 7 words before the same region. This difference is important in measuring RTs spent on this region. There should be an equal number of words between wh-fillers and the possible gap sites under investigation in the experimental sentences.

Mean RTs for each word of the 7 sentences by each participant were calculated, excluding RTs beyond ±2.5 standard deviations from each participant’s mean RTs. Then, RT data were checked for normal distribution and for outliers. It was found that the RT data were normally distributed for each word of the wh-sentences in each language group. As in the previous analyses, only extreme values and participants who did not achieve an overall accuracy over 60% were excluded from the analysis, which caused 4.8% loss in Spanish speakers’ data; 3.2% loss in Turkish speakers’ data. Mean RTs for each word of the ungrammatical wh-sentences with adjunct island violation are presented in Table 14a, beginning with the second word “*does*” and ending with tenth word “*the*” as in the example (**What did Ann eat her dinner after she watched on TV?*).

Table 14a. Mean RTs for the ungrammatical *wh*-sentences with AI violation

Groups	Word 2	Word 3	Word 4	Word 5	Word 6	Word 7	Word 8	Word 9	Word 10
Spanish (n=20)									
AI (M)	392	421	517	587	518	556	472	387	406
(SD)	(125)	(144)	(210)	(201)	(148)	(178)	(185)	(86)	(158)
Turkish (n=29)									
AI (M)	358	355	435	515	474	467	404	347	338
(SD)	(88)	(79)	(126)	(221)	(112)	(160)	(127)	(106)	(103)
English (n=30)									
AI (M)	393	429	440	502	467	495	425	363	373
(SD)	(115)	(153)	(148)	(208)	(145)	(234)	(149)	(114)	(125)

AI: Adjunct Island

The individual regions in an adjunct island sentence are given below:

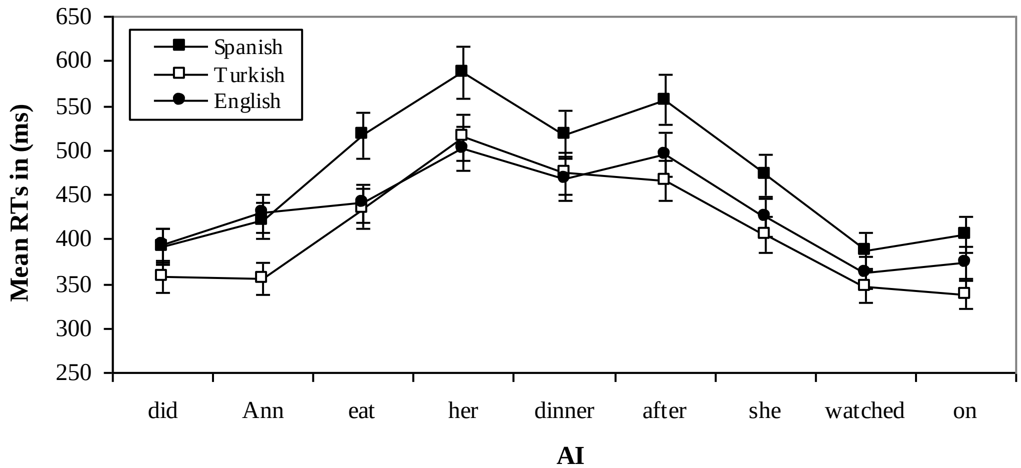
Table 14b. Adjunct island

<u>1</u>	2	3	4	5	6	7	8	9	10	11
*What	did	Ann	eat	her	dinner	after	she	watched	on	TV

Mean RTs on Table 14a show that all three groups spent longer RTs for the main verb 'eat' and the following word 'her' (Word 5), were high in three groups. This suggests that all participants linked the wh-filler '*what*' with the main verb '*eat*' via an object gap, reflected in long RTs at '*eat*' and then they experienced a filled-gap effect at '*her*', where they were expecting an object trace. This finding suggests that they linked the wh-filler '*what*' with the main verb '*eat*' via an object gap. RTs for the embedded verb '*watched*' (Words 9) in the embedded clause did not show an increase in three groups. This result suggests that they did not expect an object gap inside the adjunct island to link the wh-filler with it. This means that the participants in three groups are sensitive to the adjunct island constraint.

Figure 35 presents RTs for each word of the ungrammatical wh-sentences with adjunct island, excluding the first and the last words of the sentences.

Figure 35. Word-by-word reading times for the ungrammatical wh-sentences with AI violation



In Figure 35, the spikes between the words '*eat*' and '*her*' show that all participants slowed down in reading this part of the sentence compared to the other parts. In addition, there is no increase in RTs by three groups for either at the embedded verb '*watched*' or the illicit gap site the preposition '*on*'.

A two-way ANOVA per region was conducted to examine whether three groups spent significantly longer RTs for the word the main verb '*eat*' (Word 4), the following word '*her*' (Word 5). The results of the first ANOVA revealed a significant main effect for region ($F (1, 80) =31.17; p<.05$), and a significant interaction between language and region involving '*Ann*' (Word 3) and the main verb '*eat*' (Word 4). Further analyses of region and language showed that the L2 learners spent significantly longer RTs at the main verb than the English native speakers. This suggests that the native English speakers were faster in filler-gap formation than the two L2 groups. The results of the second ANOVA demonstrated only a significant main effect for region involving the embedded verb ‘*eat*’ (Word 4) and the following word ‘*her*’ (Word 5) ($F (1, 79) = 16.58; p < .05$). Further analysis of region revealed that the three groups spent significantly longer RTs on ‘her’ (Word 5) where they expected an object gap. This finding confirms that all participants experienced a filled-gap effect at ‘*her*’.

To explore the differences in RTs for the words ‘*watched*’ (Words 9) and ‘*on*’ (Word 10) inside the adjunct island, a two-way repeated measure ANOVA was conducted with language as between-subject factors and region (Words 9 and 10) as the within subject-factors. The results of this analysis did not show any significant main effect for region ($F (1, 80) = .348; p > .05$) for language ($F (1,80) = .190; p > .05$) or for language and region ($F (2,80) = .555; p > .05$). These results show that none of the three groups was significantly different from each other in terms of RTs spent on the gap site in the adjunct island. This suggests that they did not attempt to integrate the wh-filler ‘*what*’ with the embedded verb ‘*watched*’ and an embedded object gap. This suggests that both the native speakers and the two L2 groups are sensitive to adjunct island constraint during processing of wh-sentences.

To sum up, results from RTs for each word of the ungrammatical wh-sentences with adjunct island showed that all three groups have the same reading pattern in the sense that they all had a filled gap effect at the first word (here, “*her*”) following the main verb, and they did not expect an object gap inside the adjunct island. This finding suggests that they obey adjunct island constraint. Remember that these results should be interpreted cautiously due to the absence of a control (non-island) condition for the ungrammatical wh-sentences with adjunct-island violation.

In the last section, I will discuss word-by-word reading for the ungrammatical wh-sentences with complex NP island violation.

5.3.4.5. Word-by-word reading times for ungrammatical wh-sentences with complex NP island violation

This category involves ungrammatical wh-sentences with complex NP islands as in "*Who does Clair know the fact that Jane hates at school?*" As in the previous ungrammatical types, extracting a wh-phrase from inside a complex NP/DP, whose head N (*fact*) takes a sentential complement is blocked because they are islands for movement. Therefore, our prediction was that the parser can look for an object gap for the wh-filler '*who*' after the main verb, '*know*' and it should experience a filled gap effect upon encountering the first word of object NP '*the*'. However, it should not search for an object gap after the embedded verb 'hate' because it is inside the complex NP if it is sensitive to the complex NP island constraint on the complex NP/DP island.

For the word-by-word analysis of the ungrammatical wh-sentences with complex NP island violation, mean RTs for each word by each participant were calculated, excluding RTs beyond ±2.5 standard deviations from the mean RTs of each participant. Then, RT data were checked for normal distribution and outliers. It was observed that the RT data were normally distributed for each word in each language group. Mean RTs for the each word of the ungrammatical wh-sentences with complex NP violation are presented on Table 15a, beginning with the second word "*does*" and ending with tenth word "at" as in the example sentence (**Who does Clair know the fact that Jane hates at school?)* .

Table 15a. Mean RTs for the ungrammatical *wh*-sentences with complex NP/DP island violation

L1Groups	Word 2	Word 3	Word 4	Word 5	Word 6	Word 7	Word 8	Word 9	Word 10
Spanish (n=20)									
CNPI (M)	415	444	547	538	503	547	438	425	409
(SD)	(117)	(123)	(182)	(200)	(137)	(177)	(143)	(123)	(123)
Turkish (n=29)									
CNPI (M)	371	399	508	556	469	461	418	404	390
(SD)	(87)	(115)	(192)	(226)	(115)	(147)	(131)	(126)	(143)
English (n=30)									
CNPI (M)	391	417	457	509	479	483	394	364	385
(SD)	(112)	(161)	(179)	(178)	(206)	(187)	(113)	(106)	(133)

CNPI: Complex Noun Phrase Island

The individual regions in an adjunct island sentence are given below:

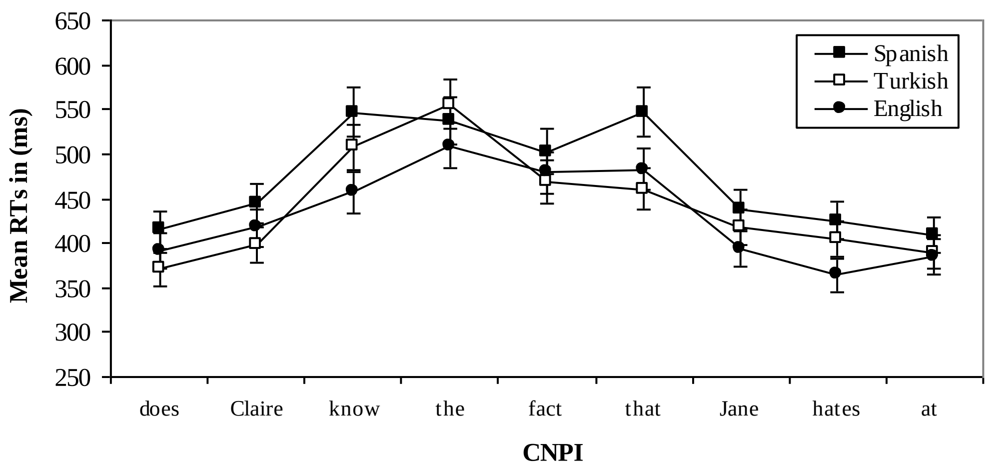
Table 15b. Complex NP/DP Island

1	2	3	4	5	6	7	8	9	10	11
*Who	does	Claire	know	the	fact	that	Jane	hates	at	school

Mean RTs on Table 15a show that RTs were increased at the main verb '*know*' (Word 4) in all groups, particularly in the L2 groups, where they associated the wh-filler '*who*' with the main verb via an object gap. RTs continued to increase at the initial word of object NP following '*know*' in English and Turkish group. This shows that English and Turkish groups had a filled gap effect at *'the'* following the main verb '*know*'. The Spanish group, on the other hand, showed a slight increase in RTs for the complementizer *that* (Word 7). RTs for the embedded verb '*hates*' (Word 9) and the following word '*at*' (Word 10) did not show an increase compared to the other words in the sentence. This result suggests that none of the groups expected an object gap inside the complex NP/DP island.

Figure 36 shows the distribution of mean RTs for each word of the ungrammatical wh-sentences with complex NP island violation.

Figure 36. Word-by-word reading for the ungrammatical wh-sentences with complex NP/DP island violation



As can be seen in Figure 35, RTs increased at the main verb '*know*' and the following article '*the*', showing a slow-down in reading in this region. However, RTs continued to decrease within the NP/DP island, including the embedded verb '*hates*' and the following preposition '*at*', where the illicit subject gap occurs.

A two-way ANOVA was conducted with language as the between-subject factor and region (Words 4 and 5) as the repeated within-subjects factor for each region. The results of this analysis showed only a main effect approaching to a significant level for region (F (1, 77) =3.10; p=.082). A further analysis of region for each language group revealed that, unlike Turkish and Spanish speakers, English native speakers spent significantly longer RTs for the definite article '*the*' following the main verb '*know*',, suggesting that the English native speakers were surprised upon finding an overt object NP at the place where they were expecting a subject gap.

To explore the difference in RTs for the embedded verb '*hates*' and the preposition '*at*', another ANOVA was conducted with language as the between-subject factor and region (Words 9 and 10) as the within-subject factors. The results did not show a significant main effect for region (F (1, 80) =.129; p>.05) '*hates*' or the following preposition '*at*' in the complex noun phrase island. This suggests that they did not expect an illicit gap for the wh-filler that would violate island constraints. Also, no significant difference was observed for language in both analyses, which shows that all language groups spent almost the same amount of RTs in these two critical regions.

To sum up, results from the RT for the sentences with the complex NP island violation indicated similar findings to those of adjunct and relative clause islands. None of the three language groups showed increased RTs at the gap site inside the complex NP/DP island, suggesting that they are sensitive to the complex NP/DP island constraint during online processing of ungrammatical wh-sentences.

In summary, overall results from accuracy judgments of five types of ungrammatical wh-extractions with island constraint violation in the moving window condition demonstrated that L2 speakers were as accurate as the native speakers in their grammaticality judgments except for *that*-trace violation. The most difficult wh-extraction type for all L2 participants was ungrammatical subject extractions with *that*-

trace. Between the two L2 groups, Spanish speakers were less successful in correctly rejecting sentence with *that*-trace violation. This suggests that their L1 background has a local effect on their acceptance of wh-extractions with *that*-trace, which are ungrammatical in English but grammatical in Spanish. This can be attributed to the fact that Spanish is a pro-drop language that allows the *that*-trace sequence. Although *that*-trace sequence is not relevant for Turkish, the Turkish group, like the Spanish group had problems in rejecting these ungrammatical constructions. Nevertheless, *that*-trace violations were also the most problematic structures or the English native speakers after the sentences with subject island violations.

With respect to RTs on the moving window condition, the analyses on ungrammatical sentences revealed that L2 speakers showed RT patterns similar to those of the native English speakers. In *that*-trace violation, the three groups displayed processing difficulty at the same region (the embedded verb following the complementizer *that).* However, the Spanish group spent shorter RTs in this region compared to the other two groups. This suggests that Spanish speakers tend to accept sentences involving subject extraction across the complementizer *that,* possibly due to the L1 Spanish influence. In wh-extractions from subject islands, unlike the Spanish group, Turkish and the native English speakers had longer RTs at the gap site inside the subject island. This showed that they expected an illicit object gap inside the island. This suggests that they are not sensitive to the subject island constraint. However, in the other three types of ungrammatical wh-extractions, namely relative clause island, adjunct island and complex NP island, neither of the three groups showed an increase in RTs for the gap site inside the islands. This clearly shows that they are sensitive to the island constraints during online processing of ungrammatical wh-dependencies.

5.3.5 Summary of the results from accuracy and reading time analyses in the moving window

In this section, we presented the results of both accuracy responses and self-paced word-by-word RTs to the five types of the grammatical and ungrammatical wh-extractions in English. The accuracy results from grammatical and ungrammatical wh-extractions in the moving window condition can be summarized as follows: all participants had a mean of 82 out of 100 in overall accuracy scores. The overall accuracy scores for the ungrammatical wh-sentences (M=41 out of 50) were slightly better than those for grammatical wh-sentences (M=40). All participants had slightly better total accuracy scores (M=83 out of 100) in the moving window condition than the full sentence condition (M=81). This may be due to the fact that in the moving window condition, all participants were tested on the wh-sentences they previously took in the full-sentence condition. In the full-sentence condition, they were more accurate on ungrammatical wh-sentences (M=42 out of 50) than grammatical wh-sentences (M=39). In the moving window condition, they were more accurate on grammatical wh-sentences (M=42) than on ungrammatical wh-sentences (M=41).

The native English speakers had a better mean score in the full-sentences condition (M=90 out of 100) than in the moving window condition (M=88). In the full-sentence condition, they were more accurate on ungrammatical wh-sentences (M=46 out of 100) than grammatical wh-sentences (M=44). In the moving window condition, their accuracy scores for ungrammatical decreased (M= 44), while the accuracy scores for grammatiacal wh-sentenecs increased (M=44). Overall, they were better on ungrammatical wh-sentences (M=45) than grammatical wh-sentences (M=44).

With respect to L2 groups, the Spanish learners were better in overall accuracy scores (M=80 out of 100) in the moving window condition, compared to the full sentence condition (M=74) in the full sentence condition. In the full sentence condition, they were more accurate on the ungrammatical wh-extractions (M=38) than the grammatical wh-sentences (M=36). In the moving window condition, they performed better on grammatical wh-sentences (M=41) than ungrammatical wh-sentences (M=39). Overall, they were equally accurate on grammatical wh-sentences and ungrammatical wh-sentences (M=39 vs. M=39)

Similarly, the Turkish group obtained higher accuracy scores in the moving window condition (M=80) than in the full-sentence condition (M=79). In the full-sentence condition, they were more accurate on ungrammatical wh-sentence (M=42 out of 50) than grammatical wh-sentences (M=37). In the moving window condition, they were equally accurate in both types (M=40 out of 50). Overall, they were better on ungrammatical wh-sentences (M=41) than ungrammatical wh-sentences (M=39).

The results of statistical analysis showed that the two L2 groups were significantly worse than the native English speakers in overall accuracy scores (i.e. total accuracy, accuracy in ungrammatical wh-sentences and accuracy in grammatical wh-sentences), but not different from each other. However, the results of the type-by-type analysis of accuracy scores revealed that the Spanish and Turkish learners were more or less as accurate as the native English speakers on grammatical wh-sentences, except for subject extractions from nonfinite clauses. They were also as successful as the native speakers on ungrammatical wh-extractions except for *that*-trace effect violation. These findings suggest that both the advanced Spanish and Turkish learners of English have the knowledge of overt wh-movement and of the constraints on wh-movement to be able to correctly judge grammatical and ungrammatical wh-sentences in L2 English.

In the full-sentence condition, RT results for grammatical and ungrammatical wh-sentences showed that the native English speakers were significantly faster than the two L2 groups. The two L2 groups were not significantly different from each other. The longest RTs were spent on subject extraction from finite clauses among the grammatical types, while they were spent on the wh-sentences with *that*-trace violation in ungrammatical types in the two L2 groups. The native speakers were similar to the two L2 groups in terms of RTs on grammatical types, but slightly different from them with respect of the RTs to ungrammatical wh-sentences. They spent longest RTs to subject island violation among the ungrammatical types, followed by the RTs to *that*-trace violation.

The results of the word-by-word RTs for grammatical and ungrammatical wh-extractions demonstrated that the Spanish and Turkish groups had a reading pattern similar to that of the native English speakers. With respect to grammatical wh-extractions, word-by-word RTs to subject extraction from finite clauses showed that all groups had most processing difficulty at the embedded NP in the embedded clauses, where they experienced a severe filled-gap effect. This effect was significantly larger in the two L2 groups. This suggests that L2 learners have more difficulty in recovering from a misanalysis in online L2 processing. That is, they have more problems with reassigning theta roles and Case during the reanalysis. In the finite clauses, RT results indicated that the locus of the difficulty for all participants was the embedded finite verb. In relation to ungrammatical wh-extractions, word-by-word RTs for wh-extractions with *that*-trace violation showed that the locus of the difficulty for all participants was the complementizer *that*. All groups spent significantly longer RTs at the embedded verb following the complementizer, showing that they were surprised at finding a finite verb after '*that*', but the Spanish learners were significantly faster than the other groups, showing that they were not as surprised as the other groups, which might be due to the fact that embedded verbs obligatorily follow the complementizer *that* in their L1 Spanish.

RT results for wh-extractions with island constraint violation showed the L2 groups were as sensitive as the native English speakers to the island constraints except for subject island condition. Neither of the three groups attempted to form an illicit wh-dependency inside the RC island, adjunct island, complex NP/DP island. However, the native English speakers and Turkish speakers associated the wh-filler with its object inside the subject island, which suggests that they are not sensitive to subject island constraint.

In the next section, I will discuss the results of the analysis on working memory tasks.

5.4. Results from working memory tasks

5.4. 1. Introduction

In this section, I will present the results obtained from the working memory (WM) tasks with respect to two critical questions: (1) Is WM a language-independent construct as suggested by Osaka & Osaka (1992)?; (2) Can the WM capacity either in the L1 or in the L2 account for the low performance of participants in processing grammatical and ungrammatical wh-dependencies in the L2 English. Recall that two online WM tasks were given in L2 English; (1) Automated Reading Span (ARSPAN); and (2) Automated Operation Span (AOSPAN).[2] In addition, a Turkish ARSPAN task for the Turkish participants and a Spanish ARSPAN task for the Spanish speakers were administered to find out whether or not the WM span is language-dependent. So, overall the four WM tasks were administered.

Recall also that, in the ARSPAN task, participants were required to read sentences while trying to remember a set of unrelated letters (B, F, H, J, M, Q, R, and X). For this task, participants read a sentence and determined whether it made sense or not (e.g., "*The prosecutors' dish was lost because it was not based on fact*"). Then, they clicked on either a 'TRUE' or a 'FALSE' button on the keyboard, depending on their judgment. On the subsequent screen, they saw a letter to be remembered. There were three sets of each set size, with the set size ranging from 3 to 7, which made for a total of 75 letters and 75

---

[1]L2 learners took the AOSPAN task only in the L2 English, the task involved simple math equations and letters to be remembered in English, which were not different in L1 Turkish or Spanish.

sentences. At the end of the task, the software reported the following scores to the experimenter: (1) *ARSPAN absolute score*, which was the sum of all perfectly recalled sets; (2) *ARSPAN total number correct*, which was the total number of letters recalled in the correct order; (3) two types of sentence errors: a) *speed errors,* in which the participant run out of time while judging a given sentence; b) *accuracy error*, in which the participant judge the sentence incorrectly.

In the AOSPAN task, participants were required to solve simple math operations while trying to remember a set of unrelated letters as in the ARSPAN task. They first saw a math operation (e.g., (1*2) +1=?)). The participants were required to solve the problem as quickly as possible and click the mouse to move to the next screen. On the next screen, a number (e.g., 3) was presented and the participants were required to click either a 'TRUE' or a 'FALSE' box that appears on the screen, depending on their answer. There were three sets of each set size, with the set size ranging from 3 to 7, which made for a total of 75 letters and 75 math operations. The same scoring procedure was used as in the ARSPAN.

The Spanish and Turkish ARSPAN tasks were the translated versions of the English ARSPAN task. The results of these tasks will be discussed in the final section of this chapter.

The reason for using two WM tasks in the L2 English is that a single task of WM cannot perfectly measure the construct it ostensibly represents (Cowan, et. al., 2005). For example, the automated operation span task is believed to measure WM capacity and also to tap the mathematical ability, motivation, and word knowledge among other constructs. Similarly, the ARSPAN task measures WM capacity as well as verbal ability. It has been suggested that despite being valid and strongly reliable measures of the WM capacity, the WM span tasks including operation span, reading span, and counting span are not perfect or process pure (see Engle et al., 1999; Kane et al., 2004; Cowan et al., 2005). Therefore, an optimal research strategy is to administer multiple WM span tasks and then to derive a latent variable from the common variance among those measures. Operation span, reading span, and counting span are particularly suited for latent variable analysis because they are moderately correlated with one another, suggesting that they are indeed tapping a similar construct, yet are not mere replicas of one another. Correlations among the three span tasks typically range from .40 to .60, suggesting that they are indeed taping some common process or ability, but also suggesting that they are not identical. In terms of measurement, this is an ideal situation; when a construct is measured with imperfect tools, it is best to use multiple reliable measures that do not replicate one another. One main advantage of latent variable is that a more reliable measure of WM capacity can be derived from two span tasks rather than one. As such the predictive power of latent variable is better than that of individual tasks. Another benefit of having multiple measures per construct is that multivariate outliers can be detected.

Conceptually, the latent variable represents only the variance between the two tasks and removes task-specific factors. In addition, the Turkish and Spanish versions of the English ARSPAN tasks were also prepared and used in this study to measure participants' L1 WM capacity with the assumption that the L1 WM data will reveal whether the WM capacity is language-independent as suggested by Osaka & Osaka (1992) or not. If the WM capacity is language-independent, then participants' memory spans should be similar in L1 and L2 WM measures. That is, a participant should get more or less the same memory span score from both L1 and L2 WM tasks. Also, these tasks should significantly correlate with each other. If, on the other hand, the WM capacity is language-dependent, different memory span scores should be obtained from the L1 and L2 WM tasks. Furthermore, participants are predicted to show better performance in their L1 measures.

To analyze the data from the L1 and L2 WM tasks, we used both correlational statistics and an extreme-group design because both procedures have some shortcomings. For example, the extreme group design has serious problems such as; (1) information and power loss because less variability is obtained by categories (i.e., leaving out the participants with mid-span); (2) overestimation of the relation between variables; and (3) tendency for misclassifying subjects due to measurement error (see Roberts & Gibson, 2003; and Conway et al., 2005 for a review). The correlational approach also has some shortcomings. For example, multiple comparisons require large numbers of participants in order to be reliable. Furthermore, with a large number of participants, statistically significant correlations may account for only a small amount of variance on a given test. Nevertheless, the correlational approach is still superior to the extreme group design approach if scatter plots do not show a natural grouping of data.

In the rest of this chapter, I will first compare the L1 and L2 ARSPAN scores of L2 groups to examine whether or not the WM capacity is language-independent. That is, the English ARSPAN, Turkish ARSPAN, and Spanish ARSPAN will be compared to each other. These analyses will also include a comparison of the L2 English AOSPAN scores with the L1 and L2 ARSPAN scores to explore the correlation between the AOSPAN and the ARSPAN tasks.

I will then examine whether or not the WM measure in the L1 and L2 WM correlate with the overall accuracy performance of the participants on five types of grammatical and ungrammatical long-distance wh-extraction presented in the full-sentence reading condition. More specifically, I will look at the relationship between the WM capacity and accuracy for the grammatical subject extraction from a nonfinite clause because it was found to be the most difficult sentence type to process. In addition, among the ungrammatical types, I will examine the relationship between the WM capacity and the ungrammatical wh-extraction with *that*-trace violations and subject island violations since they appeared to be the two most difficult categories for all participants. Similar analyses will be conducted with grouping participants according to their WM scores.

In the section 5.4.4, I will examine the relationship between the WM capacity and RTs on the critical regions in certain types of grammatical and ungrammatical wh-extraction (i.e., *that*-trace and subject island constraints). This will be followed by extreme-group analyses.

5.4. 2. Results of the correlations among the WM measures

In this section, I will first examine whether the L1 ARSPAN scores in Spanish and Turkish correlate with the L2 English ARSPAN scores. The prediction is that if the WM capacity is language-independent, then, the L1 ARSPAN scores of Turkish and Spanish participants should correlate with their L2 ARSPAN scores. In addition, if the ARSPAN task measures the same construct as the AOSPAN, as suggested by Unsworth et al., (2005), then the L1 and L2 ARSPAN scores should correlate with the AOSPAN scores as well.

In the ARSPAN task, the software collected five scores for each participant: Rspan score, total number correct, sentence errors, speed errors, and accuracy errors. The Rspan score was the sum of all perfectly recalled sets. So, for example, if an individual correctly recalled 3 letters in a set size of 3, 4 letters in a set size of 4, and 3 letters in a set size of 5, his/her Rspan score would be 7(3+4+0). The second score, 'total number correct' was the total number of letters recalled in the correct position. Three types of errors were reported: Sentence errors were total number of task errors, which was then broken down into 'speed errors', in which the participants ran out of time in attempting to solve a given sentence problem, and 'accuracy errors,' in which the participants solved the sentence problem incorrectly.

Similarly, in the AOSPAN task, the software collected five scores for each participant: Ospan score, total number correct, math errors, speed errors, and accuracy errors. The Ospan score was the sum of all perfectly recalled sets. So, for example, if an individual correctly recalled 3 letters in a set size of 3, 4 letters in a set size of 4, and 3 letters in a set size of 5, his or her Ospan score would be 7(3+4+0). The second score, 'total number correct' was the total number of letters recalled in the correct position. Three types of errors were reported: Math errors were total number of task errors, which was then broken down into 'speed errors', in which the participants ran out of time in attempting to solve a given mat operation and 'accuracy errors,' in which the participants solved the math operation incorrectly.

Table 15 below displays mean scores, standard deviations, minimum and maximum scores for each language group, which were obtained from the L1 Turkish and L1 Spanish ARSPAN, L2 English ARSPAN and AOSPAN tasks.

Mean scores on Table 15 indicate that Turkish speakers had the highest AOSPAN mean scores among the three groups. Their English ARSPAN mean scores are almost the same as the native English speakers' English ARSPAN scores (and higher than those of the Spanish speakers). The L1 ARSPAN mean scores of the Spanish speakers seem similar to the native English speakers' ARSPAN mean scores.

Table 16 below presents average scores of Rspan scores and Rspan total for L1 ARSPAN, L2 ARSPAN and average scores of Ospan scores and Ospan total for AOSPAN in each group.

Table 16 Mean scores of L1 ARSPAN, L2 ARSPAN and AOSPAN

	Spanish (n=24)				Turkish (n=30)				English (n=31)			
	M	SD	Min.	Max.	M	SD	Min.	Max.	M	SD	Min.	Max.
L1 ARSPAN	50.48	12	29	71	57.63	13	35	75	50.37	13	10	75
English ARSPAN	45.42	12	14	71	50.78	14	19	71	-	-	-	-
English AOSPAN	51.23	12	31	71	60.97	10	27	72	54.29	14	3	75

As shown in Table 16, mean scores on L1 ARSPAN and the L2 ARSPAN appear similar in three language groups. Among the three WM measures, AOSPAN has the highest mean scores in three groups and the Turkish group obtained the highest mean score in this task. Also, mean scores on AOSPAN are closer to mean scores on L1 ARSPAN than L2 ARSPAN in the L2 groups.

A one-way ANOVA conducted with mean scores on L1 ARSPAN in three groups showed a significant effect for language (F (2, 83) =3.16, $p<.05$). However, a post-hoc analysis revealed that the Spanish group's L1 ARSPAN scores were marginally lower than the Turkish group's (p=.69). This suggests that the Spanish group has marginally lower WM capacity than the Turkish in terms of the L1 ARSPAN scores.

Another one-way ANOVA which was conducted with mean scores on L2 ARSPAN in three groups did not show a significant difference among the language groups with respect to the L2 ARSPAN groups. This implies that the language groups are not different from one another with respect to L2ARSPAN scores.

The third ANOVA was carried out with mean scores on AOSPAN in three groups. This analysis showed a significant effect for language groups (F (2, 82) =4.32, p<.05). A post-hoc analysis revealed that the Spanish group's AOSPAN scores were significantly lower than those of the Turkish group (p=.016). No significant difference was found in the AOSPAN scores between the native English and the Turkish learners.

To sum up, these results suggest that the language groups are more or less similar in terms of L1 and L2 ARSPAN scores, although the Spanish group appears to have lower scores in both tasks than the other two groups. . The Turkish and the native English speakers are similar with respect to the AOSPAN scores. However, the Turkish group is significantly better than the Spanish speakers in term of AOSPAN scores.

To explore whether or not mean scores on the L1 ARSPAN given in Turkish and Spanish were significantly different from the mean scores on the L2 ARSPAN given in English and the AOSPAN in English, paired-samples t-tests were run. The first paired-samples t-test carried out with mean scores of the Spanish participants' L1 and L2 ARSPAN indicated that the Spanish participants had significantly better scores on L1 ARSPAN than L2 ARSPAN (t(22) =3.26, p<.05). Similarly, the Turkish participants were found better on the L1 ARSPAN compared to L2 ARSPAN (t (29) =3.18, p<.05).

With respect to AOSPAN scores, t-test analysis on mean scores of Spanish speakers' L1 ARSPAN and AOSPAN did not show a significant difference between these measures. However, both the native English and Turkish speakers had better mean scores on AOSPAN than those on their L1 ARSPAN.

To sum up, these results showed that for all three language groups, the AOSPAN scores were higher than their ARSPAN scores. Also, L2 participants' L1 ARSPAN scores were better than their L2 ARSPAN scores, but different from their OSPAN scores.

Although there are some differences in memory span scores among the WM measures, it is important to know whether or not these scores significantly correlate with each other. To examine this, the Pearson's correlation coefficient analysis was conducted on the WM measures (L1 Turkish ARSPAN, L1 Spanish ARSPAN, L2 English ARSPAN and English AOSPAN).

First, we looked at the correlation among the WM measures without breaking down groups by language. We found that the AOSPAN scores were significantly correlated with both the L1 ARSPAN scores ($r = .744$, $p < .01$) and the L2 ARSPAN scores ($r = .605$. $p < .01$). These results suggest that the AOSPAN is similar to ARSPAN in the way it measures the WM capacity. In other words, ARSPAN and AOSPAN reflect a common construct. This finding is consistent with findings of Turner and Engle (1989), Engle et al., (1999) Conway et al., (2002) and Unsworth et al. (2005). To note that, in WM research, correlations among the WM tasks such as counting span, reading span and operations span tasks typically range from .40 to. 60. This has been interpreted as evidence that these tasks are tapping some common process, but also suggests that they are not identical. Below, Tables 17, 18, and 19 present correlations among the WM measures in the three language groups.

Table 17 Correlations among the WM measures of English speakers

WM measures	ARSPAN	AOSPAN
ARSPAN	1	
AOSPAN	.768**	1

N=31. *p<.05 **p<.01

As shown in Table 17, the native English speakers' English ARSPAN scores are significantly correlated with their AOSPAN scores (r =. 768, p<.01). This suggests that both the ARSPAN and AOSPAN tasks are strongly related to each other and are valid indicators of WM capacity.

Table 18 Correlations among the WM measures of Spanish speakers

WM measures	L1 ARSPAN	L2 ARSPAN	L2 AOSPAN
L1 ARSPAN	1		
L2 ARSPAN	.681**	1	
AOSPAN	.657**	.582**	1

N=24. *p<.05 **p<.01

Correlation coefficients in Table 18 show that the Spanish learners' L1 ARSPAN scores were significantly correlated to their L2 ARSPAN scores (r =.681, p<.01) and their L2 AOSPAN scores (r =.657, p<.01). Their L2 ARSPAN scores were also significantly correlated with their AOSPAN scores (r =.582, p<.01). These results suggest that all WM measures of the Spanish group were strongly related to one another.

Table 19 Correlations among the WM measures of Turkish speakers

WM measures	L1 ARSPAN	L2 ARSPAN	L2 AOSPAN
L1 ARSPAN	1		
L2 ARSPAN	.619**	1	
AOSPAN	.753**	.425*	1

N=30. *p<.05 **p<.01

Correlation coefficients in Table 19 show that the Turkish learners' L1 ARSPAN scores were significantly correlated with their L2 ARSPAN (r =.619, p<.01) and with the AOSPAN scores (r =.753, p<.01). Their L2 ARSPAN scores were also significantly correlated with their AOSPAN scores (r =.425*, p<.05)

All in all, these results suggest that L1 ARSPAN tasks (Turkish and Spanish tasks) are as successful as (if not more) L2 English ARSPAN tasks in measuring the WM capacity of the participants. This implies that the WM capacity is language-independent as predicted by (Osaka & Osaka, 1992). In addition, strong correlation between L1 and L2 ARSPAN and AOSPAN scores suggests that the AOSPAN task is as successful as ARSPAN in measuring WM capacity construct. This supports the findings of Turner and Engle (1989) showing that WM capacity is independent of the specific nature of the processing component of the span task.

In the next section, I will discuss the relationship between the WM measures and accuracy and RTs obtained in the OGJT presented in the full-sentence condition.

5.4.3. WM measures and accuracy and RTs for wh-extractions in the full-sentence condition

In this section, I will look at the relationship between the WM measures and the total accuracy in the OGJT presented in the full-sentence condition to examine the role of WM capacity in correct judgments of the grammatical and ungrammatical *wh*-extraction.

Before I begin to discuss this relationship, however, I would like to briefly summarize the accuracy and RT results obtained in the full-sentence condition. Recall that the two L2 groups were significantly worse than the native English speakers in overall accuracy scores (i.e. total accuracy, accuracy in ungrammatical wh-sentences and accuracy in grammatical wh-sentences), but not different from each other. However, the results of the type-by-type analysis of accuracy scores revealed that the source of the difference in accuracy between the native English speakers and the Spanish and Turkish learners mainly stems from L2 learners' significantly lower accuracy in judging subject extractions from nonfinite clauses and *that*-trace violations. All participants, particularly L2 speakers were significantly less accurate and slower in online processing of subject extraction from nonfinite clauses than in processing of object extraction from nonfinite clauses. In judging ungrammatical sentences, L2 speakers, particularly Spanish speakers, were significantly less accurate on *wh*-extraction with *that*-trace violation, whereas the native English speakers were significantly worse and slower at *wh*-extraction with subject island constraints.

With respect to the RT results for grammatical and ungrammatical wh-sentences, in the full-sentence condition, it was found that the native English speakers were significantly faster than the two L2 groups. The two L2 groups were not significantly different from each other. In grammatical items, the longest RTs were spent on subject extraction from finite clauses. In ungrammatical items, on the other hand, the longest RTs were found in sentences with *that*-trace violation in the two L2 groups. The two L2 groups were similar to the native speakers in terms of RTs on grammatical types, but they were slightly less accurate with respect of the RTs on ungrammatical wh-sentences. In the native speakers' data, the longest RTs were found on subject island violation and on *that*-trace violation.

For the WM analysis, first of all, a z-score was computed for each participant on L1 ARSPAN and AOSPAN so that each test contributed equally to the latent (or composite) score. Then, a latent/composite variable was formed by taking the average of z-scores on English ARSPAN and AOSPAN tasks as the measure of the WM measure and was labeled as the '*L2 WM measure*". Also, latent variable was derived from the average of z-scores on L1 ARSPAN and AOSPAN tasks of each L2 groups as the measure of the WM capacity in the L1 and it is labeled as the '*L1 WM measure'.*

The first correlation was performed to examine whether or not L1 WM and L2 WM composite measures are significantly correlated with the overall accuracy scores (i.e., total accuracy in grammatical wh-extractions and accuracy in ungrammatical wh-extractions).

No significant correlation was found between the WM measures and the overall accuracy scores or accuracy scores in grammatical wh-sentences. However, the L2 WM measures were significantly correlated with the accuracy scores in ungrammatical wh-extractions ($r = .224$, $p < .05$). Further analysis of ungrammatical wh-extractions revealed that accuracy scores on wh-extractions with adjunct island violation were significantly correlated with both the L2 WM measures (r =265, p<.05) and the L1 WM measure (r =.222, p<.05). These results suggest that there is a moderate correlation between the WM measures and the accuracy scores for ungrammatical wh-sentence, particularly for wh-sentences with adjunct island violation.

I also looked at the relationship between the WM measures and RTs spent on five types of grammatical and ungrammatical wh-extractions. I first performed a correlation analysis between L1 and L2 WM measures and RTs for grammatical wh-extraction types. This analysis showed that RTs for object extraction from a finite clause significantly correlated with the L2 WM (r =-.224, p<.05) and the L1 WM (r =-.233, p<.05). More importantly, RTs for subject extraction from a nonfinite clause significantly correlated with the L2 WM (r =-.232, p<.05) and with the L1 WM (r =-.204, p=.074) measures. These results show that L1 and L2 WM capacity interact with RTs spent on these types.

In another correlation analysis, I looked at the relations between L1and L2 WM measures and RTs for five ungrammatical wh-extraction types. The results indicated that both L1 and L2 WM measures significantly correlated with all types of ungrammatical wh-extractions except for *that*-trace violation. More specifically, it was found that RTs obtained on adjunct island violation significantly correlated with the L2 WM measures (r =-.311, p<.01) and with the L1 WM measure (r =-.303, p<.01). Also, RTs on complex NP/DP island violation significantly correlated with the L2 WM (r =-.242, p<.05) and with the L1 WM measures (r =-.242, p<.05). There was also a significant correlation between RTs on relative clause island violation and the L2 WM (r =-.307, p<.01) and with the L1 WM measures (r =-.292, p<.01). In addition, RTs on subject island violation significantly correlated with the L2 WM (r =-.240, p<.05) and with the L1 WM (r =-.240, p<.05). These correlations indicate a moderate relation between the WM capacity and accuracy performance.

To sum up, analysis of correlations among the WM measures and RTs for grammatical and ungrammatical wh-sentences showed that the WM measures had a reliable relationship with RTs for object extraction from finite clauses and with subject extraction from nonfinite clauses. They also reliably correlated with all ungrammatical wh-extraction types except for RTs for wh-sentences with *that*-trace violation. Note, however, that correlations among the WM measures and RTs for grammatical and ungrammatical wh-extractions range from -.22 to -.30. This suggests that there is a moderate relationship between the WM measures and RTs.

In a grouping analysis, the participants were classified into three groups on the basis of their L1 and L1 WM measures. Independent-samples t-tests were carried out to examine whether the high span participants were significantly better than the low span participants in terms of accuracy overall accuracy scores, accuracy scores for grammatical wh-sentences, and ungrammatical wh-sentences. For the L1 WM measure, the results of the three independent-samples t-tests did not indicate a significant effect for span groups, except for accuracy scores for ungrammatical wh-sentences (p=.082), which had a marginal effect. Similar results were obtained from the analyses of the L2 WM measures and accuracy scores. These results imply that the high-span participant were not better than the low-span participant in terms of accuracy in grammatical sentences and marginally better in ungrammatical ones.

Similar analysis was conducted with RTs for grammatical and ungrammatical types. The independent-samples t-tests conducted to compare L1 WM (high- and low-span) participants' RTs to five types of grammatical wh-extractions did not show a significant difference between the two span groups. There was only a marginal effect for RTs on object extraction from finite clauses (p=.083). The analysis conducted with the L2 WM span groups showed three marginal effects: (1) for RTs on object extraction from finite clauses (p=.073), and (2) for object extraction from nonfinite clauses (p=.088) and (3) subject extraction from finite clauses (p=.088). These results are in line with those of accuracy analyses above, and suggest that the high-and low spans were not significantly different from each other in the amount of RTs they spent on grammatical wh-extractions except for some marginal effects.

The independent-samples t-tests carried out to compare the L1 and L2 WM span groups in terms of RTs to ungrammatical types indicated that the high-span participants were significantly faster than the low-span participants in processing all ungrammatical types: adjunct island (p=.002), complex NP/DP island (p=.015), relative clause island (p=.010), subject island (p=.017). The difference between the span groups in RTs for *that*-trace violation was marginal (p=.93). Similar findings were obtained from the analysis for the L1 WM span groups in terms of RTs. These results support the findings of the accuracy analysis, and suggest that the high span participants are better than the low span participants in terms of speed in reading ungrammatical wh-extractions, except for sentences with *that*-trace violation.

To sum up, the results of the grouping analysis showed some marginal effects of memory span in terms of accuracy on ungrammatical wh-extractions, except for those with *that*-trace violations. Also, the analysis showed significant effects for span groups in terms of RTs for both grammatical and ungrammatical types. Among the grammatical types, the high spans were faster than the low spans in processing object extraction from finite (without *'that'*) and nonfinite clauses, and subject extraction from nonfinite clauses at a marginal level. They were significantly different from the low spans in RTs for all ungrammatical wh-extractions except for *that*-trace violation.

To conclude, results from correlations suggest that there is a moderate relation between the participants' WM capacity and their accuracy scores for only ungrammatical wh-extractions except for those with *that*-trace violation, and that the WM capacity moderately correlates with RTs for grammatical wh-extractions involving subject extractions from nonfinite clauses, and ungrammatical wh-extractions (except for *that*-trace items). The grouping analyses support these correlations.

Table 20 below presents correlations between the English speakers' L1 WM measure, including the ARSPAN and the AOSPAN scores and their overall accuracy scores (i.e., total accuracy, accuracy in grammatical sentences and accuracy in ungrammatical sentences).

Table 20 Correlations among the WM measures and accuracy scores for English native speakers

	1	2	3	4	5	6
1. WM[1]	-					
2. ARSPAN	.940**	-				
3. AOSPAN	.940**	.768**	-			
4. Total Accuracy	.009	.095	-.078	-		
5. Accuracy in grammatical sentences	.006	.035	-.024	.874**	-	
6. Accuracy in ungrammatical sentences	.009	.126	-.110	.782**	.322	-

N= 31. *p<.05 **p<.01 [1]= WM is the latent variable derived from the average of z-ARSPAN and z-AOSPAN scores.

As presented on Table 20, no significant correlation was observed among the WM measures and the accuracy scores of English native speakers. This suggests that there is not a reliable relationship between the participants' WM capacity and their performance in accuracy judgment of grammatical and ungrammatical wh-extractions in English. In addition to this analysis, the native English speakers were divided into three span groups: low-span, mid-span, high-span group on the basis of their composite WM scores. The one-way ANOVA conducted on mean scores of total accuracy, grammatical accuracy and ungrammatical accuracy did not show a significant difference among the three span groups. This implies that individual differences in the WM capacity do not affect the native English speakers' performance in accuracy judgment of grammatical and ungrammatical wh-extractions.

Similarly, the analysis carried out to examine the relationship between the native English speakers' WM measures and their RTs for the grammatical and ungrammatical wh-sentences did not display a significant correlation. In addition, the comparison of the span groups did not show a significant difference between the low and the high span participants in terms of speed. These results support the findings of the accuracy analyses with respect to the relationship between the WM capacity and RTs. In other words, the results imply that the high-span participants are not significantly faster than the low-span participants in processing both the grammatical and ungrammatical wh-extractions.

These findings suggest that the native English speakers' performance in accuracy and RTs for grammatical and ungrammatical wh-extractions is not influenced by their WM capacity.

Table 21 presents correlations between the L1 and L2 WM measures and overall accuracy score of Spanish speakers.

Table 21 Correlations among the WM measures and accuracy scores of Spanish speakers

	1	2	3	4	5	6	7	8
1. L1WM	-							
2. L2WM	903**	-						
3. L1ARSPAN	.914**	769**	-					
4. L2ARSPAN	.657**	.889**	.681**	-				
5. AOSPAN	.907**	.889**	.657**	.582**	-			
6. Total Accuracy	.609**	.487**	.493*	.311	.555**	-		
7. Accuracy in GR. Sentences[1]	.552**	.327	.423*	.137	.445*	.846*	-	
8. Accuracy in UNGR. Sentences[2]	.375	.436*	.336	.382	.394	.646**	.138	-

N= 22. *p<.05 **p<.01 [1]=Accuracy in grammatical sentences, [2]=Accuracy in ungrammatical sentences

Correlation coefficients in Table 21 show the Spanish speakers' total accuracy scores strongly correlated with L1 WM measure (r =.609, p<.01) and L2 WM (r =.487, p<.01). Their accuracy scores for grammatical wh-dependencies are significantly related to their L1 WM measure (r =.552, p<.01). Their accuracy scores for the ungrammatical wh-sentences significantly correlated with the L2 WM (r=.436, p<.01).

To explore which types of grammatical and ungrammatical wh-sentences contributed more to the significant correlation between the WM measures and the accuracy scores, I performed a correlation analysis for each type. Among the grammatical types, object extractions from finite clauses with the complementizer *that* significantly correlated with the L1 WM measure (r =.462, p<.05) and with the L2 WM measure (r =.400, p=.053). Also, there was a significant correlation between the L1 WM and the subject extractions from finite clauses (r =.441, p<.05) and with object extraction from nonfinite clauses (r =.466, p<.05). Since L1 ARSPAN scores were slightly better than the L2 ARSPAN scores, this caused a slight increase in the correlation between the L1 WM measure and grammatical accuracy. Surprisingly, accuracy scores on subject extraction from nonfinite clauses which were the lowest in five types did not significantly correlate with the any of the WM measures. This implies that the Spanish speakers' low accuracy on this type is not due variation in their WM capacity.

Among the ungrammatical types, accuracy scores for adjunct and relative clause island violations significantly correlated with the L1 measure (AI; r =.605, p<.01and RCI; r =.508, p<.05) and L2 WM measures (AI; r =.655, p<.01 and RCI; r =.558, p<.01). These results suggest that there is a strong positive relationship between the Spanish speakers' accuracy scores for adjunct and relative clause islands and their WM capacity. The accuracy scores for the wh-extractions with *that*-trace did not significantly correlate with the WM measures although it was the type received the least accuracy by the Spanish speakers. This implies that the low-accuracy of the Spanish learners on this type is not related to their WM capacity.

To sum up, correlations among the WM measures and overall accuracy scores (i.e., total accuracy, accuracy in grammatical sentences and accuracy in ungrammatical sentences) showed that the L1 WM measure strongly correlated with both total accuracy scores, and accuracy scores for grammatical wh-sentences. However, among the grammatical types, subject extraction from nonfinite clauses did not have a reliable relationship with the WM measures in the L1 and L2. Similarly, no reliable relation was observed for wh-extractions with that trace violation.

We also divided the Spanish participants into three span groups according to their L1 and L2 WM scores: low-, mid-, and low span groups, and then, compared the three groups using ANOVAs on overall accuracy, accuracy in grammatical sentences and accuracy in ungrammatical sentences.

First, we examined the difference between the high- and low-span groups using one-way ANOVAs. The results revealed that the high-span Spanish participants were significantly better than the low-span Spanish participants in overall accuracy (F (2, 23) $=6.16$, $p<.05$) and accuracy in grammatical sentences (F (2, 23) $=3.85$, $p<.05$), but they were not different in accuracy in ungrammatical sentences (F (2, 23) $=2.38$, $p>.05$). This implies that individual differences in the WM capacity significantly affect the Spanish participants' performance in overall accuracy and accuracy in grammatical sentences, but not in ungrammatical wh-sentences.

We conducted repeated measures 3x2 ANOVAs, with span group as between-subjects factor and grammatical type as the within-subjects factor. The results showed a significant effect for the L1 WM span groups (F (1, 21) $=3.85$, $p<.05$), and a marginal effect for the L2 WM span groups (F (1, 21) $=2.90$, $p=077$). A post-hoc analysis revealed that the high-span Spanish participants were significantly better than the low-span Spanish participants in accuracy on grammatical types.

In addition, one-way ANOVAs were conducted on accuracy scores of the three span groups per type. This analysis demonstrated that the high-span participants were significantly more accurate than the low-span participants on object extractions from finite clauses with the complement *that* ($p=.040$), and subject extraction from finite clauses ($p=.078$). However, no significant difference was observed among the span groups in accuracy on subject extraction from nonfinite clauses. Remember that subject extraction from nonfinite clauses received the lowest accuracy in all grammatical types. This suggests that low accuracy on this type is not due to the individual differences in the WM capacity of the Spanish participants.

With respect to the difference in accuracy on ungrammatical types, the repeated measures ANOVA, with span as the between-subjects factor and the type as the within-subjects factors demonstrated a significant effect for L2 span ($F(1, 21) = 3.54$, $p < .05$) . The results of one-way ANOVAs per type indicated that the high-span Spanish participants were significantly better than the low-span participants on ungrammatical wh-extractions with adjunct island ($p = .011$), relative clause island ($p = .049$), complex NP/DP island ($p = .047$) and subject island ($p = .071$) violations. However, no significant difference was found among the span groups in terms of accuracy on for the ungrammatical wh-extraction with *that*-trace violation. These results are in line with the findings of correlation analyses above.

In the same vein, I looked at the relationship between the Spanish participants' WM capacity and their RTs to grammatical and ungrammatical wh-sentences. More specifically, I examined whether the difficulty in subject extraction from nonfinite clauses and the wh-extractions with *that*-trace violation is related to the Spanish learners' WM capacity. We first examined the relationship between the WM measures (in L1 and L2) and RTs for the grammatical wh-extractions. The results did not show a significant effect of memory span on RTs obtained in five types of grammatical wh-extractions.

In terms of RTs for ungrammatical types, the results of the 3x2 ANOVA were similar to the findings of RT analysis of the grammatical types in that no significant effect was observed for span. That is, there was no significant relationship between the Spanish speakers' RTs to ungrammatical wh-sentences and their WM capacity.

To sum up, the results from the analysis of the relationship between the WM capacity and accuracy scores and RTs for grammatical and ungrammatical wh-sentences showed that the Spanish speakers' low accuracy on subject extraction from nonfinite clauses or on wh-extractions with *that*-trace violation cannot be accounted for by the variation in their ' WM capacity.

Table 22 below presents correlation coefficients between the L1 and L2 WM measures and accuracy scores of the Turkish participants.

Table 22 Correlations among the WM measures and accuracy scores of Turkish speakers

	1	2	3	4	5	6	7	8
1. L1WM	1							
2. L2WM	.885**	1						
3. L1ARSPAN	.936**	.813**	1					
4 L2ARSPAN	.557**	.844**	.619**	1				
5. AOSPAN	.936**	.844**	.753**	.425*	1			
6. Total Accuracy	.196	.076	.167	-.071	.199	1		
7. Accuracy in GR. Sentences[1]	.041	-.061	.082	-.097	-.005	.812**	1	
8. Accuracy in UNGR. Sentences[2]	.280	.211	.176	.009	.314	.621**	.047	1

N= 22. *p<.05 **p<.01 [1]=Accuracy in grammatical sentences, [2]=Accuracy in ungrammatical sentences

As can be seen in Table 22, Turkish speakers' WM spans did not significantly correlate with their accuracy scores, suggesting that, WM spans of the participants did not affect their accuracy performance in the OGJT in the full sentence condition as in the case of English native speakers.

However, correlations performed for RTs on grammatical and ungrammatical wh-extractions indicated significant relations between the WM measures and RTs. Among the grammatical types, RTs for subject extraction from nonfinite clauses were significantly related to the L1 ($r = -.382$, $p = .065$) and the L2 WM measures ($r = -.466$, $p < .05$). This shows a moderate relationship between RTs for this type and the WM capacity and suggests that variations in Turkish learners' WM capacity can partly account for the difficulty they had in this type. Among the ungrammatical types, RTs for adjunct and relative clause islands had a moderate relationship with the WM measures in L1 and L2. No reliable relationship was observed between RTs for wh-sentences and the WM capacity.

As in the other language groups, we also analyzed the WM data dividing the Turkish participants into three span groups. A repeated measures 3x2 ANOVA, with span as the between-subjects factor and the type as the within-subjects factor did not show a significant effect for span in the L1 and L2, which means that span groups are not significantly different from one another in terms of accuracy scores for grammatical types. Similarly, no significant difference was found among the span groups in terms of accuracy for ungrammatical types. These results are consistent with those obtained from correlations and suggest that the high-span participants are not better than the low-span participants in accuracy scores for grammatical and ungrammatical wh-extractions in L2 English.

With respect to RT data, the results of 3x2 ANOVAs, with span group as the between-subjects factor and the grammatical type as the within-subjects factor did not show any significant effect of span on both the L1 and L2 WM measures. This suggests that the high-span Turkish participants are not faster than the low-span participants in processing any type of ungrammatical wh-extractions.

In sum, results from the online WM tasks revealed that the correlations between the WM span scores and accuracy scores were not statistically significant in Turkish and English native speakers. This suggests that the WM capacity did not affect their performance in overall accuracy in grammaticality judgments. However, in the Spanish group, it did, to some extent. This can be accounted for by the differences in WM span scores in three groups. Recall that although three groups were not very different from one another in terms of L1 ARSPAN scores, the Spanish group's L2 ARSPAN and AOSPAN scores were lower than those of the native English speakers and of the Turkish learners. Their L1 ARSPAN scores were also lower than the Turkish learners. Also, recall that the Spanish learners' overall mean accuracy score was 74 out of 100, compared to that of the native English speakers (M=90) and of the Turkish learners (M=79). This suggests that variation in WM capacity among the language groups affects their performance in processing wh-sentences in English. More specifically, the low accuracy of the Spanish learners in processing wh-dependencies can be accounted for by their low WM capacity.

A further analysis was conducted to examine whether there was a significant relation between Spanish speakers' WM capacity and their low performance in judging the ungrammatical wh-sentences with *that*-trace violation in Spanish. However, no significant correlation was found. This result suggests that Spanish speakers' WM capacity cannot account for their low accuracy in processing wh-sentences with *that*-trace violation. Since no significant correlation was found among the WM measures and accuracy judgments of the grammatical and ungrammatical wh-types in the other two language groups, no further analysis was conducted for the types within these groups.

In the next section, I will look at the results of the WM measures with respect to accuracy and RTs in the moving window condition.

5.4.4 WM measures and RTs for wh-extraction in the self-paced word-by-word reading condition

In this section, I will examine the relationship among the WM measures and RTs spent on critical words in certain wh-extractions, where the processing load was the highest and where the WM capacity would most likely show an effect (Just et al. 1996). Recall that among the sentences involving grammatical wh-extraction, subject extraction from a nonfinite clause was the most difficult sentence type for all groups. In the ungrammatical sentence types, wh-extraction with *that*-trace (for Spanish and Turkish groups) and subject island violation (for the English group) were the two most difficult categories. Recall that in the full sentence condition, it was found that all participants were less accurate and slower in processing subject extraction from a nonfinite clause than object extraction from a nonfinite clause.

Results from RTs for each word of sentences involving subject extraction from a nonfinite clause (e.g. *Who does the manager expect to meet the job applicants today?*) revealed that all participants slowed down at a point where they encountered the determiner '*the*' of the embedded overt NP (*the job applicants*) after the embedded verb '*meet*', suggesting that they had a filled-gap effect at '*the*'. This finding showed that all participants were expecting an embedded object NP trace after the embedded verb '*meet*' for the wh-filler '*who*'. The longer RTs on the determiner '*the*', suggests that as soon as the parser encounters an overt object NP at the place of the embedded object trace, it revises this analysis and posits an embedded subject trace for the wh-filler to link. This assumption was supported by findings of the RT analysis in object extraction from nonfinite clauses. RTs for the preposition '*at*' following the embedded verb '*meet*' in object extractions from nonfinite clauses as in (*Who does the manager expect to meet at work this morning?)* were significantly shorter than the determiner '*the*' in the subject extraction from a nonfinite clause. This suggests that the participants did not get a filled gap effect at word (here '*at*') following the embedded verb in the object extraction.

To examine whether or not the difference in RTs for '*the*' and '*at*' in two types is related to the WM resources, we performed a correlation for overall RTs for the words '*the*' and '*at*' and the L1 and L2 WM measures. It is well known that online WM effects are most detectable when the parser is exposed to intense pressure from processing ambiguity (.e.g., MacDonald et al., 1992).

The first correlation analysis was conducted to explore the relation between the WM measures and RTs for the determiner of embedded NP in subject extraction from nonfinite clauses, which was the locus of processing difficulty. The results of this analysis did not show a significant relationship between the WM measures and RTs for '*the*'. I also examined whether the high-span group was faster than the low-span group in processing the determiner '*the*' in subject extraction, no significant difference was found between the span groups in terms of RTs. This suggests that overall, long RT for '*the*' does not tax large WM resources, or exceed the WM capacity of the participants.

To examine the differences among the three language groups in terms RTs for the overt NP in subject extraction from nonfinite clauses, I carried out correlation and extreme-group analyses in each language group. Recall that the two L2 groups were slower and less accurate than the native speakers in processing subject extraction from finite clauses. First, I performed a correlation analysis for the Spanish group's RTs on '*the*' and their WM measures in the L1 and L2. The results (ARSPAN in English) showed a significant moderate positive correlation (.478, p=.045) between the WM measures and the RTs in subject extractions. Then, I compared span groups with respect to the RTs on '*the*' in subject extraction and found that the high-span participants spent longer RTs on '*the*', which is the critical region in subject extraction from finite clauses (.p=,50), compared to the low spans. These results suggest that the Spanish speakers' slow RTs are due to variation in their WM capacity.

Second, we performed correlations and extreme-group analyses for the Turkish participants' RTs on '*the*' in subject extraction from finite clauses, but the results did not show any significant difference or correlation between the WM measures and RTs. Similarly, no significant effect or a significant correlation was observed in the RTs of the native English speakers for the determiner '*the*' in the subject extraction from finite clauses.

In sum, these results showed that the WM measures of both the English and Turkish speakers were not significantly related to their RTs for the determiner '*the*', which was the locus of the difficulty in subject extraction from finite clauses. This suggests that the difficulty in subject extraction cannot be attributed to variation in WM capacity of the participants. However, a significant positive relationship was found between the Spanish speakers' L2 Rspan scores and their RTs for the critical region in subject extractions, suggesting that the high-span participants spent longer RTs than the low-span participants for the determiner in this type.

We also conducted a correlation analysis to examine the relationship between the WM measures and RTs for the critical word 'witnessed' following the complementizer '*that'* in wh-extraction with *that*-trace violation (e.g. *Who do the police think that witnessed the accident on Sunday?*). The result of the correlation was not significant. The results of the extreme group analysis for the English native speakers showed a marginal effect for group (t(13) =1.88, p=.082), suggesting that the high-spans were marginally faster than the low-spans in processing the embedded verb '*witnessed*' following the complementizer *that*. Unlike the native English speakers, there was a significant positive correlation between the Turkish learners' WM measures and their RTs for the embedded verb '*witnessed*' (r=.532, p<.05), suggesting that high-span Turkish participants spent longer RTs than the low-span participants on '*witnessed*'. In the analysis of the Spanish group, no significant correlation was found between the Spanish group's WM measures and their RTs to the verb following the complementizer *that*.

The last correlation analysis was conducted for the examination of the relationship between the WM measures and the RTs on the embedded verb '*amuses*' in *wh*-extraction with subject island violation (e.g. *Who does the teacher believe a story by* *amuses the children?*). The results of the correlations and extreme-group analyses did not show a significant effect or relationship between the WM measures and RTs to the embedded verb '*amuses*' in subject island violation for native speakers although the high-spans were faster than the low-spans. No significant effect or correlation was observed for the two L2 groups' RTs on the embedded verb in the sentences with subject island violation.

To sum up, the results revealed no significant correlation between the WM measures and the RTs spent on the critical regions in sentences involving subject extraction from finite clauses, *that*-trace, and subject island violations, except for two cases: (1) Spanish speakers' data, which displayed moderate positive correlation between the RT for subject extraction from nonfinite clauses, and their WM measures, and (2) Turkish speakers' data, which showed a moderate correlation between RTs for the critical region in *that*-trace and the WM measures. This suggests the WM capacity cannot fully account for the difficulty the participants had in processing wh-sentences in this study.

5.4.5. Summary

Overall results of correlation analyses indicate that participants' English ARSPAN scores strongly correlate with their English AOSPAN scores. This suggests that the ARSPAN and AOSPAN tasks measure the same construct as suggested by Unsworth et al. (2005). Also, the scores obtained from the Turkish and Spanish versions of the English ARSPAN task strongly correlate with the scores of the English ARSPAN and AOSPAN tasks. This implies that the WM capacity is language-independent as suggested by Osaka and Osaka (1992).

Results from the analyses of correlation between the WM measures and accuracy scores obtained from the OGJT indicate no significant relation between these variables in Turkish and English native speakers. However, in the Spanish group, the WM capacity is found to be strongly related to overall accuracy scores. A type-by-type analysis indicates that the Spanish learners' WM capacity correlates with accuracy scores on object extraction from finite clauses without '*that*', and subject extraction from finite clauses. However, their WM capacity does not correlate with accuracy scores on subject extraction from nonfinite clauses. These results suggest that to some extent the WM capacity can account for the differences between the Spanish learners on the one hand and the native English and Turkish speakers on the other in accurate judgments of grammatical and ungrammatical items (except for sentences involving subject extraction from nonfinite clauses and *that*-trace violation).

Also, it cannot account for the processing load that all participants had in subject extraction from finite clauses, and the ungrammatical wh-extraction with *that*-trace violation and subject island violation. These results can be evidence for the Separate Language Interpretation Resource Hypothesis proposed by Caplan and Waters (1999), which suggest that the WM capacity does not affect online syntactic processing because there is a separate pool of verbal WM resources for interpretive processing which involves automatic, first-pass reading using all sources of information such as syntactic information, semantic and discourse information.

CHAPTER 6

DISCUSSION AND CONCLUSION

6.1. Introduction

In this chapter, I will discuss to what extent the results obtained from the two online experiments and the WM tests in the L1 and the L2 can inform us about the following research questions:

1. Is there a subject/object asymmetry in processing *wh*-extraction from finite and non-finite grammatical sentences? If so, which wh-extraction type causes more difficulty for native and non-native groups?

2. Do L1 Turkish and L1 Spanish learners of L2 English pattern similarly to native English speakers in their acceptance of grammatical *wh*-extractions and rejection of ungrammatical *wh*-extractions with island constraint violation in English?

3. Does the L1 of the learner play a role in real-time processing of wh-extraction in L2 English? In other words, is a wh-in-situ L1, which allows overt movement in scrambling, and which exhibits certain island constraints, as effective as Spanish— a language similar to English in the way it exhibits overt movement and island constraints, in processing L2 English wh-dependencies?

4. Does the WM play a role in processing wh-dependencies in English?

    a) Is WM capacity language independent?

    b) Do differences in the WM capacity affect L2 learners' speed and accuracy in online processing of L2 wh-dependencies? More specifically, is there a relationship between L2 learners' WM capacity in the L1 and L2; and is there a relationship between L2 learners' WM capacity in the L2 and their L2 syntactic parsing performance in wh-extraction from both finite and non-finite grammatical sentences?

Recall that we tested advanced Spanish and Turkish learners and a group of native English speakers on an online grammaticality judgment task involving: (1) long-distance grammatical subject and object extractions from finite and non-finite clauses, and (2) long-distance ungrammatical wh-extractions with island constraint violations. The stimuli were first presented in the full-sentence condition in which participants saw a full sentence on the screen at a time and judged whether it was a grammatically correct sentence in English or not. Meanwhile, the computer collected accuracy scores and RTs for each sentence. In addition, the same stimuli were presented on a self-paced word-by-word reading condition in a complementary experiment to collect RTs on each word of each sentence to explore the locus of the processing problems during real-time processing of grammatical and ungrammatical wh-extractions in English. To measure the WM capacity of the participants in their L1 and L2 English, ARSPAN and AOSPAN tasks were administered in English. The L2 learners were also tested on ARSPAN in their L1.

6.2. Discussion

The first issue investigated is a potential subject/object asymmetry in processing wh-extraction from both finite and non-finite grammatical sentences. The study also tries to identify which wh-extraction type causes more difficulty for native and non-native groups. Accuracy results in the full sentence condition and the self-paced word-by word reading condition reveal that for all participants, subject extraction from non-finite clauses is the most difficult type to process. This difficulty is more apparent in the two L2 groups. In contrast, object extraction from non-finite clauses poses no problem for any groups. As a matter of fact, this is the easiest type to process for the L2 groups. RT results are similar to the accuracy results in the sense that subject extractions from non-finite clauses elicit the longest, and object extractions from non-finite clauses elicit the shortest RTs in the full sentence condition. This difference in subject-object extractions is not, however, observed in finite clauses. This disconfirms the findings of Juffs (2005), who also reveal a subject-object difference in finite clauses. Nevertheless, a subject-object dissociation found in non-finite clauses in the present study is consistent with the findings of White and Juffs (1998) and Juffs and Harrington (1995).

Following Juffs and Harrington (1995), we can account for this asymmetry through Pritchett's (1992) 'Generalized Theta Attachment Theory', according to which during subject extraction from non-finite clauses such as "*Who does the manager expect to meet the job applicants*?", the parser makes at least three reanalyses: (1) from matrix object trace to subject trace; (2) from subject trace to PRO and object trace; (3) from PRO to subject trace and end of the parse. However, in object extraction from non-finite clauses such as "*Who does the manager expect to meet at work this morning*?", the parser needs to do a reanalysis of the structural position and theta /Case assigner but does not require reanalysis of the type of the Theta role and Case because they remain the same. Therefore, subject extraction from a non-finite clause becomes costlier than object extraction from non-finite clauses.

Our findings are the first to confirm the claim that both native and non-native speakers have more difficulty during online processing of subject extractions from non-finite clauses but not in object extractions from non-finite clauses. Results from the self-paced word-by-word reading in the moving window condition reveal that parsing steps are identical in both subject and object extractions from non-finite clauses until the embedded verb '*meet*' (as in the above example) but these steps dramatically change just after the embedded verb. RTs on the determiner '*the*' following the embedded verb '*meet*' in subject extractions significantly increase, but RTs on the preposition '*at*' following the embedded verb in object extraction do not. Following a principle-based parsing account (e.g., Pritchett, 1991; 1992; Gibson, 1991; Weinberg, 1999; Gibson, Hickok and Schütze, 1999), which proposes an active gap creation, we assume that the parser initially posits a matrix object trace gap as soon it encounters the matrix verb '*expect*' for the wh-filler '*who*', (as reflected in longer RTs at '*expect*'). However, this analysis fails when the parser encounters '*to*' and the embedded verb '*meet*'. Thus, RTs starts to increase. The parser revises the matrix object gap analysis at the embedded verb '*meet*', and posits a PRO and an embedded object gap at '*meet*', reflected in slow RTs at '*meet*'. Upon encountering the determiner '*the*' following the embedded verb '*meet*', the parser experiences a severe filled gap effect because it expects an embedded object gap after the embedded verb '*meet*' but not an overt NP like '*the job applicants'* . These results show that the locus of the difficulty in subject extractions for the native speakers and L2 learners is the embedded object NP. The difference between L2 learners and the native speakers in accuracy scores for this type can be attributed to the fact that for an L2 speakers recovering from a misanalysis during online processing is costlier. Thus, they are slower and less accurate than native speakers. This finding is consistent with the findings of Williams et al. (2001), which suggest that L2 learners may have more difficulty than native English speakers in recovering from a misanalysis which requires simultaneously changing theta roles and Cases.

Note, however that due to the nature of our materials, our results can also be interpreted within the Gapless (i.e., the Gap-free) account of Pickering and Barry (1991), according to which the parser directly associates a wh-filler with its subcategorizer and gets theta and Case without making use of gaps (traces). As the purported gap site is adjacent to the subcategorizing verb in languages like English, the slowdown elicited at the gap site following the subcategorizing verb may also be due to readers' attempt to link the moved wh-filler directly with its subcategorizer (see Chapter 3 for more information). Thus, our findings are compatible with both the gap-based and gapless accounts on sentence processing.

In sum, the findings of this study show that there is a subject and object asymmetry in wh-extractions from non-finite clauses, but not from finite clauses. In line with the findings of Juffs and Harrington, (1995; 1996), the findings suggest that the locus of the difficulty in subject extraction from non-finite clauses is the overt embedded NP. Both the native speakers and L2 learners experience a filled-gap effect as soon as they encounter the embedded object NP, where they expect an embedded object NP trace.

In addition, these results suggest that L2 learners process wh-dependencies in the same way as the native speakers, using similar processing strategies. Although there were differences between native and non-native speakers in terms of RT scores, it is important to note that the patterns of processing wh-sentences were the same in native and non-native groups (Williams, 2006; Dussias & Cramer Scaltz, 2007). Thus these findings diverge from the Shallow Processing Hypothesis (e.g., Clahsen & Felser, 2006; Felser & Roberts, 2007; Marinis et al., 2005), which argues for a fundamental difference between native and non-native structure-building processes during online L2 sentence comprehension.

The above mentioned parallelism between native speakers and L2 groups in sentence processing patterns also relates to the second question addressed in this study, namely the potential differences between the two L2 groups as well as the differences between the L2 groups and native-speakers. When we look at the overall results from accuracy responses to grammatical and ungrammatical wh-extractions in the full sentence and the self-paced word-by-word reading conditions, we see that the participants are accurate 83% of the time in their judgments. They are slightly more accurate on ungrammatical wh-extractions (M=42 out of 50) than grammatical wh-extractions (M=41). In the full sentence condition, the native English speakers obtain significantly higher scores in overall accuracy (M=90) than the Turkish learners (M=79), and the Spanish learners (M=74). They are also more accurate on grammatical wh-sentences (M=44) and on ungrammatical wh-sentences (M=46) than Turkish learners (M=37; M=42) and the Spanish learners (M=36; M=38). However, the results from the type-by-type analysis of all items reveal that the source of the difference between the native and non-native speakers is the low accuracy of L2 learners on subject extraction from a non-finite clause and wh-extractions with *that*-trace violation.

In the full-sentence condition, it is also found that L2 learners demonstrate an accuracy hierarchy in wh-processing similar to that of the native English speakers. In judging grammatical wh-extractions, the most difficult type for all participants appears to be subject extraction from non-finite clauses. This is followed by object extraction from finite clauses with the complementizer *that* and subject extraction from finite clauses. These results also support previous research findings, which show that in grammatical wh-extractions, L2 learners experience more processing difficulty with subject extractions from non-finite than subject extraction from finite clauses (Juffs and Harrington, 1995; 1996; White and Juffs, 1998; cf. Juffs, 2005).

RT results from grammatical wh-extractions in the full sentence condition support this hierarchy in that although the L2 learners are slower than the native speakers, for both native and non-native groups, RTs for subject extraction from non-finite clauses incurs longest RTs among five types. This is followed by object extraction from finite clauses with complementizer *that*.

With respect to ungrammatical items in the full-sentence condition, we see that the Turkish and Spanish learners pattern with the native speakers in rejecting ungrammatical wh-extractions as well. The only difference between them occurs due to L2 learners' significantly lower accuracy in judging extractions including *that*-trace violation

With respect to RTs to five types of ungrammatical wh-extractions in the full sentence condition, it is observed that wh-extraction with *that*-trace violation takes the longest RTs within the L2 groups. This was followed by wh-extractions with subject island violation. Overall, accuracy and RT scores in the full sentence condition suggest that L2 are as sensitive as the native English speakers to most island constraints in the L2 English.

In the self-paced word-by-word reading condition, we see similar results. First of all, in grammatical wh-extractions, while the highest accuracy is observed in object extraction from non-finite clauses, the lowest accuracy is obtained in subject extraction from non-finite clauses.

RT results in the word-by-word moving window condition display a similar RT pattern among the three groups. In grammatical wh-dependencies, they seem to spend longer RTs at a region, where the filler is integrated with its potential gaps. For example in subject extractions from finite clauses, all of the groups spend longer RTs at the embedded verb '*shot*' following the main verb '*believe*' as in ("*Who do the police believe shot the editor in the street?"*). This shows that L2 learners have the same processing problem in subject extraction from finite clauses. Also, all groups spend longer RTs at the embedded object NP '*the job applicants'*, following the embedded verb '*meet*' in subject extraction from non-finite clauses as in ("*Who does the manager expect to meet the job applicants today*?"). This shows that all groups experience a filled-gap effect at a region, where there is an overt embedded object NP.

In ungrammatical wh-extractions, the two most problematic sentence types are wh-extraction with *that*-trace violation and wh-extraction with subject island violation, which trigger the highest error rates for all groups. The most correctly judged three ungrammatical sentence types include wh-extractions involving the relative clause island violations, adjunct island and the complex NP island violations.

In ungrammatical wh-extractions, neither native English speakers nor L2 learners spent longer RTs at the potential gap sites inside the islands, suggesting that the two L2 groups are as sensitive as the native speakers in judging wh-extraction from adjunct island, relative clause island and complex NP/DP island in the L2 English. However, in sentences with subject island violations, all groups spend longer RTs on the embedded verb '*amuses*' following' the preposition '*by*' as in (**Who does the teacher believe a story by amuses the children*?). This shows that both the native speakers and L2 learners posit an illicit gap inside the subject island to associate with the wh-filler '*who*', suggesting that they are not sensitive to subject island constraint. In wh-sentences with *that*-trace violation, RTs results show that the locus of the difficulty for both the native speakers (particularly the L2 learners) is the embedded verb '*attacked*' following the complementizer *that* as in (**Who do the police believe that attacked the man in London?*)

These results show that L2 learners are as successful as the native English speakers in accuracy in judging grammatical and ungrammatical wh-sentences, which suggests that they have abstract knowledge of overt-wh-movement and relevant constraints that restrict extraction out of islands. They also reveal that L2 learners are basically similar to the native English speakers in the way they process grammatical and ungrammatical wh-extractions. More specifically, they use similar processing strategies to those of the native speakers in processing wh-dependencies in the L2 English even if they are slower in processing items they can judge correctly.

These results are consistent with the findings of previous studies (i.e., White & Juffs, 1998; Juffs & Harrington, 1995; 1996) in that they provide evidence that adult L2 learners have the knowledge of overt-wh-movement to correctly accept long distance wh-extraction from finite and non-finite clauses, and of the constraints to successfully reject ungrammatical wh-extractions with island constraint violations. The findings also provide evidence for the claim that the difference observed between the native and non-native speakers in accuracy is due to a processing problem that L2 learners experience with certain types of wh-extractions rather than a deficit in L2 competence (grammar). This finding also suggests that the end-state L2 speakers are capable of achieving similar competence to native speakers even if they may take longer to access that competence as observed by their slower processing speed (White & Juffs, 1998).

The third research question examines whether the L1 might explain any potential non-native processing of wh-sentences in the L2 English. More specifically, the study examines whether the L1 Spanish (an overt wh-movement language) will contribute more than the L1 Turkish (a wh-in-situ language) to successful processing of wh-sentences in the L2 English. The study also explores potential positive impact of overt movement through scrambling in Turkish on processing wh-dependencies in English. It is conceivable that the presence of scrambling-induced local and long distance overt movement with certain constraints in Turkish can be equally helpful for Turkish-speaking L2 learners of English in processing of wh-sentences and various island constraints in the L2 English. With this background, we assume that like Spanish learners, Turkish learners should behave similarly to the native English speakers in processing wh-extractions in L2 English thanks to similarities between their L1 and L2 languages.

Results from the accuracy responses to grammatical and ungrammatical wh-questions demonstrate that Turkish learners are significantly more accurate than the Spanish learners in overall accuracy in judging grammatical as well as ungrammatical sentences in full-sentence reading. They are also as accurate as the native English speakers except for wh-extractions with *that*–trace violation. As in the other groups, subject extraction from a finite clause is the most difficult type for Turkish learners.

RTs for grammatical and ungrammatical wh-extractions also display that Turkish learners are similar to the Spanish learners. The two L2 groups display the same accuracy order and RT scores in processing of grammatical and ungrammatical wh-extractions. In certain domains, the Turkish group is found to be more successful than the Spanish group. For example, the Turkish-speaking group is found to be more accurate and faster than the Spanish-speaking group in processing wh-extractions with *that*-trace violation.

In the self-paced word-by-word reading condition, the differences between the two L2 groups in accuracy are smaller, but the difficulty hierarchy is the same. The Turkish and Spanish learners are as accurate as the native English speakers on grammatical wh-extractions (except for subject extraction from finite non-finite clauses), and ungrammatical wh-extractions (except for the type with *that*-trace violation).

In word-by-word reading, we see that the Turkish and Spanish L2 learners display a similar RT pattern to that of the native speakers. As we discussed above, the locus of the difficulty in subject extraction from finite and non-finite clauses is the same in the two groups. In ungrammatical items, the two L2 groups did not spent longer RTs at the critical region inside the islands, suggesting that like native speakers they are sensitive to island constraints.

These results show that Turkish learners are as good as (in some cases better than) the Spanish learners in correctly judging grammatical and ungrammatical wh-dependencies in L2 English in real-time, and that they have a similar processing pattern (i.e., similar accuracy and RT pattern) for grammatical and ungrammatical wh-extractions.

Equally successful results observed in the Spanish and the Turkish groups might be interpreted in different ways. As a first account, we might say that the L1 has no particular role in L2 sentence processing. In other words, irrespective of the presence or absence of overt wh-movement in the L1, end-state L2 learners can have access to constraints on wh-movement in the L2. UG might be implicated in this successful acquisition.

Alternatively, we can assume that UG is operative in the L2 and additionally the presence of overt wh-movement via scrambling sensitizes the Turkish-speaking learners of English to wh-dependencies in the L2 English. Recall that research in L1 Turkish suggests that overt movement via scrambling in Turkish shows reconstruction effect, weak crossover effects and scope freezing as in overt movement in languages like English (see Chapter 2 for more information). Thus, overt movement via scrambling exhibits island constraints on movement in Turkish as in English. This explains why the Turkish group is as successful as the Spanish group. Within this perspective, however, there is, naturally, no way to disentangle the UG role and the L1 role (Belikova & White, 2009).

Before moving to the fourth research question of this study, there is one more issue which requires to be discussed in relation to L1 influence. Recall that the L2 groups are significantly worse than the native speakers at correctly rejecting wh-extractions with *that*-trace violation. The Spanish group is less accurate than the Turkish group on this type. We assume that the low accuracy of the two L2 groups on this sentence type can be accounted for by negative L1 transfer In Spanish, unlike English *tha*t-trace does not cause an ECP violation in subject extractions from embedded clauses. Moreover, the complementizer '*that*' obligatorily precedes a subject trace in an embedded clause without violating the ECP (see Chapter 2 for more information about *that*-trace effect and the ECP violation). As a result, encountering a finite verb after the complementizer in the embedded clause as in (**Who do the police believe that attacked the man last night*?), in L2 English does not surprise the Spanish speakers as much as the native English and the Turkish speakers. This effect can be observed in Spanish participants' faster RTs at the embedded verb. This suggests that the Spanish learners incorrectly accept ungrammatical wh-sentences with *that*-trace violation in the L2 English due to the L1 Spanish.

The last research question that this study examines is the extent to which the WM capacity contributes to differences in processing of grammatical and ungrammatical wh-extractions. I specifically looked at the two critical issues. One of them is whether or not the WM capacity is language-independent as suggested by Osaka and Osaka (1992), and the other one is whether the WM capacity can explain differences among language groups in accuracy and RTs in online wh-processing (i.e., is there a relationship between L2 learners' WM capacity and their parsing performance in online processing of wh-extractions?). Recall that subject extraction from non-finite clauses is the most difficult type for all participants to judge, but it is significantly more difficult for L2 learners than the native speakers. If the difficulty is due to the WM limitations, then the WM measures should correlate with the difference in difficulty between the subject and object extraction from non-finite clauses, as predicted by the 'shared resource account' (e.g., Just & Carpenter, 1992). However, according to the alternative 'separate resource account', the WM measures should not correlate with the difficulty observed in these types since there is a separate WM resource dedicated to syntactic processing in verbal WM resources. We ask the same question for the difficulty in wh-extractions with *that*-trace violation: Does the WM capacity correlate with the difficulty in *that*-trace violation?

We first looked at the correlations among the WM measures in the L1 and the L2. In general, the English AOSPAN scores significantly correlate with the English ARSPAN scores (r=.605, p=.01) and with the L1 ARSPAN scores (r=.744, p=.01). This correlation is higher than the correlations between WM span measures observed in the past, which range from (.49 to.60). Thus we observe a strong relationship between AOSPAN and ARSPAN tasks, suggesting that AOSPAN is as valid as ARSPAN in measuring WM capacity and that both tasks reflect a common construct. This finding is consistent with the findings of Turner and Engle (1989), who showed that they could predict reading ability with a WM span task that does not involve reading of sentences, but solving mathematical operations while trying to remember words. The underlying assumption of Turner and Engle's proposal is that WM capacity is independent of the specific nature of the processing component (e.g., reading of sentences, or solving mathematical operations while trying to remember words) of the span task and that a highly demanding processing component is necessary to engage the processing functions of WM and draw individual differences in task performance. This result also supports the findings of Unsworth et al., (2005), who have recently replicated Turner and Engle's findings revising their Operation Span task.

Is working memory language-independent? Results from the correlations among the WM measures in L1 and L2 showed that the Turkish and Spanish learners' ARSPAN scores in their L1s significantly correlate with their L2 ARSPAN scores ((r=.619, p=.01 and (r=.681, p.01)). These correlations are similar to the correlations observed by Osaka and Osaka (1992) for Japanese. In the Osaka and Osaka's study, the average correlation between Japanese reading span and ESL version reading span test was .84, while the average between the Japanese reading span test and Daneman and Carpenter (1980) reading span was .72. This suggests that the WM capacity is in general language-independent. It also confirms the claim that the higher the span, the more language-independent the reader tends to be.

The Turkish and the Spanish learners' ARSPAN scores also significantly correlate with their AOSPAN scores ((r=.753, p.01) and (r=.657, p.01)). These results show that there is also a strong relationship between the participants' L1 ARSPAN and their AOSPAN scores. This suggests that the ARSPAN in L1 is as good as the AOSPAN in measuring the WM capacity.

Can WM capacity explain the differences between subject and object extraction from non-finite clauses in grammatical wh-extractions and the difficulty in *that*-trace violation in ungrammatical wh-extractions during online processing? To answer this question, I first looked at the correlation between the WM capacity and overall accuracy. The results of this analysis indicate that the WM capacity does not significantly correlate with overall accuracy scores. Only a low correlation is found between the WM measures and ungrammatical wh-extraction (r=.224, p.>.05). In further analyses, it is observed that this is due to a correlation between the WM measures and the accuracy scores for adjunct island violation. These results are replicated in an extreme-group analysis, which shows that the high-span participants are not significantly better than the low-span participant in processing wh-extractions.

More interestingly, no significant correlation is observed between the WM capacity and accuracy or RTs for subject extraction from non-finite clauses, or for wh-extractions with *that*-trace violation. These findings are supported by the results of the grouping analysis, which reveals no significant difference between the high- and low-span participants in accuracy or RTs in these sentence types.

Similar results are obtained for the difference between subject and object extraction from non-finite clauses and for the difficulty in *that*-trace violation in three languages groups. In the analysis of word-by-word RTs in each language group, it is found that RTs on the words, which are the loci of difficulty in subject extraction and *that*-trace violations do not strongly correlate with the WM capacity of the participants. This finding is also observed in the grouping analyses, which show no difference between the high-and low-span participants in terms of the RTs for the critical region in these two types.

These results show that WM capacity cannot account for the difference in difficulty between subject and object extraction from non-finite clauses. Neither can it account for the difference between the native English speakers and L2 learners in the accuracy rate on this type. These results are compatible with the prediction of the separate resource theory of Waters and Caplan (1996) and suggest that the difference in the WM capacity does not help escape from the difficulty in processing subject extraction from non-finite clauses. In other words, having larger WM capacity does not prevent from experiencing the difficulty in this type of sentences. These results are consistent with the findings of Juffs (2004 and 2005), and Marinis et al. (2005), and Felser and Roberts (2007).

If WM capacity does not cause a difference between subject and object extraction from non-finite clauses, what does? The difference can be attributed to the load of linguistic processes themselves such as the assignment of Case and theta roles rather than maintenance and retrieval of wh-filler and integration of the filler with its gap in the WM. The interpretation of the subject extraction is complicated by the application of multiple Case and theta roles to the wh-filler. Recall that there are more changes in theta roles and Case, compared to object extraction from non-finite clauses. In the case of L2 processing, this load is twice as big as the load in L1 processing.

WM capacity is not able to account for the difference among the language groups in accuracy on *that*-trace violation. As we discussed above, the difference between the native speakers and the L2 learners in accuracy on this type can be attributed to the L1 influence. There is a local influence of the L1 Spanish on wh-extractions with *that*-trace in L2 English.

There is no significant correlation with the WM capacity and accuracy or RTs for other types of grammatical and ungrammatical wh-extractions. This can be attributed to the fact that OGJ task used in this task does not impose substantial WM load, therefore, we may not detect a strong relation between these variables. In other words, sentence in the OGJ task is within the limit of the participants' WMC; therefore, they do not require extra WM resources. Recall that overall accuracy rate for grammatical and ungrammatical wh-extractions are 83%. Note however that, unlike native speakers and Turkish learners, the Spanish learners' WM measures significantly correlate with their total accuracy scores (.60), accuracy on grammatical wh-extractions (.55) and accuracy on ungrammatical wh-extractions (.44). These results suggest that the WM capacity has a strong positive relationship with accuracy on grammatical wh-sentences, and a moderate relationship with ungrammatical wh-sentences. In further analyses, we see that among grammatical wh-extraction types, there is a moderate relationship between their WM capacity and accuracy on object extraction from finite clauses (with *that*) and non-finite clauses, along with subject extraction from finite clauses. However, no such relationship is seen between their WM capacity and their accuracy on subject extraction from non-finite clauses. Among the ungrammatical wh-extractions, there is a strong relationship between their WM capacity and their accuracy on adjunct and relative clause island violation, but not their accuracy on *that*-trace violation.

Results form grouping analyses indicate that the high-span Spanish learners do better than the low-span Spanish learners in accuracy on grammatical wh-extractions but not different from each other on accuracy on ungrammatical wh-extractions. No significant difference is found between the high- and low-span Spanish learners in accuracy on subject extraction from non-finite clauses and *that*-trace violation. There is also no significant difference between the high- and low-span groups in terms of RTs for grammatical and ungrammatical wh-sentences.

These findings show that the WM capacity can account for the Spanish speakers' low accuracy on grammatical and ungrammatical wh-extractions to some extent. Recall that the Spanish group has lower accuracy than the other two groups. However, their WM capacity cannot explain their low accuracy on subject extraction from non-finite clauses and *that*-trace violation.

In sum, the results of the WM measures show that there is strong relationship between the ARSPAN and AOSPAN, suggesting that AOSPAN is as valid as ARSPAN in measuring the WM capacity. They also demonstrate that the L1 ARSPAN tasks strongly correlate with the L2 ARSPAN task, which suggests that the WM capacity is a language–independent construct. With respect to the difference in difficulty between the subject and object extraction from non-finite clauses or *that*-trace violation, the results support the separate WM resource account, which assumes that WM capacity is independent of general processing preferences. This may explain why the participants with different WM capacities all experience the same difficulty in these wh-sentence types.

6. 3. Conclusion

In this study, I have tested the Turkish and the Spanish learners on OGJT involving both grammatical and ungrammatical wh-extractions in the L2 English presented in two conditions: the full sentence and the self-paced word-by-word reading conditions. The main aim is to find out whether the Turkish and the Spanish learners are similar to the native English speakers in the way they process wh-dependencies in English. In relation to this question, I have examined the influence of the L1 on online processing of grammatical and ungrammatical wh-extractions in the L2 English. Turkish is a wh-in-situ language and exhibits overt wh-movement via scrambling whereas

Spanish has overt movement as in English. Nevertheless, both languages exhibit island constraints on movement as in English. I have also looked at whether there is a subject-object asymmetry in wh-extractions from finite and non-finite clauses. A final issue that I have examined is whether WM capacity can account for the differences in accuracy and RT for grammatical and ungrammatical wh-extractions.

Findings in both conditions reveal some differences between the L2 learners and native English speakers in terms of accuracy and the RTs in processing grammatical and ungrammatical wh-extractions in L2 English. The difference in accuracy between the native and non-native speakers appears to be mainly due to the lower accuracy of the L2 learners on two sentence types: subject extractions from non-finite clauses in grammatical wh-extractions, and *that*-trace violation in ungrammatical wh-extractions. A subject-object asymmetry observed in all groups in processing non-finite clauses can also be seen as evidence for similar processing pattern in native and non-native data. More specifically, the finding that the native speakers and L2 speakers are equally significantly less accurate and slower in subject extraction from non-finite clauses than object extraction from non-finite clauses suggest that L2 learners pattern with native-speakers in on-line processing of wh-extractions even if they fall behind in terms of processing speed.

Similarly, word-by-word RTs show that the loci of the difficulty in subject extraction from non-finite clauses and in sentences with *that*-trace violation are the same for all groups. This suggests that end-state L2 learners are as sensitive as the native speakers to island constraints on movement.

With respect to L1 influence, no significant difference is found between the Spanish and the Turkish groups. The Turkish group is found to be similar to (in some cases better than) the Spanish learners in correctly accepting grammatical wh-extractions and rejecting ungrammatical wh-extractions. The presence of overt movement via scrambling in Turkish as well as availability of UG in end-state L2 acquisition may be implicated in this success. Nevertheless, the results show that the L1 influence is not always positive as we see in the Spanish participants' processing data of *that*-trace violation, where they are found to accept ungrammatical wh-extractions with *that*-trace violation more than the other groups.

With respect to the WM resources, a strong correlation between the AOSAPN and ARSPAN suggests that AOSPAN is as valid as the ARSPAN in measuring the WM capacity. Also, a strong correlation among the L1 ARSPAN given in Spanish and in Turkish and the ARSPAN given in L2 English show that the WM capacity is a language-independent construct. Nevertheless, the relationship between the WM capacity and the differences in accuracy and RTs does not turn out to be significant, suggesting that the WM capacity cannot account for the differences in accuracy on grammatical and ungrammatical wh-extractions, particularly for the difference between subject and object extraction from non-finite clauses in terms of processing difficulty.

As for possible implications of this study for L2 teaching, we can suggest that some linguistic forms such as subject extractions from nonfinite clauses which are difficult to process in the L2 in real time can be taught using explicit, form-focused instructions. In addition, findings suggest that even advanced L2 speakers continue to make L1 transfer errors. In some cases, L2 speakers can keep on making transfer from their L1 to their L2 (i.e., that-trace violation in L2 English due to L1 Spanish). Providing negative evidence through explicit, form-focused instructions on these structures can raise awareness and contribute to reduce L1 influence.

As a final note, this study is an attempt to explore the complex phenomenon of L2 sentence processing in reference to L1 transfer and WM effects. To my knowledge, this is the first study that provides data for L2 English wh-processing by L1 Turkish speakers. In that sense, I believe that it will make some unique contribution to the field. However, this study is not without limitations. For example, this study does not provide non-island conditions for ungrammatical wh-extractions with island violations; therefore we can only provide suggestive evidence in relation island sensitivity. Also, this study examined L2 learners' knowledge of wh-movement in comprehension data, but it does not provide information about this knowledge in production data.

Further research should look into these issues more closely to identify the intricate factors playing a role in L2 sentence processing.

APPENDIX A

OGJT Experimental items

Grammatical items

Object extraction from finite clauses (OEFF)

1 What do the teachers think the students did at the conference?
2 What does the boy think his friends brought to the party?
3 What does the teacher say the students took from the library?
4 What does the woman believe the secretary sent to the boss?
5 What does the woman think the plumber stole from the garage?
6 Who did the lawyer suppose the judge called to the court?
7 Who did the police believe the terrorists kidnapped last week?
8 Who do the police believe the lawyer shot in the street?
9 Who does the journalist claim the manager gave a bribe last month?
10 Who does the nurse believe the doctor kissed in the room?

Object extraction from finite clauses with *that* (OEFFT)

1 What did the woman believe that her husband bought for her?
2 What does the inspector think that the boy stole from home?
3 What does the journalist believe that the man offered to John?
4 What does the man believe that his wife said to her mother?
5 What does the student say that the professors announced on Friday?
6 Who do the police believe that the man killed in London?
7 Who do the senators think that the president invited to dinner?
8 Who does the nurse believe that the doctor examined on Tuesday?
9 Who does the nurse believe that the man visited last Friday?
10 Who does the teacher believe that the student met at school?

Object extraction from nonfinite clauses (OEFNONF)

1 Who did the girl want to introduce to her family yesterday?
2 Who did the lawyer want to defend in court last week?
3 Who did the police expect to visit in prison yesterday morning?
4 Who did the woman expect to kiss at the party yesterday
5 Who do the parents want to take to school every morning?
6 Who does the boy expect to meet at the party tonight?
7 Who does the man expect to drive to the concert tomorrow?
8 Who does the manager expect to meet at work this morning?
9 Who does the manager expect to see at work tomorrow morning?
10 Who does the professor want to visit in class on Tuesday?

Subject extraction from finite clauses (SEFF)

1 Who do the parents believe kidnapped their baby from the hospital?
2 Who do the police believe shot the editor in the street?
3 Who does the boy think invited the clown to the party?
4 Who does the journalist claim gave a bribe last month?

5 Who does the lawyer suppose attacked the judge in the court?
6 Who does the nurse believe kissed the doctor in the room?
7 Who does the teacher say took the books from the library?
8 Who does the woman believe sent the letters to her boss?
9 Who does the woman say knows the author of the book?
10 Who does the woman think stole the bicycle in the garage?

Subject extraction from nonfinite clauses (SEFNONF)

1 Who did the girls expect to kiss Jennifer at the wedding?
2 Who did the manager want to introduce Jane to his family?
3 Who did the murderer want to defend him in court yesterday?
4 Who did the rector want to drive her daughter to work?
5 Who do the parents want to take their children to school?
6 Who do the police expect to visit the woman in prison?
7 Who does the boy expect to meet Jane at the party?
8 Who does the manager expect to meet the job applicants today?
9 Who does the manager expect to see the lawyer on Friday?
10 Who does the professor want to visit the students in class?

Ungrammatical items

Adjunct Island:

1 What did Ann eat her dinner after she watched on TV?
2 What did Ann stay at home because she hurt on Tuesday?
3 What did John take a shower after he cleaned last night?
4 What did Sam see a doctor because he broke at school?
5 What does John have a drink before he starts on Sundays?
6 Who did Alison go to work after she took to school?
7 Who did Jack leave the restaurant after he met last night?
8 Who did the children return home when they saw at school?
9 Who did the police arrest Sam because he attacked last night?
10 Who did the girl cook dinner because she invited for dinner?

Complex NP Island

1 What does Ann believe the fact that Tom bought in London?
2 What does James believe the fact that Alison saw at work?
3 What does Jane deny the claim that she bought in Paris?
4 Who did Mark hear the rumor that Ann visited in London?
5 Who did Matt hear the rumor that Jack kissed last night?
6 Who does Ann believe the claim that Tom helped at work?
7 Who does Claire know the fact that Jane hates at school?
8 Who does Jane believe the fact that Gerry met at school?
9 Who does John believe the claim that Tom packed at home?
10 Who does Mary believe the claim that Terry saw on TV?

Relative Clause Island

1 What did Ted visit the professor who teaches at the university?
2 What does Jane visit the architect who designed for her friend?
3 What does John trust the man who hires for the company?
4 What does Kate know the child who broke in the market?
5 What does Mark meet the girl who borrowed from the library?
6 Who did Jane see the doctor who examined at the hospital?
7 Who did John meet the lawyer who attacked in court yesterday?
8 Who does Ann admire the teacher who helps with her homework?
9 Who does Matt know the boss who interviewed for the job?
10 Who does Sally hate the lawyer who visited in prison yesterday?

Subject Island

1 What did the teacher think the results of surprised the students?
2 What do the ministers think an article about annoyed the president?
3 What do the parents think a book about interests their children?
4 What does the author believe a movie about pleased some women?
5 Who do the professor believe a letter from worried the student?
6 Who does the editor think the rumors about upset the actress?
7 Who does the journalist believe a gift from pleased the president?
8 Who does the man think the news about shocked his manager?
9 Who does the professor know a speech by motivates the students?
10 Who does the teacher believe a story by amuses the children?

That-trace

1 Who do the children believe that brings nice gifts at Christmas?
2 Who do the police believe that attacked the man last night?
3 Who do the police think that witnessed the accident on Sunday?
4 Who does the girl believe that lives in that old building?
5 Who does the man believe that sends flowers to his wife?
6 Who does the man think that saw the thief last night?
7 Who does the nurse think that told the truth to Jennifer?
8 Who does the student think that saw John in the restaurant?
9 Who does the teacher think that visited the school last week?
10 Who does the woman think that takes the child to school?

APPENDIX B

Sentences used in English ARSPAN

1. Practice sentences

1	Andy was stopped by the policeman because he crossed the yellow heaven.	FALSE
2	During winter you can get a room at the beach for a very low rate.	TRUE
3	People in our town are more giving and cheerful at Christmas time.	TRUE
4	During the week of final spaghetti, I felt like I was losing my mind.	FALSE
5	After final exams are over, we'll be able to take a well-deserved rest.	TRUE
6	After a hard day at the office, Bill often stops at the club to relax.	TRUE
7	No matter how much we talk to him, he is never going to change.	TRUE
8	The prosecutor's dish was lost because it was not based on fact.	FALSE
9	Every now and then I catch myself swimming blankly at the wall.	FALSE
10	We were fifty lawns out at sea before we lost sight of land.	FALSE
11	Throughout the entire ordeal, the hostages never appeared to lose hope.	TRUE
12	Paul is afraid of heights and refuses to fly on a plane.	TRUE
13	The young pencil kept his eyes closed until he was told to look.	FALSE
14	Most people who laugh are concerned about controlling their weight.	FALSE
15	When Lori shops she always looks for the lowest flood.	FALSE

2. Experimental sentences

1	When I get up in the morning, the first thing I do is feed my dog.	TRUE
2	After yelling at the game, I knew I would have a tall voice.	FALSE
3	Mary was asked to stop at the new mall to pick up several items.	TRUE
4	When it is cold, my mother always makes me wear a cap on my head.	TRUE
5	All parents hope their list will grow up to be intelligent.	FALSE
6	When John and Amy moved to Canada, their wish had a huge garage sale.	FALSE
7	In the fall, my gift and I love to work together in the yard.	FALSE
8	At church yesterday morning, Jim's daughter made a terrible plum.	FALSE
9	Unaware of the hunter, the deer wandered into his shotgun range.	TRUE
10	Since it was the last game, it was hard to cope with the loss.	TRUE
11	Because she gets to knife early, Amy usually gets a good parking spot.	FALSE
12	The only furniture Steve had in his first bowl was his waterbed.	FALSE
13	Last year, Mike was given detention for running in the hall.	TRUE
14	The huge clouds covered the morning slide and the rain began to fall.	FALSE
15	After one date I knew that Linda's sister simply was not my type.	TRUE
16	Jason broke his arm when he fell from the tree onto the ground.	TRUE
17	Most people agree that Monday is the worst stick of the week.	FALSE
18	On warm sunny afternoons, I like to walk in the park.	TRUE
19	With intense determination he overcame all obstacles and won the race.	TRUE
20	A person should never be discriminated against based on his race.	TRUE
21	My mother has always told me that it is not polite to shine.	FALSE

22	The lemonade players decided to play two out of three sets.	FALSE
23	Raising children requires a lot of dust and the ability to be firm.	FALSE
24	The gathering crowd turned to look when they heard the gun shot.	TRUE
25	As soon as I get done taking this envy I am going to go home.	FALSE
26	Sue opened her purse and found she did not have any money.	TRUE
27	Jill wanted a garden in her backyard, but the soil was mostly clay.	TRUE
28	Stacey stopped dating the light when she found out he had a wife.	FALSE
29	I told the class that they would get a surprise if they were orange.	FALSE
30	Jim was so tired of studying, he could not read another page.	TRUE
31	Although Joe is sarcastic at times, he can also be very sweet.	TRUE
32	Carol will ask her sneaker how much the flight to Mexico will cost.	FALSE
33	The sugar could not believe he was being offered such a great deal.	FALSE
34	I took my little purple to the ice cream store to get a cone.	FALSE
35	Kristen dropped her parents off at the love for their annual vacation.	FALSE
36	The firefighters sour the kitten that was trapped in the big oak tree.	FALSE
37	Peter and Jack ruined the family carwash when they burned the turkey.	FALSE
38	Martha went to the concert, but ate to bring a thick sweater.	FALSE
39	Sara wanted her mother to read her a window before going to sleep.	FALSE
40	Our dog Sammy likes to greet new people by joyful on them.	FALSE
41	Wendy went to check her mail but all she received were cats.	FALSE
42	Realizing that she was late, Julia rushed to pick up her child from speaker.	FALSE
43	Paul likes to cry long distances in the park near his house.	FALSE
44	The sick boy had to stay home from school because he had a phone.	FALSE
45	The judge gave the boy community sweat for stealing the candy bar.	FALSE
46	Women fall in jump with their infants at first sight or even sooner.	FALSE
47	Jason's family likes to visit him in Atlanta during the cherry every year.	FALSE
48	The doctor told my aunt that she would feel better after getting happy.	FALSE
49	The printer sprinted when he tried to print out his report last night.	FALSE
50	Nick's hockey team won their final game this past weekend at the shoes.	FALSE
51	My mother and father have always wanted to live near the cup.	FALSE
52	The prom was only three days away, but neither girl had a dress yet.	TRUE
53	The children entered in a talent contest to win a trip to Disney World.	TRUE
54	They were worried that all of their luggage would not fit in the car.	TRUE
55	The seventh graders had to build a volcano for their science class.	TRUE
56	The college students went to New York in March and it snowed.	TRUE
57	She had to cancel the appointment because she caught the flu yesterday.	TRUE
58	Doug helped his family dig in their backyard for their new swimming pool.	TRUE
59	The dogs were very excited about going for a walk in the park.	TRUE
60	In the spring, the large birdfeeder outside my window attracts many birds.	TRUE
61	Before Katie left for the city, she took a self-defense class at the gym.	TRUE
62	Mary was excited about her new furniture that she had bought on sale.	TRUE
63	The class did not think the professor's lecture on history was very interesting.	TRUE
64	Jane forgot to bring her umbrella and got wet in the rain.	TRUE
65	Dan walked around the streets posting signs and looking for his lost puppy.	TRUE
66	The couple decided that they wanted to have a picnic in the park.	TRUE
67	The girls were very excited about moving into their new house next week.	TRUE

68	Joseph told his mother that he was probably going to fail sixth grade math.	TRUE
69	We like to eat eggs and bacon for breakfast in the morning.	TRUE
70	Harry plans to play a lot of golf when he retires from his job.	TRUE
71	His stereo was playing so loud that he blew out the speakers.	TRUE
72	It was a clear night, and we could see the stars in the sky.	TRUE
73	At the party, Randy got out the camera to take some pictures.	TRUE
74	Catherine dressed up as a scary witch for the Halloween pencil on Friday.	FALSE
75	Spring is her favorite time of year because flowers begin to bloom.	TRUE
76	Herb rode his bike to the store because his dog had a flat tire.	FALSE
77	Bill worked hard at his new job and hoped to get a promotion.	TRUE
78	The committee could not come to a decision about what to do.	TRUE
79	Heather told me that cat's wings may span up to six feet.	FALSE
80	The bride shopped for her wedding dress while her mother looked for cakes.	TRUE
81	The bear came to our house to fix our pipes after they burst.	FALSE

APPENDIX C

Sentences used in Spanish ARSPAN

1. Practice sentences

1	Andy fue detenido por un policía porque cruzó el cielo amarillo.	FALSE
2	Durante el invierno puedes conseguir una habitación en la playa a un precio muy barato.	TRUE
3	La gente en nuestra ciudad es mas generosa en Navidad	TRUE
4	Durante la semana final de los espaguetti, sentí que iba a perder la cabeza.	FALSE
5	Cuando se acaben los exámenes finales podremos tomarnor un buen merecido descanso.	TRUE
6	Después de un duro día en la oficina, bill a menudo pasa por el club para relajarse.	TRUE
7	No importan cuánto hablemos con él, nunca va a cambiar.	TRUE
8	El fiscal perdió su plato porque no estaba basado en hechos.	FALSE
9	De vez en cuando me sorprendo a mí mismo nadando embobado a la pared.	FALSE
10	Estábamos a cincuenta céspedes mar adentro cuando dejamos de ver la costa.	FALSE
11	Durante la terrible experiencia, los secuestrados nunca perdieron la esperanza.	TRUE
12	Paul tiene miedo a las alturas y se niega a volar en avión.	TRUE
13	El joven lápiz mantuvo sus ojos cerrados hasta que le dijeron que podía mirar.	FALSE
14	La mayoría de la gente que ríe está preocupada con controlar su peso.	FALSE
15	El oso vino a casa a arreglar las tuberías que se habían reventado.	FALSE

2. Experimental sentences

1	Cuando me levanto por las mañanas, lo primero que hago es dar de comer a mi perro.	TRUE
2	Tras gritar en el partido sabía que iba a tener la voz alta.	FALSE
3	Le pidieron a Mary que pasara por el centro comercial a recoger varias cosas.	TRUE
4	Cuando hace frío mi madre siempre me hace ponerme una gorra en la cabeza.	TRUE
5	Todos los padres esperan que sus listados resulten inteligentes.	FALSE
6	Cuando John y Amy se mudaron a Canadá, sus deseos tuvieron una venta ene l rastro enorme.	FALSE
7	En otoño, a mi regalo y a mí nos encanta trabajar justos en el patio.	FALSE
8	Ayer por la mañana en la iglesia, la hija de Jim hizo una ciruela enorme.	FALSE
9	Sin percibir al cazador, el ciervo se puso al alcance de su rifle.	TRUE
10	Como fue el último juego, fue difícil superar la pérdida.	TRUE
11	Como llega al cuchillo temprano, Amy normalmente consigue un espacio para aparcar.	FALSE
12	El único mueble que Steve tenía en su primer cuenco era su cama de agua.	FALSE
13	El año pasado a Mike lo castigaron por correr en el vestíbulo.	TRUE
14	Las nubes cubrían el tobogán de la mañana y empezó a lloverThe huge clouds co.	FALSE
15	Con sólo una cita ya sabía que la hermana de Linda no era mi tipo.	TRUE
16	Jason se rompió un brazo cuando se cayó de un árbol al suelo.	TRUE

17	La mayoría de la gente está e acuerdo en que el Lunes es el peor palo de la semana.	FALSE
18	Me gusta pasear en las cálidas tardes soleadas.	TRUE
19	Con gran determinación superó todos los obtáculos y ganó la carrera.	TRUE
20	Nadie debería ser discriminado por su raza.	TRUE
21	Mi madre siempre me ha dicho que es de mala educación relucir.	FALSE
22	Los jugadores de limonada decidieron jugar dos de los tres sets.	FALSE
23	Críar a niños requiere mucho polvo y la capacidad de ser severo.	FALSE
24	La multitud se volvió a mirar cuando oyó el disparo.	TRUE
25	En cuanto acabe de coger esta envidia me voy a casa.	FALSE
26	Sue abrió su monedero y se dio cuenta de que no tenía dinero.	TRUE
27	Jill quería un jardín en el patio, pero el suelo era más que nada arcilla.	TRUE
28	Stacey dejó de verse con la luz cuando se enteró de que tenía una esposa.	FALSE
29	Le dije a la clase que tendrían una sorpresa si fueran naranjas.	FALSE
30	Jim estaba tan cansado de estudiar que no podía leer otra página más.	TRUE
31	A pesar de que Joe es sacástico a veces, también puede ser muy amable.	TRUE
32	Carol le preguntará a su zapatilla que cuánto cuesta un vuelo a México.	FALSE
33	El azúcar no se podía creer que le ofrecieran tan buen negociol.	FALSE
34	Llevé a mi moradito a la heladería a por un cucurucho.	FALSE
35	Kristen llevó a sus padres al amor para sus vacaciones anuales.	FALSE
36	Los bomberos amargaron al gatito que estaba atrapado en un gran roble.	FALSE
37	Peter y Jack arruinaron el lava-coches familiar cuando quemaron el pavo.	FALSE
38	Marta fue a un concierto pero comió de traer un suéter grueso.	FALSE
39	Sara quería que su madre le leyera una ventana antes de dormirse.	FALSE
40	A nuestro perro Sammy le gusta saludar a la gente vivaracho con ellos.	FALSE
41	Wendy fue a mirar su correo pero sólo había recibido gatos.	FALSE
42	Al darse cuenta de que llegaba tarde, Julia se dio prisa en recoger a su hijo del altavoz.	FALSE
43	A Paul le gusta llorar largas distancias en el parque cerca de su casa.	FALSE
44	El niño enfermo tuvo que volver de la escuela a casa porque tenía un teléfono.	FALSE
45	El juez sentenció al niño a sudor en la comunidad por robar caramelos.	FALSE
46	Las mujeres se saltan de sus hijos a primera vista o incluso antes.	FALSE
47	A la familia de Jason le gusta visitarlo en Atlanta todos los años en cereza.	FALSE
48	El médico le dijo a mi tía que se encontraría mejor depués de tomarse alegría.	FALSE
49	La impresora aceleró cuando intentó imprimir el informe anoche.	FALSE
50	El equipo de hóckey de Nick ganó la final el fin de semana pasado en los zapatos.	FALSE
51	Mis padres siempre han querido vivir derca de la copa.	FALSE
52	Faltaban sólo tres días para el baile pero ninguna chica tenía vestido todavía.	TRUE
53	Los niños participaron en un concurso de talento para ganar un viaje a Disneylandia.	TRUE
54	Estaban preocupados de que todo su equipaje no cupiera en el coche.	TRUE
55	Los niños de séptimo curso tuvieron que construir un volcán para la clase de ciencias.	TRUE
56	Los universitarios fueron a Nueva York en Marzo y nevó.	TRUE
57	Tuvo que cancelar la cita porque cogió la gripe ayer.	TRUE

58	Dough ayudó a su familia a excavar en el patio para la piscina nueval.	TRUE
59	Los perros estaban muy excitados por ir a pasear en el parque.	TRUE
60	En primavera, el nido cerca de mi ventana atrae a muchos pájaros.	TRUE
61	Antes de marcharse a la ciudad, Katie tomó un curso de defensa personal en el gimnasio.	TRUE
62	Mary estaba ilusionada con los muebles nuevos que compró en rebajas.	TRUE
63	Los alumnos no pensaban que la clase de historia fuera muy interesante.	TRUE
64	Jane se olvidó de coger un paraguas y se empapó bajo la lluvia.	TRUE
65	Dan anduvo por las calles poniendo carteles y buscando a su cachorro perdido.	TRUE
66	La pareja decidió que quería hacer un picnic en el parque.	TRUE
67	A las chicas les hacía mucha ilusión mudarse a una casa nueva la semana siguiente.	TRUE
68	Joseph le dijo a su madre que probablemente suspendería las mates.	TRUE
69	Nos gusta comer huevos y bacon para desayunar por las mañanas.	TRUE
70	Harry tiene planeado jugar al golf cuando se jubile de su trabajo.	TRUE
71	Tenía la música tan alta que reventó los altavoces.	TRUE
72	Era una noche rasa y se podían ver la estrellas en el cielo.	TRUE
73	Durante la fiesta, Randy sacó su cámara y tomó varias fotos.	TRUE
74	Catherine se disfrazó de bruja fea para el lápiz de Halloween el Viernes.	FALSE
75	La primavera es su estación favorita porque las flores empiezan a florecer.	TRUE
76	Herb llevó su bici a la tienda porque tenía un perro pinchado.	FALSE
77	Bill trabajó duro en su nuevo trabajo con la esperanza de ascender.	TRUE
78	El comité no se decidía sobre qué hacer	TRUE
79	Heather me dijo que las alas de su gato se abren hasta dos metros.	FALSE
80	La novia fue a comprar su traje de novia mientras que su madre buscaba tartas.	TRUE
81	El oso vino a casa a arreglar las tuberías que se habían reventado.	FALSE

APPENDIX D

Sentences used in Turkish ARSPAN

1. Practice sentences

1	Ali polis tarafından durduruldu çünkü o sarı cenneti geçti.	FALSE
2	Kış süresince çok düşük bir ücrete deniz kıyısında bir oda kiralayabilirsiniz.	TRUE
3	Kasabamızdaki insanlar Noel zamanı daha neşeli ve cömerttirler.	TRUE
4	Yıl sonu makarna haftası boyunca aklımı kaybediyorum sandım.	FALSE
5	Yıl sonu sınavları bittikten sonra, hak ettiğimiz bir tatil yapabileceğiz.	TRUE
6	Ofisteki yoğun bir günden sonra, Murat rahatlamak için sık sık kulübe uğrar.	TRUE
7	Onunla ne kadar konuşursak konuşalım, o asla değişmeyecek.	TRUE
8	Savcının tabağı dün kabul edilmedi çünkü o gerçeklere dayanmıyordu.	FALSE
9	Zaman zaman kendimi boş boş duvarda yüzerken buluyorum.	FALSE
10	Karayı gözden kaybetmeden önce denizde elli çimen açılmıştık.	FALSE
11	Bir ay süresince, rehineler asla umutlarını kaybetmiş görünmediler.	TRUE
12	Efe yüksekten korktuğu için uçağa binmeyi reddediyor.	TRUE
13	Genç tabak açması söylenene kadar gözlerini kapalı tuttu.	FALSE
14	Gülen insanların çoğu kilolarını kontrol etme uğraşı içindeler.	FALSE
15	Ayşe alışveriş yaparken her zaman en ucuz seli arar.	FALSE

2. Experimental sentences

1	Sabah uyandığımda, yaptığım ilk iş köpeğimin karnını doyurmaktır.	TRUE
2	Maçta bağırdıktan sonra, uzun bir sesim olacağını biliyordum.	FALSE
3	Fatma'dan bir kaç şey alması için yeni alışveriş merkezine uğraması istendi.	TRUE
4	Hava soğukken, annem her zaman başıma bir bere takar.	TRUE
5	Bütün anne babalar listelerinin büyüyünce zeki olmasını ümit ederler.	FALSE
6	Mehmet ve Fatma İzmir'e taşındıkları gün, dileklerinin büyük bir eşya satışı vardı.	FALSE
7	Sonbaharda, yeteneğim ve ben bahçede birlikte çalışmayı severiz.	FALSE
8	Dün sabah kilisede, Ali'nin kızı korkunç bir erik yaptı.	FALSE
9	Avcıdan habersiz, yavru geyik onun av menziline girdi.	TRUE
10	Son oyun olduğu için, kaybetmeye tahammül etmek zordu.	TRUE
11	Bıçağa erken geldiği için, Ebru genellikle iyi park yeri bulur.	FALSE
12	Ahmet'in ilk kasesindeki tek mobilya onun su yatağıydı.	FALSE
13	Geçen yıl, Mehmet'e koridorda koştuğu için okul tarafından uyarı cezası verildi.	TRUE
14	Büyük bulutlar sabah kutusunu kapladı ve yağmur yağmaya başladı.	FALSE
15	Tek bir buluşmadan sonra, Zeynep'in kız kardeşinin benim tipim olmadığını anladım.	TRUE
16	Murat dün ağaçtan yere düştüğünde sağ kolunu kırmış.	TRUE
17	Birçok kişi Pazartesi'nin haftanın en kötü sopası olduğu konusunda hemfikirdir.	FALSE
18	Ilık, güneşli öğleden sonraları parkta yürüyüş yapmayı severim.	TRUE

19	Büyük bir kararlılıkla bütün zorlukların üstesinden geldi ve yarışı kazandı.	TRUE
20	Bir kişi ırkı nedeniyle asla ayrımcılığa maruz kalmamalı.	TRUE
21	Annem bana her zaman beyazlamanın hoş olmadığını söyler.	FALSE
22	Bavul oyuncuları üç setten ikisini oynamaya karar verdiler.	FALSE
23	Çocuk büyütmek çok toz ve azimli olma yeteneği gerektirir.	FALSE
24	Toplanan kalabalık silah sesini duyduklarında dönüp arkalarına baktılar.	TRUE
25	Bugün bu kıskançlığı almayı bitirir bitirmez eve gideceğim.	FALSE
26	Filiz cüzdanını açtı ve hiç parası olmadığını fark etti.	TRUE
27	Ayşe evinin arkasında bir bahçe yapmak istedi ama toprak zemin büyük ölçüde balçıktı.	TRUE
28	Sibel bir eşi olduğunu keşfettiği zaman lambayla flört etmeyi bıraktı.	FALSE
29	Sınıfa eğer portakal olurlarsa, sürpriz bir ödül alacaklarını söyledim.	FALSE
30	Ali çalışmaktan o kadar yorulmuştu ki başka bir sayfa daha okuyamadı.	TRUE
31	Mustafa bazı zamanlar alaycı olmasına rağmen, bazen çok tatlı olabiliyor.	TRUE
32	Fatma spor ayakkabısına Meksika'ya uçuşun ne kadara mal olacağını soracak.	FALSE
33	Pirinç kendisine böyle büyük bir teklif yapıldığına inanamadı.	FALSE
34	Küçük morumu külah almak için dondurma dükkanına götürdüm.	FALSE
35	Canan anne ve babasını senelik tatilleri için yüzüğe bıraktı.	FALSE
36	İtfaiyeciler iki duvar arasında sıkışıp kalmış olan kedi yavrusunu ekşitti.	FALSE
37	Ahmet ve Can hindiyi yaktıklarında ailenin anahtarını mahvettiler.	FALSE
38	Zeynep konsere gitti, fakat kalın bir kazak getirmeyi yedi.	FALSE
39	Sibel annesinden ona uyumadan önce bir pencere okumasını istedi.	FALSE
40	Köpeğimiz Sammy insanların üstlerine neşelenerek onları selamlamaktan hoşlanır.	FALSE
41	Ayşe postasını kontrol etmeye gitti ama aldığı tek şey kaplanlardı.	FALSE
42	Ebru geç kaldığını fark edince, çocuğunu kağıttan almak için acele etti.	FALSE
43	Mehmet evinin yanındaki parkta uzun mesafe ağlamayı seviyor.	FALSE
44	Hasta çocuk okula gidemeyip evde kaldı çünkü telefonu vardı.	FALSE
45	Yargıç çocuğa şeker çaldığı için toplum teri verdi.	FALSE
46	Kadınlar bebeklerine ilk görüşte hatta daha önce atlama olurlar.	FALSE
47	Barış'ın ailesi her yıl ağaç süresince onu İstanbul'da ziyaret ederler.	FALSE
48	Doktor teyzeme sıkıcı olduktan sonra kendini daha iyi hissedeceğini söyledi.	FALSE
49	Dün gece raporunu basmaya çalıştığında yazıcı koşmaya başladı.	FALSE
50	Ahmet'in tuttuğu futbol takımı bu geçen hafta ayakkabılardaki final maçını kazandı.	FALSE
51	Annem ve babam her zaman kupaya yakın yaşamak istemişlerdir.	FALSE
52	Baloya sadece üç gün kalmıştı ama henüz iki kızın da elbisesi yoktu.	TRUE
53	Çocuklar Disney World'e bir seyahat kazanmak için yetenek yarışmasına katıldılar.	TRUE
54	Bütün bavulların arabaya sığmayacağından korktular.	TRUE
55	Yedinci sınıflar fen dersi projesi için bir volkan inşa etmek zorunda kaldı.	TRUE
56	Üniversite öğrencileri Mart ayında Ankara'ya gittiklerinde kar yağdı.	TRUE
57	Burak randevusunu iptal etmek zorunda kaldı çünkü dün grip oldu.	TRUE
58	Ali yeni bir yüzme havuzu için ailesinin evin arka tarafında bir çukur kazmasına yardım etti.	TRUE
59	Hasan'ın köpeği o sabah parkta yürüyüşe çıkacağı için yerinde duramıyordu.	TRUE

60	Penceremin önündeki kuşyemliğine baharda bir sürü kuş gelir.	TRUE
61	Pınar şehre taşınmadan önce spor salonunda karate dersleri aldı.	TRUE
62	Esra geçen hafta ucuzluktan aldığı yeni mobilyaları çok beğeniyordu.	TRUE
63	Öğrenciler profesörün verdiği tarih dersinin çok başarılı geçtiğini düşünmüyordu.	TRUE
64	Pınar sabah şemsiyesini almayı unuttu ve yağmurda ıslandı.	TRUE
65	Elif ilanlar asıp kayıp yavru köpeğini arayarak caddelerde dolaştı.	TRUE
66	Yaşlı çift parkta piknik yapmak istediklerine karar verdi.	TRUE
67	Kızlar gelecek hafta yeni evlerine taşınacakları için çok sevinçliydi.	TRUE
68	Tamer annesine muhtemelen altıncı sınıf matematiğinden kalacağını söyledi.	TRUE
69	Her sabah kahvaltıda yumurta ve pastırma yemeyi severiz.	TRUE
70	Kemal işinden emekliye ayrıldıktan sonra bol bol golf oynamayı planlıyor.	TRUE
71	Müzik seti o kadar yüksek sesle çalıyordu ki hoparlörleri patlattı.	TRUE
72	Bulutsuz bir geceydi ve gökyüzünde yıldızları rahatça izleyebildik.	TRUE
73	Mehmet partide bir kaç resim çekmek için fotoğraf makinasını çıkardı.	TRUE
74	Filiz Cuma günkü cadılar kaleminde büyücü kıyafeti giydi.	FALSE
75	Meltem'e göre bahar yılın en güzel zamanıdır çünkü çiçekler açmaya başlar.	TRUE
76	Fuat mağazaya bisikletle girdi çünkü köpeğinin lastiği patlamıştı.	FALSE
77	Suat yeni işinde çok sıkı çalışıyordu ve terfi etmeyi umuyordu.	TRUE
78	Komite ne yapılacağı konusunda henüz bir karara varamadı.	TRUE
79	Ayşe bana kedi kanatlarının altı metre olabileceğini söyledi.	FALSE
80	Annesi evde pasta yaparken, kız gelinlik almak için çarşıya çıktı.	TRUE
81	Martı borular patladıktan sonra onları tamir etmeye evimize geldi.	FALSE

REFERENCES

Abney, S.P. (1989). A computational model of human parsing. *Journal of Psycholinguistic Research, 18,* 129-144.

Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition, 30,* 191-238.

Altmann, G. (1998). Ambiguity in sentence processing. Trends in Cognitive Sciences, 2, 146-152.

Aoshima, S., Philips, C., & Weinberg, A. (2004). Processing filler-gap dependencies in a head-final language. *Journal of Memory and Language, 51,* 23–54

Aoun, J. & Li, Y. –H. A. (1993). *Wh*-elements in situ: Syntax or LF? *Linguistic Inquiry, 24,* 199-238.

Arslan, Z.C. (1999). *Approaches to wh-structures in Turkish*. Unpublished M.A. Thesis. Boğaziçi University

Aygen, G. (2000). T-to-C: Extractable Subject and EPP in Turkish. Proceeding of Western Conference on Linguistics, 12, 65-80. California State University, Fresno.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4,* 417–423.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G.A. Bower (Ed.), The psychology of learning and motivation, Vol. 8, 47-90 *Advances in research and theory*. San Diego, CA: Academic Press.

Baddeley, A. D. & Hitch, G. J. (1994). Developments in the concept of working memory. *Neuropsychology,* 8, 485–93.

Belikova, A. & White, L. (2009). Evidence for the Fundamental Difference Hypothesis or not? *Studies in Second Language Research*, 31, 199-223

Bever, T.G. (1970). The cognitive basis for linguistic structures. In J.R. Hayes (Ed.), *cognition and the development of language* (pp. 279-352). New York, NY: John Wiley.

Bley-Vroman , R. , Felix , S. , & Ioup , G . (1988). The accessibility of Universal Grammar in adult language learning . *Second Language Research, 4,* 1– 32.

Bley-Vroman , R . (1989). What is the logical problem of foreign language learning? In

S. M. Gass & J. Schachter (Eds.), *Linguistic perspectives on second language acquisition* (pp. 42 – 68). New York: Cambridge University Press .

Boeckx , C. (2008). *Bare syntax*. Oxford: Oxford University Press.

Boland, J., Tanenhaus, M.K., Garnsey, S. M., & Carlson, G. N. (1995). Verb argument structure in parsing and interpretation: Evidence from wh-questions. *Journal of Memory and Language,* 34, 774-806.

Bourdages, J. S. (1992). Parsing complex NPs in French. In Helen G., and Michael S. R., (eds.), *Island constraints: Theory, acquisition and processing,* 61–87. Dordrecht, ed.: Kluwer

Brysbaert, M., & Mitchell, D.C. (1996). Modifier attachment in sentence parsing: Evidence from Dutch. *Quarterly Journal of Experimental Psychology, 49A,* 664-695.

Burgess, C. (1991). The interaction of syntactic, semantic, and visual factors in syntactic ambiguity resolution. Unpublished PhD Dissertation, University of Rochester.

Caplan, D., & Waters, G. S. (1995). Aphasic disorders of syntactic comprehension and working memory capacity. *Cognitive Neuropsychology, 12,* 638–649

Caplan, D., & Waters, G. S. (1996). Syntactic processing in sentence under dual-task conditions in aphasic patients aphasic patients. *Language and Cognitive Processes*, 11, 525-551.

Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences 22,* 77–126.

Caplan, D., Waters, G., & Dede, G. (2007). Specialized verbal working memory for language comprehension. In A.R.A. Conway, C. Jarrold, M. J., Kane, A. Miyake, and J.N. Towse (Eds.), *Variation in working memory,* 272-302. New York: Oxford University Press.

Case, R., Kurland, E. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology, 33,* 386–404.

Carnie, A. (2007). *Syntax. A Generative introduction*. Oxford: Blackwell Publishing.

Cebreiros, R. (1996). Resetting bounding nodes in acquiring Spanish. In M. Goodell, & D. Choi, (eds). *Kansas Working Papers in Linguistics*, 21, 1-19. Linguistics Graduate Student Association, University of Kansas.

Chambers, C.G., Tanenhaus, M.K., & Magnuson, J.S. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 687-696.

Chen, E., Gibson, E., & Wolf, F. (2005). Online syntactic storage costs in sentence comprehension. *Journal of Memory and Language*, 52(1), 144–169.

Cheng, L.L. (2009). *Wh*-in-situ, from the 1980s to Now. *Language and Linguistics Compass,* 3 (3). 767–791.

Choe, J. W. (1987). LF movement and pied-piping. *Linguistic Inquiry, 18*, 348-353.

Chomsky, N. (1957). *Syntactic structures*. Mouton, The Hague.

Chomsky, N. (1964). Current issues in linguistic theory. In Fodor, J. A. and J. J. Katz (eds.), *The structure of language: Readings in the philosophy of language*, 50-118. Englewood Cliffs, Prentice Hall

Chomsky, N. (1973). Conditions on transformations. In S. Anderson and P. Kiparsky (Eds.) *A festschrift for Morris Halle,* 86-232. New York: Holt, Rinehart and Winston. New York, NY: North-Holland

Chomsky, N. (1976). Conditions on rules of grammar. *Linguistic Analysis* 2, 303-351.

Chomsky, N. (1981). *Lectures on government and binding.* Dordrecht : Foris .

Chomsky, N. (1986). *Barriers.* Cambridge, MA : MIT Press .

Chomsky, N. (1995). *The minimalist program.* Cambridge, MA : MIT Press .

Chomsky, N. (2000). Minimalist inquiries: the framework. In R. martin, D. Michales, and J. Uriagereka (eds.) *Step by step: essays on minimalist syntax in honor of Howard Lasnik*, 89-155. Cambridge, MA: MIT Press.

Chomsky, N. (2005). *On phases*. Ms. MIT, Cambridge, Mass.

Chomsky, N, & Lasnik, H. (1977). Filters and Control. *Linguistic Inquiry* 8, 425-504.

Chomsky, N., & Lasnik, H. (1993). Principles and parameters theory. Reprinted in N. Chomksy (1995). *The minimalist program*, (pp. 13-127). Cambridge, MA: MIT Press.

Cinque, G. (1990). *Types of A-bar Dependencies*. MIT Press, Cambridge, Mass.

Clahsen, H., &. Featherston, S. (1999). Antecedent priming at trace positions: evidence from German scrambling. *Journal of Psycholinguistic Research*, 28, 415-437.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, *2* 7(1), 3-42.

Clifton, C., & Frazier, L. (1989). Comprehending sentences with long-distance dependencies. In G. Carlson & M. Tanenhaus (Eds.), *Linguistic structure in language processing* (pp. 273–317). Dordrecht: Reidel.

Clifton, C., Speer, S. & Abney, S. (1991). Parsing arguments: Phrase structure and argument structures determinants of initial parsing decisions. *Journal of Memory and Language*, 30, 251-271.

Clifton, C., Traxler, M.J., Mohamed, M.T., Williams, R.S., Morris, R.K., & Rayner, K. (2003). The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language, 49,* 317-334.

Cook, J.V., & Newson, M. (2007). *Chomsky's Universal Grammar: An Introduction*. Oxford: Blackwell Publishing.

Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences*, 7, 547-552.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12, 769-786.

Crain, S., & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological syntax processor. In D.R. Dowty, L. Karttunen, and A.M. Zwicky (eds.), *Natural Language Parsing: Psychological, Computational and Theoretical perspectives*, 320-358. Cambridge, UK: CUP.

Crain, S. and Fodor, J.D. (1985). How can grammars help parsers. In: Dowty, D., Kartunnen, D. and Zwicky, A.M., (eds.), *Natural language parsing: Psycholinguistic, computational, and theoretical perspectives*, 94–128., Cambridge, UK: Cambridge University Press.

Crocker, M.W. (1995). *Computational psycholinguistics: An interdisciplinary approach to the study of language*. Dordrecht: Kluwer.

Cuetos, F. and Mitchell, D.C. (1988) Cross-linguistic differences in parsing: restrictions on the use of the late closure strategy in Spanish. *Cognition* 30, 73–105

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19,* 450–466.

De Vincenzi, M. (1991). *Syntactic parsing strategies in Italian*. Dordrecht: Kluwer Academic Publishers.

De Vincenzi, P. & Job, R. (1993). Some observations on the universality of the late closure strategy. *Journal of Psycholinguistic Research*, 22, 189-206.

Dinctopal, N. (2007). *Syntactic parsing in second language acquisition of English*. Unpublished MA Thesis. Bogazici University

Dussias, P. (2003). Syntactic ambiguity resolution in second language learners: Some effects of bilinguality on L1 and L2 processing strategies. *Studies in Second Language Acquisition, 25*, 529–557.

Dussias, E. P., & Cramer Scaltz, T. R. (2008). Spanish-English L2 speakers' use of subcategorisation bias information in the resolution of temporary ambiguity during second language reading. *Acta Psychologica*, 128, 501-513.

Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory & Cognition, 18,* 972–992.

Fedorenko, E., Gibson, E., & Rohde, D. (2006). The nature of working memory capacity in sentence comprehension: Evidence against domain-specific working memory resources. *Journal of Memory and Language*, 54. 541-553

Fedorenko, E., Gibson, E. & Rohde, D. (2007). The nature of working memory in linguistic, arithmetic and spatial integration processes. *Journal of Memory and Language*, 56, 246-269.

Felser, C., Roberts, L., Gross, R., & Marinis, T. (2003). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics*, *24*, 453–489.

Felser, C., Clahsen, H. & Münte, T. (2003). Storage and integration in the processing of filler-gap dependencies: an ERP study of topicalization and *wh*-movement in German. *Brain and Language* 87, 345–54.

Felser, C., & Roberts, L. (2004). *Plausibility and recovery from garden paths in second language sentence processing*. Poster presented at AMLaP, Aix-en-Provence, September 2004.

Felser, C., & Roberts, L. (2007). Processing wh-dependencies in English as a second language: A cross-modal priming study. *Second Language Research, 23,* 1-28.

Fernández, E. (1999). Processing strategies in second language acquisition: Some preliminary results. In E. Klein & G. Martohardjono (Eds.), *The development of second language grammars: A generative approach* (pp. 217–240). Amsterdam: John Benjamins

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language* 25, 348–368.

Ferreira, F., & Henderson, J.M. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language, 30,* 725-745.

Fiebach, C. J., Schlesewkey, M., & Friederici, A. D. (2002). Separating syntactic memory costs and syntactic integration costs during parsing: the processing of German WH-questions. *Journal of Memory and Language*, *47*, 250-272.

Fodor, J.A. (1983). *Modularity of mind*, MIT Press, Cambridge, MA.

Fodor, J. D. (1989). Empty categories in sentence processing. *Language and Cognitive Processes*, 4, 155-209.

Fodor, J. D. (1999). Triggers for parsing with. In E. Klein & G. Martohardjono (Eds.), *The development of second language grammars: A generative approach*, 373–406. Amsterdam: John Benjamins

Frazier, L. (1985). Modularity and the representational hypothesis. In *Proceedings of NELS* 16. 131-148.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance XII,* 559-589. Hillsdale, NJ: Erlbaum.

Frazier, L. (1995). Constraint satisfaction as a theory of sentence processing. *Journal of Psycholinguistic Research*, 24, 437-468.

Frazier, L. and Clifton Jr., C., (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes* 4, 93–126

Frazier, L. & Clifton, C. (1996). *Construal*. Cambridge, MA: MIT Press.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: The new two-stage parsing model. *Cognition*, *6*, 291-325.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*, 178-210.

Frazier, L. and Flores d'Arcais, G. (1989). Filler-driven parsing: the study of gap-filling in Dutch. *Journal of Memory and Language* 28, 331–44.

Freedman, S.E., & Forster, K. I. (1985). The psychological status of overgenerated sentences. *Cognition* 19, 101-131.

Frenck-Mestre, C., & Pynte, J. (1997). Syntactic ambiguity resolution while reading in second and native languages. *Quarterly Journal of Experimental Psychology, 50A*, 119–148

Frenck-Mestre, C. (2002).An on-line look at sentence processing in the second language. In R. Heredia & J. Altarriba (Eds.), *Bilingual sentence processing*, 217–236. Amsterdam: Elsevier.

Frenck-Mestre, C. (2005). Ambiguities and Anomalies: What can eye movements and event- related potentials reveal about second language sentence processing? In J. F. Kroll & A. M. B. deGroot (Eds.), *Handbook of bilingualism,* 268–281. Oxford: Oxford University Press.

Garnsey, S. M., Tanenhaus, M .K,. & Chapman, R. M. (1989). Evoked potentials and the study of sentence comprehension. *Journal of Psycholinguistic Research, 18,* 51-60.

Garnsey, S. M., Pearlmutter, N.J., Myers, E., & Lotocky, M.A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language, 37,* 58-93.

Gathercole, S. E., & Baddeley, A. D. (1993). *Working memory and language.* Hillsdale: Lawrence Erlbaum.

Gibson, E. (1991). *A Computational theory of linguistic processing: memory limitations and processing breakdown*. Doctoral dissertation, Carnegie Mellon University.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition, 68,*1–76.

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In Miyashita, Y., Marantz, A., & O'Neil, W. (Eds.), *Image, language, brain,* 95–126. Cambridge, MA: MIT Press.

Gibson, E. and Hickok, G. (1993). Sentence processing with empty categories. *Language and Cognitive Processes* 8, 147–161.

Gibson, E., Hickok G., & Schütze, C. T. (1994). Processing empty categories: a parallel approach. *Journal of Psycholinguistic Research, 23*, 381-405.

Gibson, E., & Warren, T. (2004). Reading-time evidence for intermediate linguistic structure in long distance dependencies. *Syntax*, *7*, 55–78.

Gibson, E., Desmet, T., Grodner, D., Watson, D. & Ko,K. (2005). Reading relative clauses in English. *Cognitive Linguistics,* 16, 313-354

Goodluck, H., & Rochemont, M. (1992). Island constraints: An introduction. In Goodluck & Rochement, (Eds.), *Island constraints: theory, acquisition, and processing*, 1-33. Netherlands: Kluwer Academic Publishers.

Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 1411–1423.

Gordon, P. C., Hendrick, R., & Levine, W. H. (2002). Memory load interference in syntactic processing. *Psychological Science*, 13, 425–430.

Gorrell, P. (1993). Evaluating the direct association hypothesis: a reply to Pickering & Barry (1991). *Language and Cognitive Processes*, *8*, 129-146.

Görgülü, E. (2006). Variable wh-words in Turkish. Unpublished M.A. Thesis. Bogaziçi University

Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input. *Cognitive Science*, 29(2), 261–290.

Hagoort, P., Brown, C.M., & Osterhout, L. (1999). The neurocognition of syntactic processing. In C. M. Brown & P. Hagoort (eds.), *The neurocognition of language*, 273-316. Oxford, Oxford University Press.

Hahne, A. (2001).What's different in second-language processing? Evidence from event-related brain potentials. *Journal of Psycholinguistic Research*, *30*, 251–266.

Hahne, A.,&Friederici, A. (2001). Processing a second language: Late learners' comprehension mechanisms as revealed by event-related brain potentials. *Bilingualism: Language and Cognition*, *4*, 123–141.

Haegeman, L. (1994). *Introduction to government and binding theory*. Cambridge, Mass.: Blackwell.

Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skills. *Studies in Second Language Acquisition* 14, 25–38.

Havik, E., Roberts, L., van Hout, R., Schreuder, R., &Haverkort, M. (2009). Processing subject-object ambiguities in the L2: A self-paced reading study with German L2 learners of Dutch. *Language Learning*, 59, 73-112.

Hintzman, D .L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, 93, 411-428.

Hoover, M., & Dwivedi, V. (1998). Syntactic processing by skilled bilinguals. *Language Learning*, 48, 1–29.

Hopp, H. (2006). Syntactic features and reanalysis in near-native processing. *Second Language Research*, 22, 369–397.

Hsiao, F., & Gibson. E. (2003). Processing relative clauses in Chinese. *Cognition* 90, 3–27.

Huang, C.-T.J. (1982). Move *wh* in a language without *wh* movement. *Linguistic Review* 1, 369–416.

Huang, C.-T.J. (1995). Logical form. In Webelhuth, G. (ed.), *Government and binding theory and the minimalist program,* 125–175. Cambridge, Mass: Blackwell Publishers.

İkizoğlu, D. (2007). *Islands in Turkish.* Unpublished B.A. Thesis. Boğaziçi University.

Inoue, A. & Fodor, J.D. (1995). Information-paced parsing of Japanese. In R. Mazuka & N.Nagai (eds.), *Japanese Sentence Processing,* 9-64. Hillsdale: Lawrence Erlbaum Associates.

Jackendoff, R. and Culicover, P.(1971). A reconsideration of dative movement. *Foundations of Language* 7, 392–412

Johnson, J., & Newport, E . (1991). Critical period effects on universal properties of language: The status of subjacency in the acquisition of a second language. *Cognition, 39* , 215 – 258 .

Juffs, A. (1998a). Main verb vs. reduced relative clause ambiguity resolution in second language sentence processing. *Language Learning, 48*, 107–147.

Juffs, A. (1998b). Some effects of first language argument structure and syntax on second language processing. *Second Language Research*, *14*, 406–424.

Juffs, A. (2000). Working memory and L1 influences in ambiguity resolution in L2 English sentence processing. American Association of Applied Linguistics Annual Meeting, Vancouver, BC, March 13, 2000.

Juffs, A. (2004). Representation, processing, and working memory in a second language. *Transactions of the Philological Society*, *102*, 199–225.

Juffs, A. (2005). The influence of first language on the processing of *wh*-movement in English as a second language. *Second Language Research, 21*, 121–151.

Juffs, A., & Harrington, M. (1995). Parsing effects in second language sentence processing: Subject and object asymmetries in wh-extraction. *Studies in Second Language Acquisition, 17,* 483–516.

Juffs, A., & Harrington, M. (1996).Garden path sentences and error data in second language processing research. *Language Learning*, *46,* 286–324.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review,* 99(1), 122-149.

Kaan, E. (1997). *Processing subject-object ambiguities in Dutch*. Doctoral dissertation, University of Groningen.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes, 15, 159-201*

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuo-spatial memory span and reasoning. *Journal of Experimental Psychology: General*, 133, 189-217

Kaplan, R.M. and Zaenen, A. (1988) Long-distance dependencies, constituent structure, and functional uncertainty. In M. Baltin and A. Kroch (eds.), *Alternative conceptions of phrase structure* 17-42. Chicago: Chicago University Press

Kim, A. & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, 52, 205-225.

Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition, 2,* 15-47.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language, 30,* 580–602.

King, J. W., & Kutas, M. (1995). Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience, 7,* 376–395.

Kluender, R., & Kutas, M. (1993). Subjacency as a processing phenomenon. *Language and Cognitive Processes,* 8.573–633.

Kornfilt, J. (1996). NP-movement and restructuring. In R. Freidin (ed.), *Current Issues in Comparative Grammar*, 121-147. Dordrecht: Kluwer

Krashen, S.D. (1987). *Principles and practice in second language acquisition.* Englewood Cliffs, NJ: Prentice-Hall

Kural, M. (1993). Scrambling and Mixed Positions in Turkish. *Proceedings of NELS* 22, 259-272. GLSA Publications.

Kurtzman, H. S., & Crawford, L. F. (1991). Processing parasitic gaps. In Sherer, T *(*ed.), *Proceedings of the 21st Annual Meeting of the North Eastern Linguistics Society,* 217–231. Amherst, MA: LSA Publications

Lasnik, H., & Saito, M. (1992). *Move a, Conditions on its application and output*. Cambridge, Mass: MIT Press.

Lee, M.W. (2004). Another look at the role of empty categories in sentence processing (and grammar). *Journal of Psycholinguistic Research* 33, 51–73.

Lehto, J. (1996). Are executive function tests dependent on working memory capacity? *Quarterly Journal of Experimental Psychology* 49, 29–50.

Lewis, R. L. (1996). Interference in short-term memory: The magical number two (or three) in sentence processing. *Journal of Psycholinguistic Research* 25, 93–115.

Lieberman, M., Aoshima, S., & Philips, C. (2006). Nativelike biases in generation of *wh*-questions by nonnative speakers. *Studies in Second Language Acquisition*, 28, 423-448.

Love, T., and D. Swinney (1996). Co-reference processing and levels of analysis on object-relative constructions: Demonstration of antecedent reactivation with the cross modal priming paradigm. *Journal of Psycholinguistic Research* 25, 5-24.

MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology, 24*, 56–98.

MacDonald, M. (1994). Probabilistic constraints and syntactic ambiguity resolution *Language and Cognitive Processes* 9, 157–202

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101, 676–703.

MacDonald, M.C., & Christiansen, M.H. (2002). Reassessing working memory: comment on Just and Carpenter 1992 and Waters and Caplan 1996.*Psychological Review* 109, 35–54.

Mahajan, A. (1990). The A/A' distinction and movement theory. MIT Dissertation, Cambridge, MA.

Mak, W.M., Vonk, W., & Schriefers, H. (2002). The influence of animacy on relative clause processing. *Journal of Memory and Language, 47,* 50-68.

Manzini, M.R. (1992): *Locality. A theory and some of its empirical consequences*. MIT Press, Cambridge, Mass.

Marantz, A. (2006). The minimalist program. In Webelhuth, G. (ed.), *Government and Binding Theory and the Minimalist Program*, 349-381. Cambridge, Mass: Blackwell Publishers.

Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language sentence processing. *Studies in Second Language Acquisition, 27*, 53–78.

Martohardjono , G. , & Gair , J . (1993). Apparent UG inaccessibility in second language acquisition: Misapplied principles or principled misapplications? In F. Eckman (Ed.), *Confluence: Linguistics, L2 acquisition and speech pathology* (pp. 79 – 103). Amsterdam: Benjamins .

Marslen-Wilson, W.., & Tyler, L.K. (1987). Against modularity. In J. Garfield (Ed.), *Modularity in knowledge representation and natural-language understanding*. Cambridge MA: MIT Press.

McElree, B., & Griffith, T. (1998). Structural and lexical constraints on filling gaps during sentence comprehension: A time-course analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24.432–60.

McKinnon, R., & Osterhout, L. (1996). Constraints on movement phenomena in sentence processing: Evidence from event-related potentials. *Language and Cognitive Processes* 11.495–523.

McKoon, G., Ratcliff, R.& Ward, G. (1994). Testing theories of language comprehension: An investigation on the on-line lexical decision task. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20.1219–28.

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. Journal of Memory and Language, 38, 283–312.

Meral, H.M. (2010). *Resumption A' Chain and implications on clausal architecture*. Unpublished Ph. D. Thesis. Boğaziçi University.

Miller, G.A., Chomsky, N., (1963). Finitary models of language users. In: Luce, R.D., Bush, R.R., Galanter, E. (eds.), *Handbook of mathematical psychology*, 2, 419–491.Wiley, New York.

Mitchell, D.C., Cuetos, F. (1991). The origins of parsing strategies. In C.Smith (Ed.), *Current issues in natural language processing*. Austin: Center for Cognitive Science, University of Texas.

Mitchell, D.C., Cuetos, F., & Corley, M.M.B. (1992). *Statistical versus linguistic determinants of parsing bias: Cross-linguistic evidence*. Paper presented at the 5th Annual CUNY Conference on Sentence Processing. New York.

Mitchell, D. (1994). Sentence parsing. In M. A. Gernsbacher (Ed.), *Handbook of Psycholinguistics*, 375–409. San Diego, CA: Academic Press.

Mitchell, D.C., Cuetos, F., Corley, M.M.B., & Brysbaert, M. (1995). Exposure-based models of human parsing: Evidence for the use of coarse-grained (nonlexical) statistical records. *Journal of Psycholinguistic Research, 24,* 469-488.

Miyake, A. K., Just, M. A., & Carpenter, P. A. (1994). Working memory constraints on the resolution of lexical ambiguity: Maintaining multiple interpretations in neutral contexts. *Journal of Memory and Language*, *33,* 175–202.

Miyake, A., & Shah, P. (1999). *Models of working memory: Mechanisms of active maintenance and executive control.* New York: Cambridge University Press.

Miyamoto, E. T., & Takahashi, S. (2002). Typing mismatch effects in the processing subject *wh*-phrases in Japanese. Paper presented at *The 15th Annual CUNY Conference on Human Sentence Processing*, CUNY Graduate Center.

Montrul, S., Foote, R., & Perpiñán, S. (2008). Knowledge of wh-movement in Spanish L2 learners and heritage speakers. *Selected Proceedings of the 10th Hispanic Linguistics Symposium*, (ed.), Joyce Bruhn de Garavito and Elena Valenzuela, 93-106. Somerville, MA: Cascadilla Proceedings Project.

Müller , G . (2007). *On deriving CED effects from the PIC* . Paper presented at the 30th Annual Colloquium of Generative Linguistics in the Old World (GLOW XXX), University of Tromsø , Norway .

Nakano, Y., Felser, C. and Clahsen, H., (2002). Antecedent priming at trace position in Japanese long-distance scrambling. *Journal of Psycholinguistic Research* 31, 531–571

Nakatani, K., & Gibson, E. (2003). An on-line study of Japanese nesting complexity. Poster presented at the 16th Annual CUNY Conference on Human Sentence Processing, Massachusetts Institute of Technology.

Neville, H., Nicol, J., Barss, A., Forster, K. & Garrett, M. (1991). Syntactically-based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience* 3.151–65.

Nicol, J.L., (1993). Reconsidering reactivation. In: Altermann, G. and Shillcock, R., (Eds.), *Cognitive models of speech processing: The second Sperlonga meeting*, 321–350. Hillsdale, NJ: Lawrence Erlbaum Associates.

Nicol, J. L. and Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research* 18, 5–20.

Nicol, J., Fodor, J. D., & Swinney, D. (1994). Using cross-modal lexical decision tasks to investigate sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 1229-1238.

Nishigauchi, T. (1990). *Quantification in the theory of grammar*. Dordrecht: Kluwer Academic Publishers

Osaka, M. & Osaka, N. (1992). Language independent working memory as measured by Japanese and English reading span tests. *Bulletin of the Psychonomic Society* 30, 287–89.

Özsoy, A. S. (1996). A'-Dependencies in Turkish. In Rona, B. (ed.), *Current issues Turkish linguistics*, 139-158, Ankara: Hitit Yayievi.

Özsoy, A. S. 2009. Turkish as a (non)-wh-in-situ language. InE. A.Csato, G. Ims, J. Parslow, F. Thiesen, E. Türker.(eds). *Turcological Letters to Bernt Brendemoen*, 221-232. Oslo; Novus forlag.

Öztürk, B. (2005). *Case, Referentiality, and Phrase Structure*. Amsterdam and Philadelphia: John Benjamins Publishing Company.

Öztürk, B. (in press). Postverbal constituents in Khalkha Mongolian and Uyghur. Rightward Movement from a Cross-linguistic Perspective, H. Walker and G. Webelhuth (Ed.), Linguistik Aktuell, Amsterdam: John Benjamins.

Papadopoulou, D., & Clahsen, H. (2003). Parsing strategies in L1 and L2 sentence processing: A study of relative clause attachment in Greek. *Studies in Second Language Acquisition, 24*, 501–528.

Pesetsky, D. (1987). *Wh-in-situ*: Movement and unselectable binding. In E. Reauland and A.G.B. ter Meulen (Eds.), *The representation of indefiniteness*. Cambridge: MIT Press.

Petersen, L. & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 58, 193-198.

Philips, C., Kazanina, N., & Abada, S. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research*, 22, 407–428

Pickering, M., & Barry, G. (1991). Sentence processing without empty categories. *Language and Cognitive Processes*, 6, 229–259.

Pickering, M. J. (1993). Direct association and sentence processing: A reply to Gorrell and to Gibson and Hickok. *Language and Cognitive Processes, 8*, 163-196.

Pickering, M. J., Barton, S., & Shillcock, R. (1994). Unbounded dependencies, island constraints and processing complexity. In Charles C. J., Lyn F, and Keith R. (eds.), *Perspectives on sentence processing*, 199–224. London: Lawrence Erlbaum.

Pickering, M. J., & Traxler, M. J. (1998). Plausibility and recovery from garden paths: An eye-tracking study. Journal of Experimental Psychology: Learning, Memory, and Cognition, 24(4), 940–960.

Philips, C., Kazanina, N., & Abada, S. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research*, *22*, 407-428.

Philips, C. (2006). The real-time status of island phenomena. *Language*, 82, 795-823.

Pickering, M.J. and Traxler, M.J. (2003). Evidence against the use of subcategorization frequency in the processing of unbounded dependencies. *Language and Cognitive Processes* 18**,** 469–503.

Pickering, M.J., (1994). Processing local and unbounded dependencies: A unified account. *Journal of Psycholinguistic Research* 23, 323–352

Pickering, M., & van Gompel, R.P.G. (2006).Syntactic parsing. In M. J. Traxler & M, A, Gernsbacher, (eds). *The handbook of psycholinguistics,* 455-503. Amsterdam, Boston: Elsevier Academic Press

Pollard, C., and I. A. Sag (1994). *He ad-driven phrases structure grammar*. Chicago: University of Chicago Press.

Pritchett, B. L. (1988). Garden path phenomena and the grammatical basis of language processing. *Language*, 64, 539-576

Pritchett, B. L. (1991). Subjacency in a principle-based parser. In R. C. Berwick (Ed.), *Principle-based parsing: computation and psycholinguistics*, 301-345. Dordrecht: Kluwer Academic Publishers.

Pritchett, B.L. (1992). *Grammatical competence and parsing performance*. Chicago: University of Chicago Press.

Radò, J. (1999). Some effects of discourse salience on gap-filling. Poster presented at *The 12th Annual CUNY Conference on Human Sentence Processing*

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior*, 22, 358–374.

Rayner, K., & Frazier, L. (1987). Parsing temporarily ambiguous complements. *The Quarterly Journal of Experimental Psychology*, 39, 657-673.

Radford, A. (2000). *Syntactic theory and the structure of English. A Minimalist Approach.* Cambridge, UK: Cambridge University Press.

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

Rizzi, L. (1986). On the status of subject clitics in Romance. In O. Jaeggli, and C. Silva-Corvalán, (eds)., *Studies in Romance syntax*, 391-419. Dordrecht: Foris.

Rizzi, L. (1990). *Relativized minimality*. Cambridge, MA: MIT Press.

Rizzi, L. (1996). Residual verb second and the wh-criterion. In Belletti, A. and L. Rizzi, L. (eds.), *Parameters and functional heads*, 63-90. Oxford: Oxford University Press.

Roberts, R., & Gibson, E. **(**2003). Individual differences in sentence memory. *Journal of Psycholinguistic Research* 31, 573–598.

Roberts, L., Marinis, T., Felser, C., & Clahsen, H. (2007). Antecedent priming at trace positions in children's sentence processing. *Journal of Psycholinguistic Research,* 36.175-188.

Rochon, E., Waters, G. S. & Caplan, D. (1994) Sentence comprehension in patients with Alzheimer's disease. *Brain and Language* 46, 49-329.

Ross, J. (1967). *Constraints on Variables in Syntax*. MIT Dissertation, Cambridge, MA.

Sag, I. A., & Fodor, J. D. (1994). Extraction without traces. In R. Aranovich, W. Byrne, S. Preuss, & M. Senturia (eds.), *Proceedings of the 13th Annual Meeting of the West Coast Conference on Formal Linguistics*, 365–384. Stanford, CA: CSLI Publishing.

Schachter, J. (1989). Testing a proposed universal. In Gass, S. and Schachter, J., editors, *Linguistic perspectives on second language acquisition*. Cambridge: Cambridge University Press, 73–88.

Schachter, J. & Yip, V. (1990). Grammaticality judgments: why does anyone object to subject extraction? *Studies in Second Language Acquisition* 124, 379–92.

Schlesewsky, M., Fanselow, G., Kliegl, R. and Krems, J., (2000). The subject preference in the processing of locally ambiguous wh-questions in German. In: Hemforth, B. and Konieczny, L., (eds.), *German sentence processing*, 65–93, Dordrecht, The Netherlands: Kluwer Academic Publishers.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime.* Reference Guide. Psychology Software Tools, Inc., Pittsburgh.

Sekerina, I.A. (2003). Scrambling and processing: Dependencies, complexity, and constraints. In: Karimi, S., (ed.) *Word order and scrambling,* 301–324. Malden, MA: Blackwell Publishing

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology,* 125, 4-27.

Sharwood-Smith, M. (1993) Input enhancement in instructed SLA: theoretical bases. *Studies in Second Language Acquisition* 15, 165–179.

Spivey-Knowlton, M.J., & Tanenhaus, M.K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 1521-1543.

Spivey-Knowlton, M.J., Tanenhaus, M.K., Eberhard, K.M., & Sedivy, J.C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology, 45,* 447-481.

Steedman, M. (2000). *The syntactic process*. Cambridge, Mass.: MIT Press

Stowe, L. (1986). Parsing *wh*-constructions: evidence for on-line gap location. *Language and Cognitive Processes* 2, 227–46.

Stowe, L.A., Tanenhaus, M.K. and Carlson, G.N., 1991. Filling gaps on-line: Use of lexical and semantic information in sentence processing. *Language and Speech* 34, 319–340

Suñer, M. (1994). V-movement and the licensing of argumental wh-phrases in Spanish. *Natural language and linguistic theory*. 12-2. 335-372.

Sussman, R., & Sedivy, J. (2003). The time-course of processing syntactic dependencies: evidence from eye-movements during spoken *wh*-questions. *Language and Cognitive Processes*18, 143-163.

Swinney, D. & Zurif, E. **(**1995). Syntactic processing in aphasia. *Brain and Language* 50, 225–39.

Torrego, E. (1984). On inversion in Spanish and some of its effects. *Linguistic inquiry*. 15. 103-131.

Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. Science, 268, 1632-1634.

Traxler, M. J., &. Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language*, 35.454–75.

Trueswell, J. C. (1996). The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language*, 35, 566–585.

Trueswell, J.C., Tanenhaus, M.K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology Learning, Memory, and Cognition, 19,* 528-553.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic disambiguation. *Journal of Memory and Language*, 33, 285–318.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28,* 127–154.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods,* 37, 498-505.

Uriagereka, J. (1999). Multiple spell-out . In S. Epstein & N. Hornstein (Eds.), *Working Minimalism* (pp. 251 – 282). Cambridge, MA : MIT Press .

van Gompel, R.P.G., Pickering, M.J., & Traxler, M.J. (2001). Reanalysis in sentence processing: Evidence against current constraint-based and two-stage models. *Journal of Memory and Language, 45,* 225-258.

VanPatten, B. (1996). *Input processing and grammar instruction in second language acquisition*. Norwood, NJ: Ablex

VanPatten, B. (Ed.) (2004). *Processing instruction: theory, research, and commentary.* Mahwah, New Jersey: Lawrence Erlbaum.

Vos, S. H., Gunter, T. C., Schriefers, H., & Friederici, A. D. (2001a). Syntactic parsing and working memory: the effects of syntactic complexity, reading span, and concurrent load. *Language and Cognitive Processes*, 16, 65–103.

Vos, S. & Friederici, A.D. (2003). Intersentential syntactic context effects on comprehension: The role of working memory. *Cognitive Brain Research, 16*, 111-122.

Wanner, E. and Maratsos, M. (1978). An ATN approach to comprehension. In: Halle, M., Bresnan, J. and Miller, G.A., (Eds.). *Linguistic theory and psychological reality*, (pp. 119–161). Cambridge, MA,: MIT Press.

Warren, T., & Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition, 85,* 79–112.

Watanabe, A. (2003). Wh-in-situ languages. In Baltin, M. and Collins, C. (Eds.), *The handbook of contemporary syntactic theory*, 203-223. Oxford: Blackwell Publishing

Waters, G., Caplan, D., & Hildebrandt, N. (1991). On the structure of verbal short-term memory and its functional role in sentence comprehension: Evidence from neuropsychology. *Cognitive Neuropsychology 8*, 81–126

Waters, G., Caplan, D., & Rochon, E. (1995). Processing capacity and sentence comprehension in patients with Alzheimer's disease. *Cognitive Neuropsychology*, 12, 1–30.

Waters, G. S., & Caplan, D. (1996a). The measurement of verbal working memory capacity and its relation to reading comprehension. *Quarterly Journal of Experimental Psychology*, 49, 51-74.

Waters, G. S., Caplan, D. & Rochon, E. (1995) Processing capacity and sentence comprehension in patients with Alzheimer's disease. *Cognitive Neuropsychology*, 12,1–30

Waters, G.S., & Caplan, D. (1996b). The capacity theory of sentence comprehension: Critique of Just and Carpenter (1992). *Psychological Review*, *103,* 761-772.

Waugh, N.C., & Norman, D. A. (1965). Primary memory. *Psychological Review*, 72, 89-104.

Weinberg, A. (1999). A minimalist theory of human sentence processing. In S. Epstein & N. Hornstein (eds.), *Working minimalism*, 282-315. Cambridge, MA: MIT Press.
.

White, L. (1989). *Universal grammar and second language acquisition*. Amsterdam: John Benjamins.

White, L. (1991). Adverb placement in second language acquisition: Some effects of positive and negative evidence in the classroom. *Second Language Research, 7,* 133–161.

White, L. & Juffs, A. (1998). Constraints on *wh*-movement in two different contexts of non-native language acquisition: Competence and processing. In Flynn, S., Martohardjono, G. and O'Neill, W., (eds.), *The generative study of second language acquisition,* 111-129. Hillsdale, NJ: Erlbaum,

Williams, J. N., Möbius, P., & Kim, C. (2001). Native and non-native processing of English *wh*-questions: Parsing strategies and plausibility constraints. *Applied Psycholinguistics, 22,* 509–540.

Williams, N.M. (2006). Incremental interpretation in second language sentence processing. *Bilingualism: Language and Cognition* 9 (1), 71-88.

Yoshida, M., Aoshima, S., & Phillips, C. (2004). Relative clause prediction in Japanese. Paper presented at the 17th annual CUNY Conference on Human Sentence Processing, College Park, Maryland.

Zagar, D., Pynte, J., & Rativeau, S. (1997). Evidence for early closure attachment on first-pass reading times in French. *Quarterly Journal of Experimental Psychology, 50A*, 421–438.

Zagona, K. (2002). *The syntax of Spanish*. Cambridge: Cambridge University Press.