

UNIVERSIDADE DE LISBOA
FACULDADE DE LETRAS



Forward and Backward Anaphora Resolution in European Portuguese and Chinese
Syntactic Properties and Second Language Acquisition

Yi Zheng

Orientadoras: Professora Doutora Maria Gabriela Ardisson Pereira de Matos
 Professora Doutora Ana Maria Lavadinho Madeira
 Doutora Paula Luegi

Tese especialmente elaborada para obtenção do grau de Doutor no ramo de Linguística, na
especialidade de Linguística Geral

2019

UNIVERSIDADE DE LISBOA
FACULDADE DE LETRAS



Forward and Backward Anaphora Resolution in European Portuguese and Chinese
Syntactic Properties and Second Language Acquisition

Yi Zheng

Orientadoras: Professora Doutora Maria Gabriela Ardisson Pereira de Matos
 Professora Doutora Ana Maria Lavadinho Madeira
 Doutora Paula Luegi

Tese especialmente elaborada para obtenção do grau de Doutor no ramo de Linguística, na especialidade de Linguística Geral

Júri:

Presidente: Doutora Ana Maria Martins, Professora Catedrática e Directora da área de Ciências da Linguagem, da Faculdade de Letras da Universidade de Lisboa

Vogais:

- Doutora Ana Maria Barros de Brito, Professora Catedrática, Faculdade de Letras da Universidade do Porto;
- Doutora Maria Fernandes Homem de Sousa Lobo Gonçalves, Professora Associada, Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa;
- Doutora Maria Gabriela Ardisson Pereira de Matos, Professora Associada com Agregação, Faculdade de Letras da Universidade de Lisboa, orientadora;
- Doutora Maria Armanda Martins da Costa, Professora Auxiliar, Faculdade de Letras da Universidade de Lisboa;
- Doutora Nélia Maria Pedro Alexandre, Professora Auxiliar, Faculdade de Letras da Universidade de Lisboa.

Table of Contents

Acknowledgement.....	vii
Abstract	ix
Resumo.....	xi
Chapter 1 Introduction	1
Chapter 2 Syntactic properties concerning anaphoric resolution.....	13
2.1 The Binding Theory	13
2.2 Null subject languages	18
2.3 The Avoid Pronoun Principle	36
2.4 The Aboutness-shift Topic.....	40
2.5 Forward and backward anaphora	44
Chapter 3 Processing of anaphoric pronominals.....	59
3.1 The Accessibility Theory.....	59
3.2 The Position of Antecedent Hypothesis	65
3.3 The Centering Theory	73
3.4 The Advantage of First-mention Hypothesis.....	77
3.5 The active search mechanism.....	80
Chapter 4 Re-analysis of anaphoric resolution of Chinese	87
4.1 Analysis of Pan & Paul 2018.....	87
4.2 Re-analyze Chinese temporal adverbial clauses	95
4.3 Re-analyze Chinese null subjects.....	109
4.4 Chinese anaphoric resolution under topic chain analyses.....	119
4.5 Relations between topic and <i>pro</i>	123
4.6 Reanalyzing Chinese forward and backward anaphora	129
Chapter 5 Theories and experimental studies on L2 acquisition.....	139
5.1 L2 acquisition for syntactic properties	139
5.2 L2 acquisition for Interface properties	141
5.3 Summary of the previous experimental studies	154
Chapter 6 Experiment 1 of the study – off-line task	159

6.1 Research questions	159
6.2 The experimental study	160
6.2.1 Participants	160
6.2.2 Materials and design.....	162
6.2.3 Procedure.....	163
6.3 Analysis of the results	166
6.3.1. Chinese pre-test	166
6.3.2. EP test.....	169
6.3.2.1. Native EP.....	171
6.3.2.2. EP as L2.....	171
6.4 Discussion	173
6.5 Conclusions of the off-line task	177
Chapter 7 Experiment 2 Self-paced reading task	179
7.1 Research questions of the on-line test	179
7.2 The experimental study 2	180
7.2.1 Participants	180
7.2.2 Materials.....	182
7.2.3 Procedure.....	187
7.3 Analysis	188
7.3.1 Off-line results: comprehension question.....	189
7.3.2 Discussion	194
7.3.3 Reaction Times: forward anaphora	196
7.3.3.1 Critical region.....	196
7.3.3.2 Post-Critical region	197
7.3.4 Reaction Times: backward anaphora.....	199
7.3.4.1 Critical region.....	199
7.3.4.2 Post-Critical region	200
7.3.4.3 Post- Post-Critical region	199
7.3.5 Discussions: Reaction Times.....	202
7.3.5.1 Forward anaphora.....	202

7.3.5.2 Backward anaphora	204
7.4 Summary of the online task.....	207
Chapter 8 Conclusions	209
Bibliographic References	217
Appendix 1 Introduction for the off-line questionnaire task.....	227
Appendix 2 Introduction for the self-paced reading task.....	231
Appendix 3 Testing items for the Questionnaire task	232
Appendix 4 Testing items for the self-paced reading task	236

List of Tables

Table 1 Results in percentages of off-line task in Sorace & Filiaci 2006	145
Table 2 Results in percentages for overt pronouns from the off-line task in Madeira et al. 2009.....	148
Table 3 Results in percentages for the selection task in Madeira et al. 2012.....	149
Table 4 Results in percentages for the comprehension task in Madeira et al. 2012.....	151
Table 5 Results in percentages for the picture verification task of Lobo et al. 2017	153
Table 6 Summary of studies on Italian pronouns	155
Table 7 Summary of studies on L2 acquisition of EP pronouns	155
Table 8 Summary of the on-line studies on processing of backward anaphora	156
Table 9 L2 learners' characteristics in Experiment 1	161
Table 10 LMM effects of Chinese pre-test.....	166
Table 11 Descriptive statistics of subject preferences in all conditions and all groups	169
Table 12 LMM effects of Portuguese L1 speakers.....	171
Table 13 LMM results of B1 versus C1	172
Table 14 LMM results of B1 versus EP	173
Table 15 LMM results of C1 versus EP	173
Table 16 L2 learners' characteristics in Experiment 2	182
Table 17 Statistical analysis for native speakers in forward anaphora - answers	191
Table 18 Statistical analysis for L2 learners in forward anaphora - answers.....	191

Table 19 Comparison between native speakers and L2 learners in forward anaphora - answers.....	191
Table 20 Statistical analysis for native speakers in backward anaphora - answers.....	193
Table 21 Statistical analysis for L2 learners in backward anaphora - answers	193
Table 22 Comparison between native speakers and L2 learners in backward anaphora - answers.....	193
Table 23 Statistical analysis for native speakers in forward anaphora – critical region ...	196
Table 24 Statistical analysis for L2 learners in forward anaphora – critical region.....	197
Table 25 Comparison between native speakers and L2 learners in forward anaphora – critical region.....	197
Table 26 Statistical analysis for native speakers in forward anaphora – post critical region.....	198
Table 27 Statistical analysis for L2 speakers in forward anaphora – post critical region	198
Table 28 Comparison between native speakers and L2 learners in forward anaphora – post critical region.....	198
Table 29 Statistical analysis for native speakers in backward anaphora – critical region	199
Table 30 Statistical analysis for L2 learners in backward anaphora – critical region	200
Table 31 Comparison between native speakers and L2 learners in backward anaphora – critical region.....	200
Table 32 Statistical analysis for native speakers in backward anaphora – post critical region.....	200
Table 33 Statistical analysis for L2 learners in backward anaphora – post critical region	201
Table 34 Comparison between native speakers and L2 learners in backward anaphora – post critical region.....	201
Table 35 Statistical analysis for native speakers in backward anaphora – post post critical region	202
Table 36 Statistical analysis for L2 learners in backward anaphora	

– post post critical region	202
Table 37 Comparison between native speakers and L2 learners in backward anaphora	
– post post critical region	202

List of Graphics

Graphic 1 Results of the Chinese pre-test in Experiment 1	167
Graphic 2 General results of the EP test in Experiment 1	170
Graphic 3 Results of answers in forward anaphora in Experiment 2.....	190
Graphic 4 Results of answers in backward anaphora in Experiment 2	192
Graphic 5 Reaction time in forward anaphora – null and overt pronoun.....	197
Graphic 6 Reaction time in backward anaphora – null and overt pronoun.....	199

Abbreviation List

A-topic = Aboutness-shift Topic

ATB = Across-the-Board

BP = Brazilian Portuguese

CLLD = Clitic Left-Dislocation

EP = European Portuguese

EPP = Extended Projection Principle

FEC = Free Empty Category

GB = Government and Binding

GCR = Generalized Control Rule

IH = Interface Hypothesis

L1 = First Language

L2 = Second Language

LF = Logical Form

LMM = Linear Mixed Model

M = Mean

MP = Minimalist Program

PAH = Position of Antecedent Hypothesis

QR = Quantifier Raising

TEP = True Empty Position

SD = Standard Deviation

UG = Universal Grammar

Acknowledgement

During the past six years, I have been dedicated to my PhD program in Linguistics. After all of the arduous but meaningful works, I am finally approaching the final goal. During the process of elaboration of my thesis, many people supported me. Without their selfless and valuable helps, I would not be able to gain the success. Therefore, I would like to thank all of the people who supported me during my time for PhD degree.

First of all, I would like to thank my Supervisor, Professor Gabriela Matos, who has guided my scientific investigation for over 8 years, since my time for Master degree, until my time for PhD degree. Professor Gabriela Matos not only taught me countless knowledges about Linguistics and Scientific investigation, but also guided me on how to plan my academic career. She always thinks for me and has offered all kinds of resources to help me improve my investigation. Without her selfless support, I would not be able to finish my thesis.

Second, I would like to thank my co-Supervisor, Professor Ana Madeira, who also guided my academic career since my time for Master degree. I am grateful for all of the academic knowledge she has taught me during the past few years, especially those about second language acquisition. I would also like to thank her for giving me all kinds of support to help my investigation, especially the precious commentaries for my thesis and publications.

I would like to thank my co-Supervisor, Doctor Paula Luegi, who provided me enormous support for my thesis and publications, especially on the design and analysis for the experimental tests. Thanks for her effort, I have been able to gain a better understanding about academic investigation on the basis of experiments.

I would like to thank my Supervisor and co-Supervisors for their communication with the participants or institutes from which I collected data for the experimental tests of my thesis.

Of course, I would like to thank my parents, Xing Zheng and Xia Hu, who always support my academic career with selfless effort and always encourage me to not give up. Without my parents, I would not be able to finish my PhD course. I would also like to thank my mother for the communication she had made with the participants in China, from which I collected the data for some of my experiments.

I would like to thank all of the professors and investigators from FLUL and CLUL who have

supported and helped me during my study in Lisbon. Especially, I would like to thank Professor Armanda Costa, who helped me to get a better understanding about language processing; and Professor Nélia Alexandre, who gave me permission to apply both of the two experiments of my thesis to the Chinese students who study Portuguese at ICLP-FLUL.

I would like to thank all of the other professors who gave me the permission or helped me organize the application of test, namely Professor António Avelar from ICLP-FLUL, Professors Luís Coelho and Luís Barbeiro from Polytechnic Institute of Leiria and Professor Cristina Martins from University of Coimbra.

Last but not the least, I would like to thank all of the participants who took part in the experimental tasks, as well as those generous teachers who permitted me to spend time during their class. With their priceless support, I am finally able to accomplish the experimental part of my thesis.

Abstract

The purpose of this thesis is twofold: to compare European Portuguese (EP) and Mandarin Chinese, two null subject languages with different properties, regarding the resolution of forward and backward anaphora in temporal adverbial adjunct clauses; and to investigate the acquisition of these properties by Chinese learners who acquire EP as a second language (L2). It is observed that, in Romance null subject languages such as EP, a null subject is preferentially interpreted as referring to a subject antecedent, while an overt subject (pronoun) is preferentially interpreted as referring to a non-subject antecedent. This bias was systematized/described by Carminati 2002 as the Position of Antecedent Hypothesis (PAH).

Chinese does not show the properties predicted by PAH in forward anaphora, as an overt pronoun may easily take a subject antecedent (Yang et al. 1999, Zhao 2012). However, in backward anaphora, Chinese has a similar interpretation as EP, as a null subject is preferentially interpreted as referring to a subject antecedent, while an overt pronoun is preferentially interpreted as referring to a non-subject antecedent. The current thesis uses the theory of topic chain developed by Pu & Pu 2014 to explain the asymmetry between forward and backward anaphora in Chinese and proposes that anaphoric resolution in Chinese is constrained by discourse pragmatic factors, and not merely by syntactic properties.

Considering the similarities and differences between EP and Chinese in their pronominal anaphoric resolution, the current thesis aims to test if the Chinese learners of L2 EP will be influenced by their first language (L1) when interpreting forward and backward anaphora in EP. Experiment 1 of the thesis is an off-line comprehension test, where B1 and C1 level learners were requested to read sentences of EP in forward and backward anaphora structures, with a null or an overt pronoun in the temporal adverbial adjunct. The results show that it is easier for the Chinese learners to master the EP anaphoric resolution in backward anaphora than in forward anaphora, probably because resolution strategies are similar in both languages in backward anaphora, and, therefore, there might have been an effect of L1 on the L2.

Experiment 2 is an on-line self-paced reading task, where B2 and C1 level learners were requested to read segmented sentences of EP, which are similar to those tested in Experiment 1. The results show that the L2 learners' real time processing is influenced by their L1 in forward

anaphora conditions, but fail to explain their processing in backward anaphora conditions.

Key words: European Portuguese, Chinese, Forward and Backward Anaphora Resolution, Language Processing, Second Language Acquisition.

Resumo

A presente tese compara duas línguas de sujeito nulo de natureza diferente, a saber o português europeu (PE) e o chinês mandarim, considerando a resolução anafórica das orações adverbiais temporais em estruturas de anáfora e de catáfora. Esta tese também investiga a aquisição destas propriedades por parte dos aprendentes chineses que adquirem o PE como língua segunda (L2). O segundo capítulo da tese apresenta estudos anteriores sobre o sujeito nulo nas línguas românicas de sujeito nulo (incluindo o PE), que é considerado como *pro* e que está relacionado com a concordância verbal rica. Para as línguas de sujeito nulo sem concordância verbal, como o caso do chinês, Huang 1984 defendeu que o sujeito nulo pode ser tanto *pro*, legitimado e recuperado através da Generalized Control Rule (GCR), como uma variável gerada por movimento-A'. Trabalhos posteriores como o de Liu 2014 defendem que *pro* em chinês também pode ser recuperado através da operação de Concordância (Agree), com base na proposta de “tópico de aboutness-shift” de Frascarelli 2007. Segundo Li 2007, o sujeito nulo do chinês contém um traço uD (*uninterpretable definite*), que terá de ser valorado pela GCR. No capítulo 4 da tese, vou apresentar também as propostas mais recentes de Pan 2016, 2017, que avaliam a questão da dependência-A' do chinês relativamente às teorias de Agree e Match. Procuro mostrar que os sujeitos nulos nas orações subordinadas, sobretudo nas estruturas de ilhas, não podem ser gerados por Agree, uma vez que esta operação está sujeita a condições de ilha. Também não é adequado considerar que estes sujeitos nulos são gerados por Match, uma vez que a intervenção de um outro elemento nominal pode excluir a legitimidade do sujeito nulo. Neste sentido, proponho que *pro* nas orações subordinadas do chinês é legitimado pela GCR e tem de ser controlado pelo elemento nominal mais próximo. Quanto a *pro* nas orações matriz, considero que este pode ser gerado por Agree, em conformidade com as propostas de Liu 2014. Nas línguas Românicas de sujeito nulo, um sujeito nulo é preferencialmente interpretado como referindo-se a um antecedente sujeito, enquanto um sujeito pleno é preferencialmente interpretado como referindo-se a um antecedente não sujeito. Estas propriedades serão discutidas no Capítulo 2 da tese, em que apresento as estruturas sintáticas das condições de anáfora e catáfora, sobretudo as propostas de Canceiro 2016, que mostra que estas interpretações sobre os pronomes não estão exclusivamente sujeitas às condições de ligação,

pois são determinadas por fatores pragmático-discursivos.

Em termos de processamento, estas propriedades foram generalizadas por Carminati 2002 como a Hipótese da Posição de Antecedente (HPA), que se baseia na Teoria de Acessibilidade de Ariel 1990, 2001, que analisa esta questão considerando a estrutura da memória. Estes aspetos serão discutidos no Capítulo 3 da tese, juntamente com outras hipóteses de processamento, incluindo a Hipótese da Carga Informacional, a Teoria da Centralidade, a Hipótese da Vantagem da Primeira Referência e o Mecanismo de Active Search.

Quanto ao chinês, esta língua não mostra as propriedades previstas pela HPA em anáfora, uma vez que um pronome pleno pode ter facilmente um antecedente sujeito (Yang et al. 1999, Zhao 2012). No entanto, em catáfora, o chinês tem uma interpretação semelhante à do PE, dado que um sujeito nulo é preferencialmente interpretado como referindo-se a um antecedente sujeito, enquanto um pronome pleno é preferencialmente interpretado como referindo-se a um antecedente não sujeito. Huang 1982 e Zhao 2014 utilizam a noção de c-command cíclico para explicar a interpretação do chinês em catáfora, enquanto Lust et al. 1996 utilizam a Elevação na Forma Lógica para tal. A presente tese mostra que ambas as análises têm desvantagens.

Desenvolvendo as análises de Pan & Paul 2018 para as orações condicionais do chinês, a presente tese (no Capítulo 4) faz uma análise detalhada das orações adverbiais temporais do chinês e propõe que as predições feitas pelo c-command cíclico são na verdade o resultado de uma ambiguidade de análise da estrutura sintática do adjunto adverbial temporal. Desta forma, a presente tese utiliza a teoria de cadeia de tópico desenvolvida por Pu & Pu 2014 para explicar a assimetria entre a anáfora e a catáfora do chinês.

Ainda no Capítulo 4, a presente tese propõe que no chinês existem dois níveis de tópicos: um nível sintático e um nível abstrato. O tópico sintático ocorre em posição-A' da frase e funciona como um operador sem realização fonética, no sentido de Lobo 1994, 1995. Proponho que no chinês algumas frases ou uma frase por si só podem formar uma cadeia de tópico, cuja cabeça (tópico abstrato) será determinada pelo contexto discursivo. Sempre que a cabeça da cadeia seja estabelecida, todos os operadores (tópico sintático) dentro da cadeia também recebem o valor referencial da cabeça. Consequentemente, o operador pode ter uma operação de Agree com o *pro* da frase e atribuir-lhe o seu valor referencial.

A presente tese propõe que as frases de catáfora do chinês com pronome pleno no adjunto

adverbial temporal contêm duas cadeias de tópico separadas, que não podem ter uma cabeça de cadeia com a mesma referência, uma vez que a cabeça de tópico na oração matriz é um DP, que implica a mudança de tópico e não favorece a interpretação correferencial entre o sujeito encaixado e o sujeito matriz. Para as estruturas de catáfora com sujeito nulo, é adoptada a ideia de Biller-Lappin 1983 de que o sujeito nulo é uma forma convencional para introduzir uma entidade nova. Assim, a correferência entre o sujeito nulo encaixado e o sujeito matriz é aceitável. Desta maneira, a presente tese mostra que a resolução anafórica do chinês é determinada pelos fatores pragmático-discursivos, e não é restringida apenas pelas propriedades sintáticas.

Quanto à aquisição de L2, Sorace & Filiaci 2006 propõem a Hipótese de Interface (HI) que defende que as propriedades puramente sintáticas podem ser adquiridas completamente em L2, enquanto as propriedades relacionadas com a interface entre dois módulos (a sintaxe e o discurso-pragmática neste caso) podem não ser completamente adquiridas. Isto implica que os aprendentes de L2 podem ter dificuldades em adquirir as propriedades relacionadas à resolução anafórica. Estes aspetos serão discutidos no Capítulo 5 da tese.

Considerando as semelhanças e diferenças entre o PE e o chinês na resolução anafórica, a presente tese visa testar se os aprendentes chineses de L2 PE são influenciados pela sua língua materna (L1) quando interpretam anáfora e catáfora em PE.

A Experiência 1 (apresentada no Capítulo 6) desta tese é um questionário off-line de compreensão, em que foi solicitado aos aprendentes do nível B1 e C1 que lessem frases do PE em anáfora e catáfora, com sujeito nulo e pleno no adjunto adverbial temporal. Os resultados mostram que, tanto em anáfora como em catáfora, os falantes nativos do PE (o grupo controlo) preferem um antecedente sujeito para pronomes nulos e um antecedente não sujeito para pronomes plenos. Quanto aos aprendentes, em anáfora, eles têm uma interpretação semelhante à da sua L1, uma vez que preferem o antecedente sujeito para os pronomes encaixados nulos e plenos e não há nenhuma mudança do nível B1 para o nível C1. Entretanto, em catáfora, os aprendentes do B1 mostram uma interpretação indeterminada para pronome pleno, enquanto os aprendentes do C1 mostram uma interpretação alvo, que favorece o antecedente não sujeito. Para os pronomes nulos, os aprendentes de B1 e C1 mostram uma interpretação alvo, uma vez que preferem o antecedente sujeito tanto em anáfora como em catáfora, que é consistente com

o chinês. Os resultados da Experiência 1 indicam que os aprendentes de L2 são influenciados pela sua L1, que tem interpretações distintas para anáfora e catáfora no caso de pronome pleno. Quando a interpretação da sua L1 é semelhante ao PE, verifica-se também um desenvolvimento em direção à interpretação alvo na passagem do nível B1 para o nível C1.

A Experiência 2 (apresentada no Capítulo 7) é uma tarefa on-line de Leitura auto-Monitorada (self-paced reading), em que foi pedido aos aprendentes do nível B2/C1 que lessem frases segmentadas do PE nas estruturas de anáfora e catáfora, com pronome nulo e pleno na oração matriz (no caso de anáfora) ou na oração adverbial temporal (no caso de catáfora). O objectivo desta experiência foi comparar o processamento de anáfora e catáfora e testar se os aprendentes são influenciados pelo Mecanismo de Active Search de Kazanina et al. 2007 no processamento das estruturas de catáfora. Os resultados mostram que os falantes nativos preferem um antecedente sujeito para pronome nulo e pleno em anáfora e um antecedente sujeito para o pronome nulo em catáfora. No entanto, não se verificam efeitos significativos nas condições com pronome pleno em catáfora. Os aprendentes, em anáfora, também preferem o antecedente sujeito tanto para o nulo como para o pleno, o que é consistente com a sua L1. No entanto, em catáfora, não houve resultados significativos nesta experiência. Com efeito, a Experiência 2 só revela que o processamento em tempo real dos aprendentes de L2 é influenciado pela sua L1 nas condições de anáfora, mas não conseguem explicar o seu processamento no caso de catáfora.

Palavras chave: Português Europeu, Chinês, Resolução Anafórica e Catafórica, Processamento da Linguagem, Aquisição da Língua Segunda.

Chapter 1 Introduction

The purpose of the current thesis is to compare two null subject languages, namely European Portuguese (EP) and Mandarin Chinese, in relation to their interpretation of null and overt pronominals in forward and backward anaphora structures, as well as the acquisition of these properties of EP by L2 learners who speak Chinese as their L1.

In this thesis, ‘forward anaphora structures’ refers to sentences in which a referentially dependent expression (null or overt pronoun, in the case of EP) appears after its potential antecedent (see (1a)), while ‘backward anaphora structures’ (also known as ‘cataphora’) refers to sentences in which a referentially dependent expression appears before its potential antecedent (see (1b)).

- (1) a. O João₁ comeu uma maçã quando [-]₁/ele₂ chegou a casa.
‘John₁ ate an apple when [-]₁/he₂ arrived home.’
b. Quando [-]₁/ele₂ chegou a casa, o João₁ comeu uma maçã.
‘When [-]₁/he₂ arrived home, John₁ ate an apple.’

The phenomenon of null subject was firstly observed by Perlmutter 1971, while studies like Taraldsen 1980 considered that this property should be related to rich verbal agreement. Under the Government and Binding (GB) framework, Chomsky 1981 proposed the null subject parameter, which divides languages into null subject languages and non-null subject languages. In addition, Rizzi 1982 considered that languages with rich verbal agreement have [+pronominal] and [+referential] features, which give rise to referential null subjects in finite clauses. The null subject in finite clauses was also considered to be *pro* (Chomsky 1982), which bears [-anaphoric] and [+pronominal] features¹. In Rizzi 1986, the author also claimed that the null subject is legitimated by rich verbal agreement and identified by the phi-feature contained in the agreement. Under these theories, EP belongs to the group of null subject languages.

However, Huang 1984 reported that not all of the null subject languages have a rich verbal agreement system, since Chinese, which was considered as a topic-oriented language by Li & Thompson 1976, also allows null subjects, as well as null objects, even if it does not have a

¹ Chomsky 1981 firstly considered the null subject in finite clause as *PRO*. Due to Rizzi 1982’s proposal, Chomsky 1982 then considered such null subject as *pro*.

verbal agreement system. Huang 1984 proposed that the Chinese null subject in subordinate clauses could also be considered as *pro*, which was legitimated by the Generalized Control Rule (GCR) (2).

- (2) An empty pronoun is coindexed with the closest nominal.
(suggested by Huang 1984: 552 and reformulated by Huang, Li & Li 2009: 209)

For null subject in matrix clause and null objects, Huang 1984 proposed that such empty categories are variables generated by A'-movement.

Coming into the Minimalist Program (MP), some theories concerning null subjects also underwent some changes. Roberts & Holmberg 2010 divided null subject languages into three groups, the consistent null subject languages, the partial null subject languages and the discourse null subject languages. Under this proposal, EP, as well as other Romance null subject languages, is considered as a consistent null subject language, since it has a rich verbal agreement system, while Chinese is assumed to be a discourse null subject language, since it does not have a verbal agreement system and its null subjects are highly related to discursive information.

Adopting the feature valuing theory of Chomsky 2001, Holmberg 2010 proposed that *pro* in consistent null subject languages is in fact the result of incorporation of a null pronoun into T, while its reference value should be recovered through the aboutness-shift topic (Frascarelli 2007). Barbosa 2009 also proposed an incorporation analysis for EP, but did not apply to the aboutness-shift topic.

Theories concerning Chinese null subjects and objects also changed during the MP era, in which those Eastern Asian null subject languages without verbal agreement are considered as radical *pro*-drop languages. Studies like Tomioka 2003 proved that the null arguments in radical *pro*-drop languages may turn out to be a deleted NP anaphora, a proposal which was also adopted by Barbosa 2013 to analyze Brazilian Portuguese (BP), a partial null subject language.

However, studies like Tomioka 2003 were mainly based on analyses of Japanese, which present some differences in respect to Chinese, according to Li 2007. Li 2007 also proposed that the Chinese null subject is *pro*, which bears an uD feature² that must be checked through the GCR.

² Though Li 2007 did not clarify this, the uD feature should be interpreted as uninterpretable Definite feature,

The null object can be analyzed as generated by A'-movement, or simply as a True Empty Position (TEP).

There are other studies that proposed the agreement analysis for Chinese null subjects, namely Zhao 2012 and Liu 2014. For example, Liu 2014 used a similar analysis to the one of Frascarelli 2007 to propose that there is an agreement relation between *pro* and topic in Chinese.

Concerning the discursive use of null and overt pronominals in null subject languages, it has been reported that these two categories may not be used in free alternation. For example, in EP, Costa et al. 1998, 1999 already claimed that, in subordinate clauses, a null subject tends to refer to an antecedent in subject position, while an overt pronoun tends to refer to a non-subject antecedent. The same observation is also found for other Romance null subject languages, for example, for Italian by Calabrese 1986 and for Spanish by Luján 1986, among others. Based on the Accessibility Theory of Ariel 1990, 2001, Carminati 2002 proposed the Position of Antecedent Hypothesis (PAH), which argues that, in Italian, the antecedent of a null subject is preferentially in a subject position, while the antecedent of an overt pronoun is preferentially in other positions. This hypothesis has been proved to be true in other consistent null subject languages, though some other factors may also influence the interpretation of null and overt pronominals. For example, the results of Luegi 2012 show that structural position (which is related to the syntactic function/position) and order of reference both contribute to the interpretation of null and overt pronouns in EP.

There are also some studies that compare forward and backward anaphora structures. For example, for Italian, Sorace & Filiaci 2006 and Serratrice 2007 tested the interpretation of null and overt pronouns in both forward and backward anaphora structures. The results showed that the PAH still holds in backward anaphora, though there are some differences between forward and backward anaphora, as in the latter case the participants are more likely to accept an extra-linguistic antecedent for an overt pronoun.

Lobo & Silva 2016, 2017 also made such a comparison for EP. However, in their experimental tests, the native speakers of EP accepted a subject antecedent for the overt pronoun more easily in the backward anaphora than in the forward anaphora condition.

since it is linked to the referential value of the null category.

The interpretation of null and overt pronouns in Chinese is a little bit more complicated than in Romance null subject languages, since there is an apparent difference between forward and backward anaphora when analyzing the interpretation of overt pronouns. In both forward and backward anaphora, Chinese does not differ from Romance null subject languages in what concerns the interpretation of the null subject, since it is always preferential to refer to a subject antecedent. However, in forward anaphora, as reported by Zhao 2012 and Yang et al. 1999, Chinese, unlike the Romance null subject languages, may easily accept the coreferential reading between an overt pronoun and a subject antecedent, which goes against the PAH. Simpson et al. 2016 even found that in some contexts the Chinese overt pronoun is preferentially interpreted as referring to a subject antecedent. However, in backward anaphora, Chinese is again consistent with the Romance null subject languages, as an overt pronoun does not preferentially take the matrix subject as its antecedent.

Several studies have been conducted to explain the difference between Chinese forward and backward anaphora. The first approach was made by Huang 1982, which is followed by Zhao 2014. In this approach, Huang 1982 proposed the notion of cyclic c-command, which states that in the backward anaphora structure of Chinese, the embedded subject may cyclically c-command the matrix subject, and thus their coreference violates the Principle C of the Binding Theory, see example (3):

- (3) [CP [TP Ta₁ chi wanfan] [C de shihou]],
 he eat dinner DE when
 Zhangsan₂ dai zhe yi ding maozi.
 Zhangsan wear Prg one CL hat
 ‘When he was having dinner, Zhangsan was wearing a hat.’ (Zhao 2014: 387)

The second approach was made by Lust et al. 1996, who argued that the overt pronoun *ta* in Chinese is in fact located in a SpecNP position [NP *ta* [N[-]], while the null subject has in fact a null NP structure [NP [-] [N [-]]]. The authors then argued that the overt pronoun has to undergo quantifier raising in Logical Form (LF), while the covert one does not need to do so. After the overt pronoun in the backward anaphora structure undergoes quantifier raising, it cannot share the same index with the matrix subject because of scope limitation. Since the null subject does not undergo quantifier raising, the coreferential reading is still available.

As will be shown in Chapter 2 of the current thesis, both analyses have shortcomings. As a result, the current thesis proposes an alternative analysis to explain the asymmetry between forward and backward anaphora of Chinese. This is the first objective of the thesis.

Elaborating on the analysis for conditional clauses in Chinese made by Pan & Paul 2018, the current thesis proposes that there is some parsing ambiguity for backward anaphora structures in Chinese, which gives rise to the apparent cyclic c-command analysis. In fact, the Chinese temporal adverbial clause has two different structures, one with preposition and the other without preposition. Zhao 2014 and Lust et al. 1996 only analyzed the structure without preposition, which may easily cause parsing ambiguities³. The current thesis argues that the apparent embedded subject in (3) can be analyzed as a matrix topic, which cannot share the same index with the matrix subject. This explains why many native speakers of Chinese do not accept the coreferential reading in (3), without appealing to the analysis of cyclic c-command or quantifier raising.

The analysis of the previous paragraph seems to indicate that the interpretation of the overt pronoun in backward anaphora should not be constrained by syntactic properties, but only because of some parsing ambiguities. It should be noticed that such parsing ambiguities do not exist in the adverbial clauses with preposition. In order to test if there is a difference between structures with and without preposition, the current thesis applied a pilot test⁴ to investigate the issue. The results have shown that the native speakers of Chinese also accept coreference between the embedded overt pronoun and the matrix subject in backward anaphora, though the rate is only around 25%. However, if the disjoint interpretation is constrained by syntax, the acceptance rate should not reach 25%. There is also no significant difference between conditions with and without preposition. Such results may indicate that the interpretation of backward anaphora should be constrained by discourse-pragmatic factors.

In fact, many studies (Xu 2003, Pu & Pu 2014) have analyzed Chinese anaphora relations under the theory of topic chain. The basic idea is that, in Chinese, several sentences or clauses may form a topic chain, which can be headed by a DP or an overt pronoun. The null subject inside a

³ Here, ‘parsing ambiguity’ refers to the cases where the syntactic structure of a sentence can be analyzed in more than one way.

⁴ This preliminary test is not part of the experimental test of the thesis. In the remainder of the thesis, it will be known as ‘pilot Chinese test’. In section 2.5 there will be a more detailed description of this pilot Chinese test.

topic chain may refer to the head of the chain, which indicates that the legitimation and identification of a null subject is based on the possibility to form a topic chain. In addition, an overt pronoun may also serve as the head of a topic chain, which can inherit the referential value from the previous topic chain.

The advantage of the topic chain analysis resides in the fact that it does not rely on the syntactic position of the antecedent, which fails to explain the interpretation of overt pronouns in forward anaphora of Chinese. Since Chinese is a language closely related to discourse and topic (Tsao 1977, Li & Thompson 1976, etc.), the current thesis adopts the topic chain analysis to analyze the interpretative difference between forward and backward anaphora in this language.

Nevertheless, to adopt the topic chain analysis, one question must be properly resolved: how will the null subject be legitimated and identified by the topic head? In order to resolve this problem, the current thesis proposes an eclectic analysis to combine the analyses of Li 2007, Liu 2014 and Pan 2016, 2017, which will be described in detail in Chapter 4.

In summary, the theoretical part of the thesis proves that the Chinese interpretative preference for forward and backward anaphora is also constrained by discourse-pragmatic factors, though it may be different from that found in Romance null subject languages. Then it is interesting to investigate how this interpretative preference influences the L2 acquisition of EP by learners who speak Chinese as their L1. In this is the second objective of the thesis.

The acquisition of the interpretative biases of null and overt pronouns is an issue analyzed by many studies in the past years. Studies such as Montrul & Rodríguez Louro 2006 and Madeira et al. 2012 have proven that L2 learners usually do not have problems acquiring the use or interpretation of the null subject, which is favored to refer to a subject antecedent⁵. However, many studies such as Sorace & Filiaci 2006 and Rothman 2008, 2009 have found that L2 learners may have difficulty in acquiring the use and interpretation of an overt pronoun, since they tend to establish a coreferential relation between the overt pronoun and a subject antecedent, or show indeterminate judgement. These findings corroborate the Interface Hypothesis (IH) of Sorace & Filiaci 2006, who argue that only purely syntactic properties can be completely acquired by L2 learners, while properties which involve the interface between

⁵ On the other hand, studies like Keating et al. 2011 and Rothman 2009 found that L2 learners or near-native speakers of Spanish may also display a non-target interpretation of null subjects.

different modules (for example, interface between syntax and discourse-pragmatics) may not be completely acquired. Some studies found that the L1 of the learners also influence their L2 acquisition, for example, Madeira et al. 2009 and Lobo et al. 2017 have shown that L2 learners whose L1 is a null subject language showed quicker development towards the target interpretation than learners whose L1 is a non-null subject language⁶, though the interpretations of both L2 groups are similar in elementary levels.

Based on these previous findings on L2 acquisition, it is interesting to investigate how L1 Chinese will influence their L2 interpretation. If the IH is applied to all of the L2 learners, then the Chinese L2 learners of EP should show some difficulties in acquiring anaphoric resolution, regardless of the similarities and differences between the two languages. However, just as Madeira et al. 2009 and Lobo et al. 2017 have shown, L2 learners may show faster development if their L1 is similar to the L2 in the relevant properties. Now we see the case of Chinese: though Chinese is a null subject language, its interpretation of overt pronouns in forward anaphora is different from EP. However, Chinese is consistent with EP with respect to the interpretation of overt pronouns in backward anaphora. Will Chinese learners of EP show difficulties in acquiring the interpretative preference in this case, just as other L2 learners of Romance null subject languages do? Or will the similarity between Chinese and EP in the interpretation of backward anaphora make it easier for the L2 learners to master the interpretative preference of the target language? Another factor that should be taken into consideration is the level of the L2 learners. If L1 Chinese indeed helps their L2 acquisition, will the help already occur at a lower level, or will it only appear at a later phase of acquisition?

Moreover, in backward anaphora, since the potential antecedent for the null or overt pronoun may appear afterwards, the processing costs also increase in this case, which led some researchers (Kazanina et al. 2007) to propose the language universal active search mechanism. According to this proposal, the parser should find the antecedent of a backward anaphoric expression as soon as possible, which favors the subject antecedent interpretation. Will this factor also influence the interpretation of the L2 learners?

⁶ It should be noticed that some studies, such as Margaza & Bel 2006, which investigated Greek learners of L2 Spanish, have reported that L2 learners may still have difficulties in mastering the interpretation of overt pronouns, even if their L1 is similar to the target language.

In order to answer the previous questions, two experiments were conducted in the current thesis to analyze the acquisition of forward and backward anaphora in EP, by Chinese-speaking L2 learners.

The first task is an off-line questionnaire study, which includes three groups: a control group with native speakers of EP, two groups of Chinese L2 learners of EP, one from level B1, the other from level C1. The objective of this task is to find how will L1 Chinese will influence the learners' L2 EP at different proficiency levels.

The off-line task was designed to test the following four conditions, exemplified in (4). As shown in the examples, the current thesis only tested temporal adverbial adjunct clauses. For forward anaphora, the tested structures have a right-dislocated integrated adverbial adjunct, while for backward anaphora, there is a left-dislocated adverbial adjunct clause. The participants were asked to choose the interpretation of the embedded null or overt pronoun, between the matrix subject and an extra-linguistic entity.

(4) Condition 1: (Forward overt anaphora)

A Maria apagou a luz enquanto ela comia o bolo.

'Mary turned off the light while she was eating the cake.'

Condition 2: (Forward null anaphora)

A Maria apagou a luz enquanto [-] comia o bolo.

'Mary turned off the light while [-] was eating the cake.'

Condition 3: (Backward overt anaphora)

Enquanto ela comia o bolo, a Maria apagou a luz.

'While she was eating the cake, Mary turned off the light.'

Condition 4: (Backward null anaphora)

Enquanto [-] comia o bolo, a Maria apagou a luz.

'While [-] was eating the cake, Mary turned off the light.'

The questionnaire task demonstrates that the L2 learners are indeed influenced by their L1, since they show different interpretative preferences for the overt pronoun in forward and backward anaphora. Based on these results, the second experiment, an on-line self-paced reading task investigates how the L2 learners in B2 and C1 levels process forward and backward anaphora in real time, considering the influence of the active search mechanism. Kazanina & Philips 2010 found in Russian that the native speakers may also apply to the active search mechanism for the *poka* structures, where the coreference between the embedded pronoun and

the matrix subject is not possible in backward anaphora. So it is interesting to test if the L1 and/or L2 EP speakers will apply to the active search mechanism when processing backward anaphora in EP, where the coreference between embedded overt pronoun and the matrix subject is not the favored interpretation.

In Experiment 2, a self-paced reading task, eight conditions were tested, four for forward anaphora and four for backward anaphora, which are exemplified in (5). Unlike the off-line test, all of the testing sentences in this task have a left-dislocated temporal adverbial adjunct clause. The participants were also asked to answer a question about the sentence, while their reaction time for each region was registered as well. Gender was used as to force a possible or an impossible interpretation between anaphor and a previous or a subsequent subject (forward or backward conditions).

(5) For forward anaphora:

Condition 1 Null-Match

Quando/a Rita/contou/a história,/ficou relaxada/com a reação/dos amigos.

When/the Rita/told/the story,/became relaxed.fem.sg/with the reaction/from+the friends

Condition 2 Overt-Match

Quando/a Rita/contou/a história,/ela ficou relaxada/com a reação/dos amigos.

When/the Rita/told/the story,/she became relaxed.fem.sg/with the reaction/from+the friends

Condition 3 Null-Mismatch

Quando/a Rita/contou/a história,/ficou relaxado/com a reação/dos amigos.

When/the Rita/told/the story,/became relaxed.masc.sg/with the reaction/from+the friends

Condition 4 Overt-Mismatch

Quando/a Rita/contou/a história,/ele ficou relaxado/com a reação/dos amigos.

When/the Rita/told/the story,/he became relaxed.masc.sg/with the reaction/from+the friends

‘When Rita told the story, (he/she) became relaxed with the reaction from the friends.’

For backward anaphora:

Condition 1 Null-Match

Enquanto/estava/virada/para o restaurante,/a Soraia/apagou/o cigarro/tranquilamente.

While/was/turned.fem.sg/to the restaurant,/the Soraia/put out/the cigarette/quietly.

Condition 2 Overt-Match

Enquanto/ela estava/virada/para o restaurante,/a Soraia/apagou/o cigarro/tranquilamente.
While/she was/turned.fem.sg/to the restaurant,/the Soraia/put out/the cigarette/quietly.

Condition 3 Null-Mismatch

Enquanto/estava/virado/para o restaurante,/a Soraia/apagou/o cigarro/tranquilamente.
While/was/turned.masc.sg/to the restaurant,/the Soraia/put out/the cigarette/quietly.

Condition 4 Overt-Mismatch

Enquanto/ele estava/virado/para o restaurante,/a Soraia/apagou/o cigarro/tranquilamente.
While/he was/turned.masc.sg/to the restaurant,/the Soraia/put out/the cigarette/quietly.

‘While (he/she) was turned to the restaurant, Soraia put out the cigarette, quietly.’

The results of the online task showed that the L2 learners have a similar interpretation in forward anaphora as in the off-line task. However, the results obtained in this task are not conclusive regarding their processing strategy in backward anaphora.

The current thesis is organized in the following way:

In Chapter 2, I will make a detailed review of the previous proposals for null subject languages including both EP and Chinese, under both the GB (e.g. Chomsky 1981, 1982, Rizzi 1982, 1986, Huang 1984) and the MP (Roberts 2010, Holmberg 2010, Roberts & Holmberg 2010, Li 2007, Liu 2014). I will also present the Avoid Pronoun Principle and the properties related to this principle, which may explain the different use of null and overt pronouns. Chapter 3 reviews the previous proposals about the interpretation of null and overt pronouns considering language processing mechanisms, including the Accessibility Theory of Ariel 1990, 2001, the PAH of Carminati 2002, the Centering Theory of Grosz et al. 1995, the Advantage of First-mention Hypothesis of Gernsbacher & Hageavers 1988 and Gernsbacher 1989 and the active search mechanism of Kazanina et al. 2007. Chapter 4 proposes an advanced analysis for forward and backward anaphora structures in Chinese, based on those of Pan & Paul 2018. In this chapter, I will present the topic chain analysis for Chinese proposed by Pu & Pu 2014 and make a proposal to connect this theory to those of Li 2007, Liu 2014 and Pan 2016, 2017. Then I use the new proposal to explain why Chinese shows different interpretations between forward and backward anaphora regarding the antecedent of the overt pronoun. Chapter 5 makes a brief review of the literature about L2 acquisition. I will present those proposals for acquisition of purely syntactic

properties, as well as those which involve interfaces, under the IH of Sorace & Filiaci 2006.

From Chapter 6, I will present the experimental tests of the current thesis. Chapter 6 presents the off-line questionnaire, while Chapter 7 presents the on-line self-paced reading task. Finally, Chapter 8 concludes the entire thesis and proposes the investigation questions for future analyses.

Chapter 2 Syntactic properties concerning anaphoric resolution

In this chapter, I will review some of the literature on anaphoric resolution. Section 2.1 presents the basic principles of the Binding Theory. In sections 2.2 to 2.4, I will present the null subject parameter and the related assumptions, namely the Avoid Pronoun Principle and the Aboutness-shift Topic. Section 2.5 discusses the syntactic structures of forward and backward anaphora in EP and Chinese.

2.1 The Binding Theory

The present study aims to analyze some specific cases related to anaphoric and coreferential properties of EP and Chinese. It should be noticed that the term ‘anaphora’ may have a narrow and a broad sense meaning. The narrow sense of anaphora is only related to reflexive expressions which are defined in the Principle A in the classic version of the Binding theory⁷ (see examples below, in (3) and (4)). The broad sense of anaphora refers to any expressions which establish a coreferential relation with a constituent in the previous or upcoming contexts. The present study will assume the broad sense of anaphora in most of the theoretical review and the experimental section, though some properties relating to its narrow sense may be discussed.

One important theory concerning anaphora and coreference is the Binding Theory (Chomsky 1981). Firstly, let’s see the definition of binding.

(1) Binding:

A binds B iff A c-commands B and A and B are coindexed.

Within the Binding Theory, if an element is not bound, it is said to be free.

In this thesis, the following definition of c-commands is adopted:

(2) C-command⁸:

⁷ In fact, there are also alternative analyses for the Binding Theory, which were discussed in works such as Reinhart 1983, Reinhart & Reuland 1993, Büring 2005, Safir 2013 and Menuzzi & Lobo 2016. The proposal of Reinhart & Reuland will also be discussed in the later part of this section.

⁸ Some languages may apply other kinds of c-command, namely the cyclic c-command proposed by Huang 1982, and the weak c-command proposed by Teng 1985 and Huang, Li & Li 2009, which will be adopted to show some properties relating to the interpretation of the backward anaphora in Chinese, see section 2.5.

The node A c-commands the node B iff A and B are sisters or the sister of A dominates B.⁹

For example, in (3), *o João* binds *se* ‘self’ since it c-commands *se* (*se* is dominated by the sister of *João*) and both share the same index.

- (3) O João₁ deitou -se₁.
 the John lay down self
 ‘John went to bed.’

Binding Theory originally presented three principles, namely the principles A, B and C, as proposed by Chomsky 1981.

Principle A states that an anaphora (in its narrow sense) must be bound in its local domain. In GB theory, the local domain of an element is the governing category of the element, which can be defined as follows: the governing category of *a* is the minimal domain containing *a*, its governor, and an accessible subject (cf. Chomsky 1986). In MP of Chomsky 1995, the notion of government has been abandoned, although the notion of local domain subsists. The current study will adopt Büring’s 2005 definition of local domain, also known as binding domain, which is defined in the following way: for a reflexive, the binding domain is the minimum category that contains the reflexive, its case assigner and a subject¹⁰. For a non-reflexive element, its local domain is the domain of co-argument, which is the minimum XP that contain the non-reflexive element, its case assigner and all of the arguments of the case assigner.

Chomsky 1982 then defines anaphora as a nominal element with a [+anaphor -pronominal] feature¹¹. In this way, in (3), *se* ‘self’ is a reflexive anaphora and must be bound by *o João*, which is the closest nominal element in the local domain of *se* ‘self’. However, some languages, like Chinese, permit long-distance reflexive binding. In this case, according to Cole & Sung 1994, it is assumed that there is a successive movement of the reflexive expression, see (4).

- (4) Zhangsan₁ renwei Lisi₂ bu xihuan ziji_{1,2}.
 Zhangsan think Lisi not like self
 ‘Zhangsan thinks that Lisi does not like himself.’

⁹ In the construction of the phrase structure, two root nodes can form a new node via merge. It is believed that the new merged node contains and dominates the root nodes or the root nodes of the root nodes that form it.

¹⁰ However, according to Safir 2013, the notion of case assigner is problematic when including the A-movement trace, in the case of passive and subject fronting structures.

¹¹ In this section, anaphor takes its narrow sense.

In this case, the reflexive *ziji* ‘self’ may refer to either *Zhangsan* or *Lisi*. When *ziji* ‘self’ takes *Zhangsan* as its antecedent, a successive movement of the reflexive takes place in LF. During this movement, *ziji* ‘self’ successively moves to V, I and C of the subordinate clause, and then to V of the matrix clause. It finally moves to I of the matrix clause, where *Zhangsan* is contained in the local domain of *ziji* ‘self’.

Principle B of the binding theory states that a pronoun must be free in its local domain. Chomsky 1982 defines a pronoun as a nominal element with a [+pronominal -anaphor] feature. In (5), *o João* is in the local domain of the clitic pronoun *o* ‘him’. As a result, *o* ‘him’ cannot be bound by *o João* and the two elements must be disjoint in their reference.

- (5) O João₁ chamou-o₂.
The John called him
‘John called him.’

Principle C states that an R-expression must be free in any domain. For example, in (6), *o Rui* is an R-expression and therefore cannot be bound by the pronoun *ele* ‘him’.

- (6) Ele₁ encontrou o Rui₂.
He encountered the Rui
‘He met Rui.’

Reinhart 1983 claims that there are some exceptions to *Principle C*, which should be explained in terms of pragmatics. Nevertheless, in some Asian languages, *Principle C* may be redefined as *Principle D* (initially proposed by Lasnik 1991), which states that a less referential expression cannot bind a more referential expression. See the examples below:

- (7) *O João₁ viu o João₁.
The John saw the John
‘John saw John.’
- (8) ?Zhangsan₁ kanjian le Zhangsan₁.
Zhangsan see Pst Zhangsan
‘Zhangsan saw Zhangsan.’

In (7), for Portuguese, the second occurrence of *o João* cannot be bound by the first one, because

the second one is an R-expression (which must be free, according to principle C of Binding Theory). However, for Chinese, according to Huang, Li & Li 2009, the equivalent sentence of (7) is marginally acceptable (as 8 shows), because the occurrences of Zhangsan are equally referential and neither of them is more referential than the other. Only in cases of (9) and (10) is binding completely impossible, since the pronoun and *nanhair* ‘boy’, which functions as an NP with a less generalized referential value, are less referential than the proper name.¹²

- (9) *Ta₁ kanjian le Zhangsan₁.
 he see Pst Zhangsan
 ‘He saw Zhangsan.’
- (10) *Nanhair₁ kanjian le Zhangsan₁
 boy see Pst Zhangsan
 ‘The boy saw Zhangsan.’

However, *Principle D* of binding theory cannot explain the case of (11), where a proper name, though being more referential, fails to bind a more common expression *nanhair* ‘the boy’:

- (11) *Zhangsan₁ kanjian le nanhair₁.
 Zhangsan see Pst boy
 ‘Zhangsan saw the boy.’

It is reasonable to assume that *Principle D* of Binding Theory only predicts that a less referential expression cannot bind a more referential expression, but does not necessarily mean that a more referential expression is able to bind a less referential expression. As a result, the less referential expression *nanhair* ‘the boy’, still being an R-expression, is exempt from being bound by any other expressions.

It should be noticed that although Binding Theory is strongly related to coreference, this does not mean that every kind of coreferential relation is constrained by it. In some structures, the coreferential relation may be only determined by other factors, such as pragmatics, and have nothing to do with the binding theory. Most structures involved in this study, especially those in the experimental sections, are not constrained by Binding Theory, since we test pronouns in conditions that do not violate *Principle B*.

¹² According to the Informational Load Hypothesis, which will be discussed in section 3.1, the general expressions present fewer features, thus become less informative and less referential.

There are some alternative analyses for the Binding Theory, namely the studies by Reinhart & Reuland 1993 (henceforth, R&R 1993), who proposed the Reflexivity Theory. Following Reinhart 1983, R&R 1993 argued that *Principle C* of the binding theory is linked to a different inferential module, which presents an alternative to *Principles A* and *B* of the classic Binding Theory.

R&R 1993 distinguish local anaphor and long-distance anaphor by introducing two features: SELF for local anaphor and SE for long-distance anaphor. The authors consider that anaphors and pronouns possess different properties. In their theory, anaphors possess the reflexivizing function, with [+SELF] feature, while pronouns possess a referential independence property signaled by the [+pronoun] feature. The authors argued that the referential independence property is also related to referential expressions.

R&R 1993 then began to redefine the *Principle B* of the Bind Theory in the following way:

- (12) a. A predicate is reflexive iff two of its arguments are coindexed.
 b. A predicate (formed of P) is reflexive-marked iff either P is lexically reflexive or one of P's arguments is a SELF anaphor.

(13) Condition B

A reflexive predicate is reflexive-marked.

Reinhart & Reuland (1993: 663)

According to (12) and (13), the following sentences of (14) can be explained in an alternative binding theory.

- (14) a) *Max₁ criticized him₁.
 b) Max₁ criticized himself₁.
 c) *Max₁ / he₁ criticized Max₁.

Reinhart & Reuland (1993: 663)

For (14a), the two arguments of the sentence are coindexed, which requires the predicate to be reflexive. However, neither of the two arguments of (14a) is a SELF anaphor, thus it constitutes an ill-formed sentence. For (14b), the two arguments are also coindexed. In this case, the complement of the predicate is *himself*, a SELF anaphor, which satisfies (12b) and yields a well-formed sentence. (14c) is ill-formed in the same way as (14a).

Then the authors proposed the following *Condition A*:

(15) A reflexive-marked predicate is reflexive.

Reinhart & Reuland (1993: 671)

As a result, *Condition A* and *Condition B* constitute the alternative binding theory of R&R 1993, with the following logic: if a predicate has a reflexive-marked argument in terms of syntax, then it must be semantically marked as reflexive as well; if a predicate is semantically reflexive, then it must contain a reflexive-marked argument in terms of syntax.

The alternative approach of R&R 1993 is able to explain the properties related to the classic Binding Theory applying only to the notion of reflexivity, and avoids referring to the three categories of pronoun, anaphor and R-expression. It should be noticed that (14a-b) correspond exactly to the classic examples captured by *Principles A* and *B* of Binding Theory. However, the proposal of R&R 1993 cannot explain phenomena involving two clauses in complex sentences, which may be the case of R-expressions that may be captured in classic Binding Theory, since they claim that the *Principle C* is related to other pragmatic factors. The current study will adopt the classic Binding Theory in the following sections and chapters.

2.2 Null subject languages

Languages can be distinguished by the fact that some of them allow a covert form of a referential subject pronoun, while others do not. This phenomenon, which is designated as null subject, was first observed by Perlmutter 1971, who pointed out that some languages may allow a finite clause to omit the subject, while others do not. Taraldsen 1980 considered that this phenomenon is related to the verbal agreement.

Inspired by Taraldsen 1980, Chomsky 1981, in the GB, argues that null subject languages like Italian present a series of related properties:

- (16) (i) missing subject
- (ii) free inversion in simple sentences
- (iii) long wh-movement of subject
- (iv) empty resumptive pronouns in embedded clause
- (v) apparent violations of the that-t filter

(Chomsky 1981: 240)

Chomsky then argued that there exists a *pro*-drop parameter that distinguishes null subject and

non-null subject languages. For example, Italian and Spanish show the properties above, and belong to the group of null subject languages, while English and French do not have those properties and are called non-null subject languages.

According to Chomsky 1981, the possibility of allowing a null subject depends on the degree of richness of the verbal agreement of a certain language. Languages with rich verbal agreement permit the omission of the subject in finite clauses (e.g. Spanish and Italian). On the other hand, languages without rich verbal agreement do not accept null subjects in matrix clauses (e.g. English and French).

Chomsky 1981 considered that there are three types of empty categories: NP-trace, Wh-trace (variable) and *PRO*. He stipulated that *PRO* is ungoverned and controlled in an infinitive clause, but may also appear in a finite clause.

Chomsky 1982 changed the classification of empty category by introducing two features, namely anaphoric and pronominal, according to which empty categories can be divided in four categories:

- | | |
|--------------------------------|----------------------------|
| (17) [+anaphoric – pronominal] | <i>NP-trace</i> |
| [- anaphoric – pronominal] | <i>Wh-trace</i> , variable |
| [+anaphoric + pronominal] | <i>PRO</i> |
| [- anaphoric + pronominal] | <i>pro</i> |

(cf. Chomsky 1982: 78)

NP-trace and Wh-trace are generated by movement, while *PRO* appears in infinitive clauses. Hence, *pro* is the empty category that appears in finite clauses, (and, in languages like Portuguese, in clauses with inflected infinitive). Bearing [-anaphoric] and [+pronominal] features, *pro* is governed and receives case. However, the licensing of *pro* depends on the degree of richness of the verbal agreement of a language. Languages with rich verbal agreement may recover the phi-features (person and number) of *pro*, and allow finite null subjects, while languages without rich verbal agreement cannot allow finite null subjects.

Rizzi 1982 also related the finite null subject to the verbal inflection, but he differed with respect to the relevant inflectional features. He considered that the verbal inflection of a null subject language has pronominal and referential features. Languages with [+pronominal] and [+referential] features in their inflection allow null subjects in finite clauses, while languages

with [-pronominal] and [-referential] features in their inflection do not. Languages with [+pronominal] but [-referential] features in their inflection only permit expletive null subjects. Rizzi 1986 then argued that the null finite subject *pro* must be licensed and identified. For null subject languages, the verbal inflection may govern *pro*, thus *pro* is licensed. On the other hand, the phi-features of the verbal inflection also recover *pro* as a definite pronoun. As a result, *pro* is identified.

However, there is an asymmetry between 1st, 2nd and 3rd person in the recovery of the identification of *pro*. Studies like Matos & Duarte 1984, Eliseu 1984, Ambar 1988 and Lobo 1994, 1995 claimed that 1st and 2nd person null subjects have a deictic value and can always be recovered as referring to the speaker or the hearer. But the reference for 3rd person null subjects cannot be established if there is no context. In this sense, the authors argued that 3rd person null subjects are only legitimated when they can be recovered from the context.

A similar point of view is held by studies like Camacho's 2013, who claims that even in rich verbal agreement languages, the reference of a 3rd person null subject may not be uniquely identified by morphology. In that case, the reference of a 3rd person null subject must be identified via a discursive topic. Otherwise, an overt pronoun must be used.

Lobo 1994, 1995 also proposed that the structure of a sentence with a null subject is different from that of a sentence with an overt pronoun. Following Rigau 1988, Lobo 1994, 1995 argues that the overt pronoun in EP implies an emphasis sense, occupying the position of SpecFocus, while the null subject in EP occupies a lower position, namely SpecIP.

Lobo's 1994, 1995 proposal in fact indicates that there is a functional category Focus above IP, which is also proposed by Martins 1993 and adopted by Kato 1999, who identify this category as SigmaP, following Laka 1990.

Lobo 1994, 1995 also proposes that above the projection of Focus, there is a functional category DiscourseP that links the sentence to the discursive context. In her analysis, an operator of the nature of *pro*-variable occupies the DiscourseP position, which indicates the referential value of the null or overt subject of the sentence.

As a result, for sentences with null or overt 1st and 2nd person subjects, the structure can be exemplified as in (18), where the operator is always linked to the speaker or hearer.

- (18) Vou ao cinema.
 go-1sg to+the cinema
 ‘I will go to the cinema.’
 [Disc" Op1^ap.1 [Disc' ([F" [F')][Agr" e₁ [Agr' V+T₂+Agr [T" [T' v₂.....ao cinema]]]]]]]
 (Lobo 1995)

For 3rd person null subjects, the operator must be linked to the discourse (19a). If it fails, as in (19b), then it is not able to identify the referential value of the null subject.

- (19) a. O João não janta. Vai ao cinema.
 the John not have dinner go-3sg to+the cinema
 ‘John will not have dinner. He will go to the cinema.’
 [Disc" Op3^ap.1 [Disc' ([F" [F')][Agr" e₁ [Agr' V+T₂+Agr [T" [T' v₂]]]]]]]
 b. Vai ao cinema.
 go-3sg to+the cinema.
 ‘(He) will go to the cinema.’
 [Disc" [-] [Disc' ([F" [F')][Agr" e₁ [Agr' V+T₂+Agr [T" [T' v₂.....ao cinema]]]]]]]
 (Lobo 1995)

Barbosa 1995, 2000, 2009 also proposes that null and overt subjects in EP occupy different positions. Her basic idea is that the real position of overt subjects in EP is post-verbal, while the apparent pre-verbal subject in this language is in fact the result of clitic left-dislocation (CLLD), a hypothesis that is shared by several studies for other languages, for instance Pollock 1997, Alexiadou & Anagnostopoulou 1998.

The basic idea of Barbosa is based on the fact that in EP, the pre-verbal and post-verbal subject convey different interpretations. For example, as illustrated by Barbosa 1995 and Costa 2000, the pre-verbal subject in EP refers to an old information, which predominantly serves as the topic of the sentence. As for the post-verbal subject, it usually conveys a new information, which serves as the informational focus of the sentence. See examples (20) and (21):

- (20) Q: O que aconteceu ao Rui? ‘what happened to Rui?’
 A: O Rui comeu um bolo.
 the Rui ate one cake
 A: #Comeu um bolo o Rui.
 Ate one cake the Rui
 ‘Rui **ate one cake.**’

- (21) Q: Quem comeu o bolo? ‘Who ate the cake?’

- A: #O Rui comeu o bolo.¹³
 the Rui ate the cake
 A: Comeu o bolo o Rui.
 ate the cake the Rui
 ‘**Rui** ate the cake.’

Barbosa 1995, 2000 then argues that the real argument position for the subject in EP is post-verbal, while the apparent pre-verbal subject is the result of CLLD. In Rizzi 1982’s original proposal for post-verbal subjects in Romance languages, there is a *pro* in the SpecTP position to fulfill the Extended Projection Principle (EPP), which requires that every sentence must have a subject in its SpecTP position. However, Barbosa 1995, 2000 proposes that in EP, since the verbal inflection, which serves as a pronominal clitic, has a [+pronominal] feature, the raising of the verb from vP to TP already satisfies the EPP, thus the subject remains in-situ and there is no need to insert a *pro* in the SpecTP.

- (22) [TP [T telefonou [VP o João ~~telefonou~~]]]
 called the John

For sentences with pre-verbal subject DP, there is a *pro* in-situ (in SpecvP) to fulfill the verbal agreement, while the pre-verbal subject can be basically inserted in the SpecTP position, which is in fact an A’-position in a CLLD structure.

- (23) [A Maria₁ [TP telefonou *pro*₁]]
 the Mary called

If the pre-verbal subject is a quantifier, then it is generated by A’-movement. In that case, the pre-verbal subject is focalized and located in FP.

- (24) [FP ninguém [F’ telefonou ~~ninguém~~]]
 nobody called

In both structures, there is no A-movement, which leads the author to conclude that the pre-verbal subject in EP is not an A-position.

¹³ The sentence with the SVO order may still serve as the answer to the question, but in this case, it indicates that someone else may also eat the cake. This is not exactly the meaning that the VOS order conveys, since it indicates that only *Rui* ate the cake.

Later, in Barbosa 2009, the author reviews the EP null subject analysis on the basis of the two hypotheses of Holmberg 2005, who analyzed null subjects under the feature valuing system. I will return to this topic after presenting studies about null subjects in Chinese under the GB framework.

Li & Thompson 1976 noticed that not all null subject languages apply the same mechanism to legitimate null subjects as the Romance null subject languages. Huang 1984, 1989 pointed out that some Eastern Asiatic languages (such as Chinese and Japanese) may allow null argument while there is no verbal agreement at all. Huang 1984 then argued that *pro* in Chinese can be legitimated by the GCR, which was already discussed in Chapter 1 and will be repeated here as (25):

(25) The Generalized Control Rule

An empty pronoun is coindexed with the closest nominal¹⁴.

(suggested by Huang 1984: 552 and reformulated by Huang, Li & Li 2009: 209)

Huang 1984 argued that for languages with rich verbal agreement, the closest nominal is the verbal agreement that governs *pro*. For languages without verbal agreement, the closest nominal can be a nominal element that c-commands *pro*, which can be located in a higher domain. As exemplified in (26):

- (26) [Zhangsan₁ shuo [CP [-]₁ mai shu le].
Zhangsan say buy book Pst
'Zhangsan said that (he) bought the books.'

The empty category located in the subordinate clause is considered as a *pro* by Huang 1984. As there is no nominal element in the subordinate clause that can serve as the antecedent of *pro*, the matrix subject *Zhangsan* in (26) becomes a candidate for the antecedent of *pro*, since it is the closest nominal element to *pro*. As a consequence, *pro* is controlled by *Zhangsan* and shares its index. In this way, the Chinese *pro* (specifically those in subordinate clauses) is legitimated and identified.

¹⁴ Huang 1984:552-553 'defines 'closest' in the following manner. Following Chomsky 1980, A is closer to B than C is if A c-commands B but C does not c-command B. Furthermore, for two nodes A and C, both of which c-command B, A is closer to B than C is if A but not C occurs within the same clause as B, or if A is separated from B by fewer clause boundaries than C is.'

Huang 1984 considered that there is another kind of null argument in Chinese, which can be analyzed as a variable and is A'-bound by a topic. As shown in (27), the null subject (in (a)) and the null object (in (b)) in the subordinate clause can be considered as variables, which is bound by the topic *zhege ren* 'this person' and *zheben shu* 'this book', respectively.

- (27) a. [Top *Zhege ren*₁] *Zhangsan*₂ *shuo* [CP [-]₁ *neng ying*].
 this person *Zhangsan* say can win
 'This person, Zhangsan says that (he) can win.'
 b. [Top *Zheben shu*₁] *Zhangsan*₂ *shuo* [CP *Lisi*₃ *kan guo* [-]₁].
 this book *Zhangsan* say *Lisi* read Pst
 'This book, Zhangsan said that Lisi read (it).'

It should be noticed that the topic may also be null, as Huang 1984 (under Tsao's 1977 proposal) considered that Chinese may form a topic chain, in which only the head of the chain is overt, while all of the lower elements of the chain remain covert.

- (28) Top *Zhongguo*_i [-]_i *difang hen da*, [-]_i *renkou hen duo*, [-]_i *tudi hen feiwo*, [-]_i *qihou ye hen hao*. [-]_i *women dou hen xihuan*.
 'as for China_i, [-]_i (its) land area is very large, [-]_i (its) population is very big, [-]_i (its) land is very fertile, [-]_i (its) climate is also very good, [-]_i we all like (it).'
- (Huang 1984: 549)

Sometimes the topic can be retrieved from the context, so it can be omitted even if there is no topic chain. As in (29), the null topic of the answering sentence can be deduced as *zhege ren* 'this person', which serves as the antecedent of the null subject, which is analyzed as an A'-bound variable.

- (29) A: *Zhege ren*₁ *lai guo le ma*?
 this person arrive Pst Pst Int
 'Did this person arrive?'
 B: [Top [-]₁] *Lai guo le*.
 arrive Pst Pst
 '(Yes), (he) arrived.'

As a consequence, in Huang's 1984 analysis, all of null subjects in matrix clauses of Chinese should be conceived as variables bound by an overt or null topic.

Unlike null subjects, Huang 1984 considered that the null object in Chinese can only be

analyzed as variable, but not *pro*. Considering sentence (30), the null object can only be analyzed as a variable which is A'-bound by the topic *zhege ren* 'this person', but not as a *pro*. Huang 1984 argued that the impossibility to consider the null object as *pro* is due to the GCR, which forces *pro* to be controlled and bound by the closest nominal element. In (30), the closest nominal to the null object is the subject *Zhangsan*. If the null object is *pro*, it must be controlled and bound by *Zhangsan*, but the co-reference between *Zhangsan* and *pro* violates the *Principle B* of the Binding Theory, which states that a pronominal (here as *pro*) must be free in its local domain. As a result, the null object in Chinese should only be analyzed as a variable, but not *pro*.

(30)[Top Zhege ren₁,] Zhangsan₂ bu renshi [-]_{1/*2}.
this person Zhangsan not know
'This person, Zhangsan does not know (him).'

In a later section 4.3, I will present some studies which argue that the Chinese matrix null subject can also be analyzed as *pro*, and so do some null objects.

In recent studies (Roberts & Holmberg 2010), null subject languages are divided into three groups. The consistent null subject languages include all kinds of null subject languages which are related to the rich verbal agreement (including the Romance null subject languages); the partial null subject languages includes languages which only allow controlled finite null subjects in subordinate clauses (such as Finish and BP); and null subject languages without verbal agreement (such as Chinese and Japanese) are called discourse null subject languages or radical *pro*-drop languages.

In the MP Framework, several studies analyze consistent null subjects in the perspective of chain-reduction by agreement (e.g. Holmberg 2005, 2010, Roberts 2010 and Barbosa 2009).

Chomsky 2001 pointed out that features should be divided into two groups: interpretable features and uninterpretable features. The uninterpretable features should be valued by match with interpretable features and eliminated. This point of view raised some questions for the analysis of null subjects.

Rizzi 1986 argued that *pro* must be licensed and identified. T in the consistent null subject languages contains a [+pronominal] feature. In this way, the content of *pro* is identified through

the phi-feature of T. However, according to the MP, the uninterpretable features must be deleted. Chomsky 2001 considered that the phi-features in T are uninterpretable and must be valued by the interpretable phi-features of a DP merged in SpecTP. In this manner, the phi-features in T of consistent null subject languages should not identify the content of *pro*, since they themselves are also uninterpretable.

Holmberg 2005 proposes that there are two ways to solve this problem.

(31) Hypothesis A:

There is no pro at all in null-subject constructions. Instead Agr, the set of φ -features of I, is itself interpretable; Agr is a referential, definite pronoun, albeit a pronoun phonologically expressed as an affix. As such, Agr is also assigned a subject theta-role, possibly by virtue of heading a chain where the foot of the chain is in vP, receiving the relevant theta-role.

(Holmberg 2005: 537)

Hypothesis B:

The null subject is specified for interpretable φ -features, values the uninterpretable features of Agr, and moves to SpecIP, just like any other subject. This implies that the nullness is a phonological matter: The null subject is a pronoun which is not pronounced.

(Holmberg 2005: 538)

Holmberg 2005 admits that Hypothesis B is correct. The function of *pro* is very similar to that of an overt pronoun, the only difference being that *pro* does not have phonetic realization. In this sense, *pro* is only an unpronounced pronoun, in other words, a deleted pronoun. Holmberg 2005 pointed out that the difference between consistent null subject languages and partial null subject languages resides in the fact that the former contains a definite feature in T, designated as D-feature. He admits that pronouns may be DPs or phiPs. DPs have a more complex structure than phiPs: [DP D [phiP phi [NP N]]] while null pronouns are phiPs.

According to Holmberg 2005, T has an uninterpretable phi-feature and the phiP (null subject) has an interpretable phi-feature. The interpretable phi-feature of the phi-P values the uninterpretable phi-feature of T. As a result, T and the null subject, with a union of phi-features, form a definite pronoun, in a language that has a D-feature in T. In a language without D in T, the definite pronoun cannot be formed in the same way so an indefinite pronoun is formed instead.

Thus, the definite null subject in consistent null subject languages must be a null pronoun specified with phi-features but without D. This pronoun can only become referential when T

contains a D-feature. This is the most notable property of this group of languages. In Holmberg 2010, the author analyzed how the definite pronoun is formed by using the concept of incorporation of Roberts 2010. He then analyzes how the D in T identifies the content of *pro*. Before discussing Holmberg 2010, I will present the proposal of incorporation of Roberts 2010. Taking the Romance clitics as an example, Roberts 2010 defines the concept of probe and defective goal¹⁵. According to his proposal, there is an incorporation between a transitive verb (v^*) and the clitic form of a direct object. The v^* contains an unvalued phi-feature while the clitic contains a valued phi-feature. As both the entities possess a phi-feature, the incorporation is possible.

(32) *Incorporation can take place only where the features of the incorporee are properly included in those of the incorporation host.*

(Roberts 2010: 65)

In this way, the v^* serves as a probe, while the clitic serves as a defective goal. Following the sequence of the Agree operation, the probe copies the valued features of the clitic:

(33) a. *Trigger for Agree:*

v^* [Pers:___], [Num:___] ϕ [Pers:a, Num:b]

b. *Outcome of Agree:*

v^* [**Pers:a, Num:b**] ϕ [Pers:a, Num:b]

(Roberts 2010: 66)

In this way, the clitic is extracted to the left side of v^* , while its lower copy is deleted, as the consequence of the incorporation. However, Roberts also proposes that incorporation is not compatible with the EPP feature. If the probe contains an EPP feature, the defective goal will be moved to satisfy the EPP and cannot be incorporated with the probe. As a result, the author represents the following generalization:

(34) *A probe P can act as an incorporation host only if it lacks an EPP-feature.*

(Roberts 2010: 67)

Roberts then gives the definition of defective goal:

¹⁵ Here, the example of Romance clitics only serves to introduce the idea of incorporation. As can be seen from the later analysis, Roberts 2010 claimed that the Romance null subject cannot be analyzed in the same way as the clitics.

- (35) *A goal G is defective iff G's formal features are a proper subset of those of G's probe P.*
(Roberts 2010: 70)

Roberts then analyzed the nature of *pro*. He firstly considered the possibility that *pro* may also be a defective goal, just like a clitic, but then denied this hypothesis. According to various authors, including Cardinaletti & Starke 1999 and Holmberg 2005, *pro* is a weak pronoun, and not a strong pronoun or a clitic. Roberts 2010 then characterizes *pro* as a $D_{\min/\max}$, a determinant nature constituent that occupies the SpecTP position to satisfy the EPP.

He then assumes three central points: T in consistent null subject languages is pronominal; there is a D-feature in T in this group of languages; the D-feature is related to the rich verbal agreement. In this way, *pro* in this group of languages can be a defective goal, as it contains phi-features and a D-feature, which are also shared by T. Meanwhile, T also contains an EPP-feature, which can be satisfied by *pro*. As a result, the author proposed a generalization to characterize defective goals:

- (36) *Defective goals always delete/never have a PF realisation independently of their probe.*
(Roberts 2010: 76)

According to (34), incorporation cannot occur in this situation as the probe contains an EPP-feature, which implies that *pro* cannot be omitted like a clitic. However, Roberts considered that *pro* can be deleted as a consequence of chain reduction. Due to the agreement between the D and phi-features, *pro* can be eliminated after the copying of those features.

Now we return to Holmberg 2010, who indeed accepts the incorporation analysis for consistent null subject languages.

For this group of languages, the identification of the third person null pronoun depends on an antecedent. Frascarelli 2007 proposed that every sentence presents an aboutness-shift topic (A-topic). Holmberg 2010 suggested that the A-topic is located in an A'-position inside the C domain of the sentence, which can be overt or null. The antecedent of a null pronoun is the A-topic of the sentence. The A-topic, when it is in its null form, can be considered as a copy of the A-topic of the previous sentence. The following sentence is an Italian example that explains how the A-topic functions.

- (37) [_{CP} <Gianni₁> [questa mattina Gianni₁ ha visitato la mostra.]]
 [_{CP} <e₂> [Piu tardi ha φP₂ visitato l' università.]]
 1=2

(adapted from Holmberg 2010: 96)

In the first sentence, the A-topic is *Gianni*, while in the second sentence, the antecedent of the null subject phiP is the A-topic in CP, which is null in this case. This null A-topic is an omitted copy of the A-topic of the first sentence, which is *Gianni*. As a result, the null subject of the second sentence also refers to *Gianni*.

Holmberg 2010 considers that the antecedent of the null subject in consistent null subject languages is the A-topic. As a result, the A-topic will value the uninterpretable D-feature while D will copy the referential index of the A-topic.

Nevertheless, this is only true for the third person, as suggested by Frascarelli 2007; the first and second persons are discourse persons and the identification of the speaker and hearer is always presented in the discourse. Matos & Duarte 1984, Lobo 1994, 1995 also verified this asymmetry between first, second and third persons in EP.

Based on the idea of incorporation of Roberts 2010, Holmberg considers that the null subject of the consistent null subject languages is a consequence of incorporation, which can be explained by the Agree operation.

In a language with uD¹⁶ feature in T, the null pronoun incorporates with T in order to form a definite pronoun. The incorporation is processed in the following way.

T possesses a D-feature marked by the A-topic, a nominative case and an uninterpretable phi-feature. The subject phi-P possesses a phi-feature of third person singular and an uninterpretable case feature (step 1 in 38). T then marks the uninterpretable case feature of phi-P and the latter marks the uninterpretable phi-feature of T (step 2 in 38). In this way, T shares all of the features of the phi-P, which implies their incorporation.

Holmberg 2010 suggested that the incorporation can be explained in the following way: T and the phi-P form a chain which can be reduced. With the non-realization of the subject phi-P, the incorporation is completed (step 3 in 38). T will have a D feature and shares the referential

¹⁶ Holmberg considers that the D-feature in T is uninterpretable and must be marked by the A-topic.

index of the A-topic. Consequently, the result of the incorporation is a definite pronoun.

- (38) 1 [T, D2, u ϕ , NOM] [vP [3SG, uCase] v...]
 2 [T, D2, 3SG, NOM] [vP [3SG, NOM] v...]
 3 [T, D2, 3SG, NOM] [vP [~~3SG, NOM~~] v...]

(Holmberg 2010: 97)

If the subject is a lexical DP or a D-pronoun, then the D of the subject can mark the uninterpretable D feature of T while the latter will not adopt the referential index of the A-topic. The lexical subject DP or the D-pronoun will become the A-topic and cannot be incorporated in T. In this case, the overt subject will appear in the SpecTP position to fulfill the EPP. Holmberg did not discuss the idea of Barbosa 1995, 2000, 2009, Alexiadou & Anagnostopoulou 1998, among others, who propose that the pre-verbal overt subject should appear in an A'-position.

The problem here is how the consistent null subject languages mark the EPP in the case of incorporation. Holmberg 2010 suggests that the EPP can be fulfilled by the A-topic that values the uD of T. According to his analysis, all of the null subjects must have an A-topic as their antecedent. Without the antecedent it is always necessary to fill the SpecTP position to mark the EPP. This criterion implies that when the subject is indefinite, the SpecTP must be occupied to fulfill the EPP. One example that favors this analysis is (39), in which the subject position must be filled by a clitic pronoun *se*, in order to obtain an indefinite reading of the subject.

- (39) *É assim que se faz o doce.* (EP)¹⁷
 is like this that SE does the sweet.
 'The sweet is done like this.'

(Holmberg 2010: 92)

In summary, the EPP is fulfilled by the A-topic, which marks the uninterpretable D-feature of T, which shares the referential index of an A-topic and undergoes an incorporation operation with the null subject to value their uninterpretable features, and therefore, form a definite pronoun.

Now we come back to the proposals of Barbosa 2009. Differing from Holmberg 2005, Barbosa

¹⁷ However, Barbosa 2011 claims that EP may also accept the null generic subject, as shown in the following example: *É assim que pro fazem o bolo*, 'the sweet is done like this.' in which the *pro* has the generic meaning.

2009 proposes that Hypothesis A is superior to Hypothesis B (See (31) above), indicating that it is the verbal inflection, but not *pro*, that possesses the phi-feature and undergoes Agree operation with T.

Barbosa 2009 then proposes that the phi-feature in T is uninterpretable but valued, in the sense of Pesetsky & Torrego 2004. The author confirms also that T with valued phi-feature does not contain an EPP feature. As a result, the formation of a sentence with an overt DP subject in EP can be analyzed in the following way:

(40) [[T Tns Pers:a, Num:b] [V/vP **DP** Pers:a, Num:b / Case []]] (Barbosa 2009)

The phi-feature in T is uninterpretable, and it searches a constituent with an interpretable phi-feature, namely the DP. Then T and the DP form a relation of probe and goal, which results in an Agree operation that eliminates the uninterpretable phi-feature of T and the uninterpretable case feature of the DP.

For the case of null subjects, Barbosa 2009 proposes that *pro* in EP is a $\phi_{\min/\max}$, while T functions as an incorporation host that lacks an EPP feature.

(41) [[T Tns [Pers:a, Num:b]] [V/vP ϕ **min/max**[Pers:a, Num:b]]] (Barbosa 2009)

T serves as a probe as it contains uninterpretable and valued phi-features, while $\phi_{\min/\max}$ serves as a goal. For the sake of Agree, the content of the goal is eliminated, which results in chain reduction. In this way, the $\phi_{\min/\max}$ is not produced and the incorporation is completed, which results in a null subject.

The previous part has reviewed some studies regarding null subjects in Romance null subject languages under MP framework. Now we see the case of Chinese. According to proposals such as Roberts & Holmberg's 2010, some discourse null subject languages, such as Chinese, Japanese and Korean, which lack verbal agreement, are considered as radical *pro*-drop languages. Several previous studies have shown that radical *pro*-drop has the following properties: lack of agreement (Saito 2007), bare NP (Tomioka 2003), agglutinative morphology (Neeleman & Szendroi 2007), among others.

Tomioka 2003 argued that all radical *pro*-drop languages must allow bare NPs in the argument position. See the example of Japanese in (42), where the bare NP object can be indefinite singular, definite singular, indefinite plural or definite plural.

- (42) Ken-wa ronbun-o yon-da.
Ken-top paper-acc read-past
'Ken read a paper/papers/the paper/the papers.'
(Tomioka 2003:328)

Thus, the radical *pro*-drop is a kind of NP anaphora. The reference of the null argument is recovered by two semantic operations: Existential Closure (Heim, 1982): \exists closure and Type-shifting of a predicate to an individual (Partee, 1987): Iota.

Barbosa 2011, 2013 argues that the null subject in BP (also considered as a topic prominent language) is also NP anaphora. Barbosa 2013 also compares partial null subject languages with discourse null subject languages and concludes that the null argument in both groups of languages can be analyzed as null NP anaphora.

Neeleman & Szendroi 2007 argue that all radical *pro*-drop languages of this type must have agglutinative markers for case (Japanese) or number (Chinese).

For example, in Japanese, a case marker is required to annex to syntactic constituents, as in (43), where *ga* follows the subject *kare* 'he' to mark the nominative case, while *o* follows the direct object *kare* 'he' to mark the accusative case. It should be noticed that the pronoun remains in the same form, while it is the case marker that indicates if the pronoun is nominative or accusative, which is a relevant aspect of the agglutinative case marking, opposite to the fusional case marking, where the case information is contained in the pronoun forms¹⁸.

- (43) Kare-ga kare-o settokusuru.
he-NOM he-ACC persuades
'He persuades him.'
(Neeleman & Szendroi 2007)

As for Chinese, there is a plural marker for pronouns, rather than a case marker. In (44), the first pronoun *ta-men* 'they' carries a plural marker *men*, which is annexed to the 3-sg pronoun *ta* 'he'. This property is also the opposite of what is found in the languages with fusional number

¹⁸ Fusional case means that the case information is contained in the pronoun's lexical root. As a result, the different cases of a same pronoun may have different lexical roots, for example, in English, the nominative form of 3rd person singular is 'he', while its accusative form is 'him', a lexical root other than 'he'.

marking, where the number information is contained in the pronoun forms¹⁹.

(44) Ta-men kanjian ta le.
he-PL see he Pst
'They saw him.'

(Neeleman & Szendroi, 2007)

These agglutinative markers imply that pronouns in these languages are equal to NPs (in line with Déchaïne & Wiltschko 2003), but not DPs or KPs (K represents case), which is the key point of licensing the radical *pro*-drop in Neeleman & Szendroi's theory.

Saito 2007 proposes that radical *pro*-drop is a kind of argument ellipsis. He analyzes the *pro*-drop of Japanese in the theory of LF copying and argues that this kind of argument ellipsis is only available for languages without agreement.

He argues that each null argument must have its antecedent copied in the LF. For a language with agreement, the copied antecedent cannot fulfill the agreement, because its uninterpretable case feature is already deleted in the original position. But languages without agreement do not have this limitation, since there is no feature valuing neither in Syntax nor in LF. So according to Saito 2007, the absence of agreement is the key point for the radical *pro*-drop.

The hypotheses of Neeleman & Szendroi 2007, Tomioka 2003 and Saito 2007 may imply that there are three licensing conditions for the radical *pro*-drop languages, namely the NP nature of pronouns, the existence of bare NP in argument position and the absence of agreement. However, although these three conditions are not identical, they may be related. It is possible that the first two are a consequence of the third. In that case, the NP nature of the pronoun and bare NP in argument position should not be considered as a licensing condition of the null subject, but a related property of radical *pro*-drop, just as the relationship between free inversion and that-trace effect and the consistent null subject.

However, there may be some differences among the radical *pro*-drop languages. For example, Japanese allows coreference between an embedded null object and the matrix subject, while Chinese does not. The licit formation of (45a) indicates that the covert object of Japanese should be considered as an NP, while the illicit formation of (45b) indicates that the null object of

¹⁹ However, some languages also use a plural marker to distinguish the singular and plural form of a pronoun, as in Portuguese, the plural form of *ele* 'he' is *eles* 'they', where the plural marker 's' is attached to the singular form.

Chinese should be analyzed as a variable, which functions as an R-expression and prevents the coreference interpretation by applying Binding Principle C, according to Huang 1984.

- (45) a. *Dono gakusei-mo* [CP *Dan-ga pro buzyokushi-ta*] *to* *it-ta*.
 Which student-even Dan-NOM insult-PERF comp say-PERF
 ‘Every student₁ said that Dan insulted (him)₁.’ (Tomioka 2003: 322)
- b. **Meige xuesheng₁ dou shuo Zhangsan₂ qifu le [-]₁*.
 every student all say Zhangsan bully Pst
 ‘Every student said that Zhangsan bullied (him).’

Another difference between Chinese and Japanese is revealed by the violation of island conditions. According to Li 2007, Japanese (46a) may easily accept some island violation, while Chinese cannot (46b).

- (46) a. *Bush-wa, maikeru muua-ga senkyo-ni katta node gakkari siteita*.
 Bush-top Michael Moore-nom election-dat won because disappointed
 ‘Bush, Michael Moore was disappointed because (he) won the election.’
- b. *(*Lisi₁*,) *wo yinwei [e₁ hai bu renshi naxie ren] hen danxin*.
 Lisi I because still not know those people very concerned
 ‘(Lisi,) I am very concerned because Lisi still does not know those people.’
 (Li 2007: 97)

It has been verified that Chinese only permits a gap in subject position inside an island which functions as a subject of a sentence, but not in an object position²⁰. Furthermore, an overt antecedent of the subject must be presented (Huang & Yang 2013), see examples of (47) and (48):

- (47) a. **Zhangsan₁, wo xihuan [e₁ changge de shengyin]*
 Zhangsan I like sing DE voice
 ‘As for Zhangsan, I like the voice with which (he) sings.’
- b. *Zhangsan₁, [e₁ changge de shengyin] hen haoting*.
 Zhangsan sing DE voice very good-sound
 ‘As for Zhangsan, the voice with which (he) sings is very nice.’
- (48) **[-]top₁, [e₁ changge de shengyin] hen haoting*.
 sing DE voice very good-sound
 ‘The voice with which (he) sings is very nice.’

²⁰ In fact, Li 2007 already pointed out that Chinese may allow null objects inside an island in some situations. As it will be discussed in the chapter 4, Pan 2017 and Zhang 2002 verified that Chinese does allow null object to violate island conditions, as long as the verb that selects the null object transmits a non-episodic meaning.

The illicit formation of (47a) indicates that the subject gap inside the complex NP island should be a variable generated by movement. The licit formation of (47b) then indicates that the gap is not a variable, but a *pro*, which is controlled by the topic *Zhangsan*, according to the GCR of Huang 1984, which is defined in the introduction and in (25) of the current section, here repeated in (49). The illicit formation of (48) indicates that the GCR cannot be applied in the absence of an overt antecedent.

(49) An empty pronoun is coindexed with the closest nominal.

(suggested by Huang 1984: 552 and reformulated by Huang, Li & Li 2009: 209)

The examples listed above, (45b), (46b) and (47), reveal that the null subject and object in Chinese should not be analyzed as a null NP anaphora, since the GCR plays an important role in explaining the asymmetry between (47a) and (47b), which shows that *pro* does exist in this language. The illicit sentences of (45b), (46b) and (47a) indicate that if a null argument in Chinese is not *pro*, it must have a variable nature and be subject to island conditions. Li 2007 also revealed that there is an asymmetry between subject and object regarding island violations in Chinese, as some kinds of null objects, but not null subjects, appear to violate the island condition in some contexts²¹.

(50) (Naxie ren₂), wo yinwei [Lisi hai bu renshi e₂] hen danxin.

those people I because Lisi still not know very concerned

‘(Those people), I am very concerned because Lisi still does not know *e*.’

(Li 2007: 97)

Li 2007 then argues that there is a TEP in the null object position inside the islands. The author proposed that some transitive verbs obligatorily subcategorize for an object.

(51) Subcategorization Requirements on TEP:

a. If a head subcategorizes for an E, E must be present in the syntactic structure.

b. An E can be a true empty position only in subcategorized positions.

(Li 2007: 90)

²¹ In the GB Framework, some asymmetries between subject and object concerning empty categories can be explained by the Empty Category Principle of Chomsky 1981, which is related to the notion of government. Since the current framework of MP has eliminated government, some related phenomena (including the subject/object asymmetry) can no longer be explained. Rizzi 2016 attempts to explain some subject/object asymmetries by using the theory of Bare Phrase Structure, under the mechanism of Labeling. However, this analysis is beyond the scope of the current study.

When the object position cannot be filled by a *pro* or a variable, then it will be filled by a TEP, in order to satisfy the subcategorization requirement of the verb. In this case, a TEP cannot be considered as a *pro* or as a variable. It is present only to satisfy a subcategorization requirement and is not equivalent to any lexical items. When interpreting TEP, Li 2007 proposes that its referential value can be copied from a linguistic antecedent or an antecedent from the discourse-pragmatic context at LF.

In fact, Xu 1986 had already proposed that Chinese null arguments can be analyzed as a Free Empty Category (FEC), which does not bear any features and can be used in any situation, as long as the discourse information allows it. However, Li 2007 proposes that TEP is different from FEC, because the former is not an empty category, while the latter is.

Later work such as Pan's 2016, 2017 addressed the violation of island conditions with null objects in Chinese, concerning the internal structure of the null element. These analyses will be discussed in Chapter 4.

Concerning the nature of *pro* in Chinese, Li 2007 proposes that *pro* has a DP nature, which involves a D feature which needs to be identified by another nominal expression. Li 2007 proposes that the identification process of D corroborates the GCR. Since the Chinese *pro* must contain a D feature, it must obey the GCR and be controlled by the closest nominal. If a null object is *pro*, then its D feature must be identified via GCR, which may lead to an illicit formation: since the closest nominal element is the subject in the same domain of the null object. Thus, if the subject in the same domain controls the null object, their coreference violates the *Principle B* of the Binding Theory, which explains why Chinese null objects cannot be analyzed as *pro*. However, Li 2007 proposes that the Japanese null argument has a NP nature, which lacks the D feature. As a result, this language does not need the GCR to fulfill the D feature identification. As a consequence, both the null subject and the null object in Japanese can be analyzed as null NPs and may have a more flexible distribution than Chinese in island constructions.

2.3 The Avoid Pronoun Principle

Chomsky 1981, 1986 proposes the Avoid Pronoun Principle, which states that a pronoun should

be avoided as long as it is possible. Brito 1991: 116-117 adopts this principle for EP, defending that ‘[*e*]vitar o pronome numa língua deve ser entendido como a estratégia sintática de usar pronomes nulos interpretados correferencialmente, enquanto o uso de pronomes lexicalmente realizados conduz a uma interpretação de referência disjunta.’ (avoiding the pronoun in a language should be understood as a syntactic strategy to use null pronouns which are interpreted coreferentially, while using overt pronouns leads to a disjoint reference interpretation).

Later, Costa et al. 1998, 1999 tested this property in EP with coordinate and subordinate clauses, verifying that in EP the null subject in the second conjunct of coordination, or in subordination, is preferentially interpreted as coreferent with the subject of the first conjunct, or the matrix subject, while the overt subject is preferentially interpreted as disjoint²².

- (52) a. A Helena₁ viu a Maria₂ no cinema mas [-]₁ não a cumprimentou.
 ‘Helena₁ saw Maria₂ in the cinema but [-]₁ did not greet her.’
 b. A Helena₁ viu a Maria₂ no cinema mas ela_{1,2} não a cumprimentou.
 ‘Helena₁ saw Maria₂ in the cinema but she_{1,2} did not greet her.’
 c. O João₁ viu o Pedro₂ quando [-]_{1,2} entrou no cinema.
 ‘John₁ saw Pedro₂ when [-]_{1,2} entered the cinema.’
 d. O João₁ viu o Pedro₂ quando ele_{1,2} entrou no cinema.
 ‘John₁ saw Pedro₂ when he_{1,2} entered the cinema.’

(Adapted from Costa et al. 1998: 176-178)

Costa et al. 1998, 1999 assume that in coordination the null subject may be the trace of across-the-board (ATB) movement or *pro*, while in subordination only *pro* occurs.

Luján 1986 observed that, in Spanish, both in forward and backward anaphora, there is the same division of labor between null and overt pronouns. See examples in (53) and (54).

- (53) Cuando Juan₁ trabaja, *pro*₁/ él₂ no bebe.
 ‘When John₁ works, [-]₁/he₂ does not drink.’
 (54) Cuando *pro*₁/él₂ trabaja, Juan₁ no bebe.
 ‘When [-]₁/he₂ works, John₁ does not drink.’

²² For sentence (52c), Costa et al. 1998 considered that the matrix object may also serve as the antecedent of the embedded null pronoun, depending on the semantic properties of the verb. However, according to the results of their experimental study and other studies such as Madeira et al. 2012, Luegi 2012, the object antecedent is not the preferential interpretation in subordinate clauses of EP. The interpretative variation found in (52c) and (52d) also indicates that the pronominal resolution in subordinate clauses in EP is not exclusively constrained by syntactic factors, and is influenced by information of other linguistic levels, such as pragmatics.

In following works, Larson & Luján 1989, 1991 propose that there is quantifier raising of the overt pronoun in Spanish, which results in disjoint interpretation in backward anaphora. This property will be further discussed in section 2.5.

Calabrese 1986 also reported a similar property for Italian. As shown in (55), a null pronoun is interpreted as having its antecedent in subject position, while a stressed overt pronoun is preferentially interpreted as having a non-subject antecedent.

- (55) Quando Carlo₁ ha picchiato Antonio₂, *pro*₁/lui₂ era ubriaco.
 ‘When Carlo₁ hit Antonio₂, [-]₁/he₂ was drunk.’

Calabrese 1986 then considers that the different interpretative preferences of null and overt pronouns can be explained by the nature of the two types of pronouns. While the null pronoun is an unstressed pronoun, the overt pronouns *lei/lui* belong to the stressed pronoun type. In later studies, such as Cardinaletti & Starke’s 1999, the unstressed pronoun is defined as a weak pronoun, while the stressed one is defined as a strong pronoun. It is commonly believed that the weak pronoun is preferentially used to retrieve an old topic, while the strong pronoun is preferentially used to introduce a new topic or have a contrastive focus meaning.

Nevertheless, Chinese does not seem to present the same preference judgments as the Romance null subject languages do. According to some authors, including Huang 1984 and Yang et al. 1999, Chinese also easily accepts the coreference between the matrix subject and the subordinate overt pronoun, while the disjoint reading is also possible, see (56). Simpson et al. 2016 even found that the coreference between an overt pronoun and a subject is the preferred interpretation in some contexts. This fact implies that the Avoid Pronoun Principle in Chinese is not as restricted as in EP and other Romance null subject languages. It also implies that the pronoun interpretation in Chinese is more flexible than that in EP, in other words, for Chinese, sentences with null and overt pronouns may have the same interpretation, see (57):

- (56) Zhangsan₁ gaosu Lisi₂ ta_{1,2} de jiang le.
 Zhangsan tell Lisi he win prize Pst
 ‘Zhangsan told Lisi that he won the prize.’
 (57) Zhangsan₁ renwei [-]_{1,2} /ta_{1,2} neng ying.
 Zhangsan think he can win
 ‘Zhangsan thinks that (he) can win.’

The same situation has also been found in partial null subject languages. For example, in BP, though an embedded null subject must refer to the closest matrix subject that c-commands it, an embedded overt pronoun may convey the same interpretation, which is different from what happens in EP, see (58):

- (58) O João₁ disse que [-]₁/ele_{1,2} comprou um computador. (BP)
the João said that he bought one computer
'John₁ said that [-]₁/he_{1,2} bought a computer.' (Adapted from Barbosa 2013: 3)

Barbosa 2013, following Barbosa 1995, 2000, 2009, Pollock 1997, Alexiadou & Anagnostopoulou 1998, among others, proposes that, in EP and other consistent null subject languages, the original position of an overt subject is post-verbal, while the apparent pre-verbal subject in fact occurs in a left-dislocated clitic structure, which implies topic switch or emphasis. However, for BP, Barbosa 2013 proposes that the original position of an overt subject is exactly pre-verbal, located in the SpecTP position. Thus, the pre-verbal subject does not necessarily convey the topic switch or emphasis meaning and can be easily accepted in the topic maintenance conditions. This kind of structural difference between BP and consistent null subject languages can be adopted to explain the interpretative difference between these languages.

Another approach to explain this phenomenon is the typology of pronominals of Cardinaletti & Starke 1999²³, which separates pronouns into strong and weak pronouns. The stressed pronouns in Romance null subject languages, such as *lei* 'she' and *lui* 'he' of Italian, are considered as strong pronouns, which may involve topic switch. The unstressed pronouns in partial null subject languages (BP) or non-null subject languages (English and French) are considered as weak pronouns (Roberts 2007), which lack the function of topic switch. The null pronoun in consistent null subject languages also belongs to the category of weak pronouns, as it always conveys a topic continuation interpretation.

Both analyses may stand as an explanation for the interpretative preference of Chinese presented in (56) and (57) and connect this language to BP in terms of forward anaphoric

²³ The typology of pronominals was firstly proposed in 1994.

resolution. On the one hand, Chinese does not have the agreement system of consistent null subject languages, thus, its pre-verbal subject should appear in SpecTP, but should not be considered as occurring in left-dislocated clitic construction. Furthermore, the Chinese pronoun may function as a weak pronoun, which explains why it does not have the meaning of topic switch or emphasis.

Later, Frascarelli 2007 analyzes the difference between null and overt subjects under the concept of A-topic. Inspired by Givón 1983, Frascarelli 2007 considers that the A-topic serves to introduce for the first time or reintroduce a topic in the discourse. In this sense, the Avoid Pronoun Principle is redefined in the following way:

(59) ‘Avoid strong pronoun, whenever it agrees with the local Aboutness-shift Topic.’
(Frascarelli 2007: 719)

Frascarelli 2007 considers that in Italian the null pronoun is used when referring to the local A-topic²⁴, while the overt pronoun is used when referring to an entity which is different from the A-topic. Frascarelli’s 2007 theory corroborates what Chomsky 1981, 1986 had admitted, namely that the Avoid Pronoun Principle should have a discursive background, though it has a strict relation to the syntax. I will discuss this proposal in the following section.

2.4 The Aboutness-shift Topic

As Frascarelli’s 2007 proposal involves the notion of topic, I will firstly review some theories relating to topics and focus.

According to Lambrecht 1994, topic is the entity on which an utterance intends to comment, so it is what a sentence talks about. A topic can be (but is not always) the subject of a sentence. In Brito et al. 2003, topics which are at the same time the subject of a sentence are defined as unmarked topics, while non-subject topics are defined as marked topics (e.g. left dislocated structures). The position of a marked topic is in the left periphery of CP. This type of topic can be directly inserted into the CP position (60), or through A'-movement (61). Topics can refer to old information, but may also refer to a newly introduced entity, as in the case of the A-topic of

²⁴ According to Frascarelli 2007’s data, as long as a new A-topic is not introduced, all of the sequences of the sentences that follow an A-topic may take it as the local one.

Frascarelli 2007.

(60) Quanto ao João, gosto muito das obras dele.

‘As for John, I like a lot his works.’

(61) Os livros, já entreguei aos miúdos.

‘As for the books, I already handed them to the children.’

According to Zubizarreta 1998, focus is the non-presupposed part of a sentence. In other words, it is information that is not supposed to be shared by the speakers. As a result, focus is new information in many cases. There are two types of focus, namely contrastive focus and informational focus.

Contrastive focus implies that the information goes against the presupposition of the speakers. For example, in (62) *ao João* ‘to John’ is the contrastive focus and indicates that the books are handed exactly to ‘John’, and not to any other person. In this way, the contrastive focus may not refer to a new entity.

(62) AO JOÃO, entreguei os livros.

‘To John, I handed the books.’

The contrastive focus is normally located in the left periphery of CP. The other type of focus is informational focus, which, according to Belletti 2004, is located in the left periphery of vP. The informational focus normally conveys new information, which may serve as the answer to a question: see (63), where *o João* serves as the informational focus.

(63) Q: Quem é que comprou os livros?

‘Who bought the books.’

A: Comprou o João.

‘John bought (the books).’

Now we consider the A-topic of Frascarelli 2007. In juxtaposed structures, as in (64), the subject *o João* is a newly introduced entity which serves as the A-topic. As a result, the null subject in the following sentence may take it as its antecedent.

(64) O João₁ chegou à cidade. À tarde, [-]₁ visitou a universidade.

‘John₁ arrived to the city, In the afternoon, [-]₁ visited the university.’

According to Frascarelli 2007, it is in this way that null subjects in Romance null subject languages are identified. Holmberg 2010 also adopts this idea.

This proposal may also be extended to subordinate clauses. As in (65), the subject *o João* is assumed as the aboutness-shift topic, as a result, the embedded null subject should refer to this subject.

- (65) O João₁ disse ao Pedro₂ que [-]₁ ganhou o grande prémio.
 ‘John₁ told Peter₂ that [-]₁ won the grand prize.’

In (66), on the other hand, an overt pronoun is used in the subordinate clause to alter the topic. In this case the subordinate subject refers to a non-subject antecedent.

- (66) O João₁ disse ao Pedro₂ que ele₂ ganhou o grande prémio.
 ‘John₁ told Peter₂ that he₂ won the grand prize.’

Liu 2014 proposes that the null subject in Chinese should also function as its counterpart in Italian, meaning that the null subject should refer to the A-topic. Based on Lambrecht 1994, Liu 2014 proposes that every sentence should have a topic. For example, in (67), the subject of the first clause is the topic of the whole sentence and is located in the C domain.

- (67) [CP John₁ [TP [-]₁ hen congming]], suoyi [CP Top₁ [TP [-]₁ chang na diyi-ming]]
 John very intelligent consequently always get first-place
 ‘John is very intelligent, so he always gets the first place.’ (Liu 2014: 210)

The subject of the first clause has to agree with the closest topic. As a result, it should refer to ‘John’. Then, the null topic of the second clause also agrees with ‘John’ and that forms a topic chain. In the end, the null subject of the second clause agrees with the topic of the clause and also refers to ‘John’. Hence, Chinese is identical to Romance null subject languages in this respect.

However, the difference between Chinese and Romance null subject languages is the possibility of forming an extended topic chain. For the Romance null subject languages, once an A-topic is established, the head of the topic chain may establish an anaphoric relation long-distance, if there is no topic change. For Chinese, Liu 2014 argues that this language does not permit long-distance topic chains, as a null topic cannot agree with another null topic.

- (68) *[CP Top₁ [TP [-]₁ hen congming]], suoyi [CP Top₁ [TP [-]₁ chang na diyi-ming]]
 very intelligent consequently always get first-place
 ‘(he) is very intelligent, so he always gets the first place.’ (adapted from Liu 2014)

In (68), the topic of the first clause is null; as a result, this null topic cannot agree with the topic of the second clause. Consequently, the null subjects of the two clauses cannot be identified as well, which causes the oddness of the whole sentence. In this example, even if the topic of the first clause is identified through the previous context, it still fails to agree with the following topic, because a null topic in Chinese only agrees with an overt topic, in accordance to Liu 2014. This factor also implies that Chinese does not allow sequences of sentences in which only the first clause has an overt topic and all of the null subjects of the following sentences refer to this unique topic.

However, this phenomenon is very common in Romance null subject languages, including EP:

- (69) “Ela fechou a porta com um suspiro de alívio. [-] Tinha conseguido. [-] Olhou para cima certificando-se de que a casa continuava às escuras e silenciosa. [-] Olhou para o relógio, restava-lhe algum tempo, [-] tinha de voltar antes de darem por sua falta. A noite húmida e morna encobria-a. Só os candeeiros da rua iluminavam os seus passos. [-] Apressou-se. Estavam à sua espera dois pares de olhos famintos. [-] Tirou os poucos alimentos que [-] conseguira arranjar e deu-os às mãos ávidas que os recebiam.”

(Translation: She closed the door with a sigh of relief. [-] (she) had done it. [-] (she) looked up, making sure the house was still dark and silent. [-] (she) looked at her watch, there was some time left, [-] (she) had to go back before they noticed her absence. The humid, warm night covered her. Only the street lamps illuminated her steps. [-] (she) hurried. Two sets of hungry eyes were waiting for her. [-] (she) took the few foods [-] (she) could find and gave them to the greedy hands that received them.)

(corpus CAL2 of CLUNL, native speaker)

In Chinese, in order to reestablish the topic agreement, an overt pronoun can be used to save the wrong form of (68), under the theory of Liu 2014.

- (70) [CP Ta₁ [TP *pro*₁ mei-tian budan zhun-shi....., [CP TOP₁ [TP *pro*₁ haihui]]].
 he everyday not only on time but also
 ‘He not only (goes to school) on time every day, but also ...’
 (adapted from Liu 2014: 214)

In (70), the sentence initial pronoun *ta* ‘he’ can inherit its reference from a former context and

serves as the topic of the current sentence, thus it can agree with the *pro* of the first sentence. Afterwards, the topic *ta* ‘he’ also agrees long-distance with the null topic of the second clause, making it possible to further agree with the *pro* of the second sentence. In this way, with an overt topic, the sentence (70) is still grammatical in Chinese.

2.5 Forward and backward anaphora

Now we start to focus on the main topic of the thesis: anaphoric resolution in forward and backward anaphora. I would like to make a brief summary about anaphoric resolution in EP and Chinese.

As already shown in the previous sections, in forward anaphora of EP, a null subject prefers a subject antecedent while an overt pronoun prefers a non-subject antecedent, which has been tested in Costa et al. 1998, 1999, Luegi 2012, Madeira et al. 2012, Lobo & Silva 2016, Lobo et al. 2017, Costa & Ambulate 2010, among others.

(71) A avó₁ cumprimentou a menina₂ quando [-]₁/ela₂ chegou a casa.

‘The grandmother₁ greeted the girl₂ when [-]₁/she₂ arrived home.’

(adapted from Lobo et al. 2017)

As for backward anaphora (72), although there are not as many studies about it, Lobo & Silva 2016 and Lobo et al. 2017 have shown a similar interpretation as in forward anaphora, though the acceptance rate for the coreferential reading between overt pronoun and the subject antecedent is relatively higher than in forward anaphora.²⁵

(72) Quando [-]₁/ele₂ subiu à árvore, o polícia₁ viu o ladrão₂.

‘When [-]₁/he₂ climbed the tree, the police₁ saw the thief₂.’

(adapted from Lobo et al. 2017)

For Chinese, studies such as Huang 1984, Yang et al. 1999, Lust et al. 1996 and Zhao 2012, 2014 have shown that, in forward anaphora, both the null and overt pronoun may take a subject as their antecedent, which is different from EP, see (73):

(73) Zhangsan₁ chi le yige pingguo, zai [-]₁/ta_{1,2} kanshu de shihou.

‘Zhangsan₁ ate an apple, when [-]₁/he_{1,2} was reading the book.’

²⁵ A detailed description of Lobo et al.’ 2017 findings can be found in Chapter 5.

However, in backward anaphora, Chinese shows different interpretations for null and overt pronouns, as the former continue to take a subject antecedent, while the latter take a non-subject antecedent, as has been shown by Huang 1982, Lust et al. 1996, Zhao 2014.

- (74) Zai [-]₁/ta₂ kanshu de shihou, Zhangsan₁ chi le yige pingguo.
'When [-]₁/he₂ was reading the book, Zhangsan₁ ate an apple.'

In the remainder of this section, I will present some previous analyses of the syntactic structures of (71) to (74) and show how the structures may influence the interpretation of the null and overt pronouns.

Now we come back to the question of the relations between forward and backward anaphora in adverbial adjunct structures. One important point is that, either in forward anaphora or in backward anaphora, the coreference between the matrix subject and the pronominal subject of the subordinate clause does not violate any Binding Principles. Firstly, see the forward anaphora below.

In (75), the matrix subject may bind the pronoun in the adverbial adjunct.

- (75) O João₁ despiu o casaco quando ele_{1,2} chegou a casa.
the John took off the coat when he arrived to home
'John took off the coat when he arrived home.'

This example involves *Binding Principle B*. Since the matrix subject *o João* is in the matrix clause, which is out of the adverbial adjunct, the local domain of the pronoun *ele* 'he', the coreference between *o João* and *ele* 'he' does not violate *Principle B*, because this principle only requires the pronoun to be free in its local domain. As a consequence, in the structures of adverbial adjunct with forward anaphora, the coreference between the matrix subject and the embedded pronoun is allowed by the Binding Theory. The same is true for Chinese. However, this reading is marginal for consistent null subject languages like EP, but this property is not constrained by the Binding Theory, but by other factors (see Chapter 3).

It should be noticed that there is a structural difference between integrated and non-integrated subordinate adverbial sentences. Just as Canceiro 2016 pointed out, in forward anaphora of EP, it is more likely to establish a coreferential reading between the matrix subject and the

subordinate overt pronoun in integrated structures than in non-integrated structures. In an off-line preference judgement test, Canceiro 2016 tested the interpretation of null and overt pronouns in subordinate clauses, in both forward and backward anaphora. Here, I will only discuss the structures with overt pronouns to show the difference between integrated and non-integrated adverbial clauses. The results of Canceiro 2016 show that the native speakers of EP accept more easily the coreference reading between the matrix DP and subordinate overt pronoun in sentences such as (76) than in sentences such as (77):

(76) O Simão₁ jantou às 23h porque ele₁ saiu tarde do trabalho.

‘Simon₁ had dinner at 23h because he₁ left work late.’

(77) A Mariana₁ comprou um bolo, embora ela₁ estivesse a fazer dieta.

‘Mariana₁ bought one cake, although she₁ was on a diet.’

(Canceiro 2016: 120 and 123)

According to the data of Canceiro 2016, most of the native speakers accept the coreferential reading between the overt pronoun and the matrix subject in sentences such as (76), though around 30% participants accept both the coreferential and disjoint interpretation. However, for sentences such as (77), only 40% percent of native speakers accept uniquely the coreferential reading, while others accept uniquely the disjoint reading or accept both interpretations.

Canceiro 2016 then proposes that the integrated and non-integrated subordinate adverbial in EP have different syntactic structures. For integrated adverbials, the structure is presented in (78), with the adverbial clause as an adjunct to vP, while the structure of non-integrated adverbials is presented in (79), where the adverbial clause adjoins to TP.

(78) [TP matrix subject [T' ... [vP vP [CP [TP adverbial subject ...]]]]].

(adapted from Canceiro 2016: 136)

(79) [TP [TP matrix subject ...] [CP [TP adverbial subject ...]]].

(adapted from Canceiro 2016: 139)

However, the author noticed that, in both structures, there is a c-command relation between the matrix subject and the subordinate subject. This kind of c-command does not preclude the coreferential reading between the two subjects if the subordinate one is pronominal, since *Binding Principle B* only applies to constituents located in the same domain, which is not the case of (78) and (79). However, if the subordinate subject is a referential expression, then the

disjoint interpretation must be established, since *Binding Principle C* requires that a referential expression must be free everywhere.

There are some factors that should be noticed in Canceiro 2016. Firstly, the experimental data for integrated structures with overt pronouns may not completely support the Avoid Pronoun Principle, since the preferred interpretation is the coreferential reading, which goes against what is predicted by the Avoid Pronoun Principle or the data from studies like Costa et al 1998, 1999, Luegi 2012, Madeira et al 2012 and Lobo et al 2017. This kind of contrast may be explained by the fact that the subordinate clause analyzed in Canceiro 2016 is an adverbial causal clause, headed by *porque* ‘because’, which helps the participant to establish a coreferential reading between the matrix and the subordinate subjects.

Secondly, the sentences tested in Canceiro 2016 only have one nominal argument in the matrix clause which can serve as a potential antecedent for the embedded null or overt pronoun, while in many other studies (on EP or other null subject languages) the matrix clause contains two nominal arguments. This structural difference may also serve as an explanation for the differences in their results.

Furthermore, the test in Canceiro 2016 allows the participants to accept both the matrix subject and an extra-linguistic entity as the antecedent of the embedded pronoun, while in other studies the participants were not given this option. In this case, it is possible that some participants in Canceiro’s 2016 study have a preferred interpretation but also accept the other, so they chose to accept both interpretations. This fact may also influence the results of the test of Canceiro 2016.

Nevertheless, in any case, the structures proposed by Canceiro 2016 clearly reveal that there is a structural difference between integrated and non-integrated structures of subordinate clauses. Now we consider the case of Chinese, Lust et al. 1996 have shown that the temporal adverbial adjunct in Chinese can only be left-branched, unlike in the case of English. However, such an argument about Chinese is based on the fact that the structures tested in Lust et al. 1996 only involve temporal adjuncts without preposition, which is exemplified in (80) and (81):

- (80) Ta chifan de shihou, Milaoshu dai zhe yiding maozi.
he eat DE time Mickey Mouse wear Prg one.CL cap
‘When he was having meal, Mickey Mouse was wearing a hat.’

- (81) *Milaoshu dai zhe yi.ding maozi, ta chifan de shihou. (Afterthought)
 Mickey Mouse wear Prg one.CL cap he eat DE time
 ‘Mickey Mouse was wearing a hat, when he was having meal.’

If a preposition *zai* is inserted at the start of the adverbial adjunct, the sentence (81) can be well-formed as well, which can be exemplified in (82)²⁶:

- (82) Milaoshu dai zhe yiding maozi, zai ta chifan de shihou.
 Mickey Mouse wear Prg one.CL cap prep. he eat DE time
 ‘Mickey Mouse was wearing a hat, when he was having meal.’

The grammaticality of (82) indicates that the right-dislocated temporal adverbial adjunct can also be accepted in Chinese, since there is a preposition in the initial position of the subordinate clause. Then it is crucial to discuss if the structure in (82) is integrated or non-integrated. A negation test can be made to prove that the structure in (82) is, after all, non-integrated, see the example below.

- (83) Milaoshu meiyou dai zhe yiding maozi, zai ta chifan de shihou.
 Mickey Mouse not wear Prg one.CL cap prep. he eat DE time
 ‘Mickey mouse did not wear a hat, when he was having meal.’

A negation *meiyou* ‘not’ is inserted in the matrix clause, which forms the structure of (83). If (83) has indeed an integrated adverbial adjunct, the negation should have taken the whole sentence as its scope, negating also the subordinate clause, with the meaning of ‘Mickey Mouse did wear a hat, but not during the time when he was having dinner’. However, this interpretation is not available in Chinese, since (83) can only convey the meaning of ‘during the time he was having dinner, Mickey Mouse was not wearing a hat.’

The negation test has shown that the temporal adverbial adjunct in Chinese does not have an integrated structure, unlike in EP, since the equivalent sentence in EP with negation can convey the wide scope interpretation, see (84):

- (84) O João não comeu os bolos enquanto ele copiava os dados.
 the John not ate the cakes while he copied the data
 ‘John did not eat the cakes while he was copying the data.’

²⁶ Some detailed analyses about the structures with and without preposition in Chinese will be discussed in Chapter 4.

Sentence (84) may have two interpretations: negation with narrow scope, ‘during the time he copied the data, *João* did not eat the cakes.’; negation with wide scope which includes the subordinate clause, ‘*João* ate the cakes, but not during the time he copied the data.’

As a result, the difference between (83) and (84) illustrates that the apparently similar structures in Chinese and EP have different syntactic structures, with Chinese having a non-integrated structure and EP having an integrated structure.

Now we see the case of backward anaphora, where the adverbial adjunct clause precedes the matrix clause. In EP, taking a left-dislocated temporal adverbial adjunct clause as an example, it can be analyzed as a CP, where a temporal connector such as *quando* ‘when’ occupies the C⁰ position, while the structure below C⁰ consists of a TP, which is considered as a finite clause. It should be noted that Mória 2001 considers the temporal connector *quando* ‘when’ in EP as a relative pronoun (which functions as the head of a free relative), which is equivalent to ‘at the moment in which’²⁷.

In this case, the subordinate clause may be generated in two ways. For example, the adverbial adjunct clause in (85) may be moved from a lower position to the topic position, or be directly inserted in the topic position.

(85) Quando ele chegou a casa, o João chamou o Rui.
 when he arrived to home the John called the Rui
 ‘When he arrived home, John called Rui.’

In accordance with the proposals of Barss 1986, Duarte 1987 and Lasnik & Saito 1991 for topicalization, the fronted adverbial clause would derive the following representation, which involves movement from its original position to the left of TP:

(86) [CP [TP [CP [C Quando [TP ele chegou a casa]]]]₁,
 When he arrived to home
 [TP o João chamou o Rui [-]]₁]].
 the John called the Rui
 ‘When he arrived home, John called Rui.’

²⁷ As can be seen from Pan & Paul 2018, which will be discussed in Chapter 4, the Chinese temporal adverbial adjunct is also analyzed as a relative clause.

The temporal adverbial adjunct (CP) c-commands the matrix subject, while the embedded pronoun *ta* ‘he’ is located in a lower structure, but according to the cyclic c-command of Huang 1982, which is formulated in (90), *ta* ‘he’ may cyclically c-command the matrix subject.

(90) A cyclic c-commands B if and only if:

- a. A c-commands B, or
- b. If C is the minimal cyclic node (NP or S’) that dominates A but is not immediately dominated by another cyclic node, then C c-commands B. (Huang 1982: 394)

In Huang 1982, S’ corresponds to CP in the current framework, thus, in (89), the CP of the adjunct works as a cyclic node, which c-commands the matrix subject *Zhangsan*. The embedded *ta* ‘he’ is dominated by the CP, which is not immediately dominated by another cyclic node. As a result, the structure of (89) meets the condition (90b), which implies that the embedded pronoun *ta* ‘he’ may cyclically c-command the matrix subject *Zhangsan*. In this way, the coreferential reading between these two elements violates the *Principle C* of the Binding Theory. However, if the embedded subject is null, it is possible to obtain the coreferential reading between the embedded null subject and the matrix subject. Zhao 2014 applies another mechanism to explain this phenomenon, proposing that, in this case, the null subject is in fact a \emptyset topic, as initially proposed by Zhao 2012.

Another way to analyze the interpretation in (89) is to consider the notion of weak binding proposed by Huang, Li & Li 2009, according to which an NP₁ may bind an NP₂ if NP₁ is contained in an NP which binds NP₂. See the sentence (91):

- (91) *Ta*₁ *de* *mama*₂ *bu* *xihuan* *Zhangsan*₃.
 he Gen mother not like Zhangsan
 ‘His mother does not like Zhangsan.’

In this sentence, the pronoun *ta* ‘he’ in NP *ta de mama* ‘his mother’ does not c-command the second NP *Zhangsan*, but the coreferential reading between *ta* ‘he’ and *Zhangsan* is odd in Chinese. Huang, Li & Li 2009 propose that *ta* ‘he’ may weakly bind *Zhangsan*, since *ta* ‘he’ is contained in the NP *ta de mama* ‘his mother’, which c-commands *Zhangsan*. As a result, the coreferential reading between *ta* ‘he’ and *Zhangsan* violates Binding Principle C or D.

It is possible to consider that the left-dislocated adverbial adjunct clause is similar to the case

of (91), see the example in (92), which is equivalent to (89):

- (92) *Ta jin wu de shihou, Zhangsan tuo le waiyi.*
 he enter room DE time Zhangsan take off Pst jacket
 ‘When he entered the room, Zhangsan took off the jacket.’

Ta ‘he’ does not c-command *Zhangsan*, because *ta* ‘he’ is contained in the adverbial adjunct. But the adjunct as a whole c-commands *Zhangsan*, since it is topicalized. As a result, it is possible that *ta* ‘he’ weakly c-commands *Zhangsan*, as in (91).

In principle, (92) is different from (91) as the pronoun in (91) is contained in an NP, while the pronoun in (92) is located in an adjunct. However, in Teng’s 1985 proposal concerning weak c-command, an NP₁ contained in any maximal projection that c-commands NP₂ may weakly c-command NP₂. In this sense, *ta* ‘he’ is contained in an adjunct, which is a maximal projection. The adjunct c-commands *Zhangsan* and, as a result, *ta* ‘he’ also weakly c-commands *Zhangsan*. Consequently, the coreference between *ta* ‘he’ and *Zhangsan* should be precluded by the *Binding Principle C* or *D*.

In fact, a similar case for Spanish is analyzed by Larson & Luján 1989, who also investigated backward anaphora in Spanish applying the analysis of c-command. Firstly, they argued that the null subject of Spanish is equivalent to the unstressed pronoun of English, while the overt pronoun is equivalent to the stressed pronoun of English. For backward anaphora structures like (93a), the authors argued that the overt pronoun, which is considered as an emphatic pronoun, should be interpreted as disjoint from the matrix subject *Juan*. They proposed that the pronoun *él* ‘he’ undergoes movement in LF and is located in a higher position, see (93b).

- (93) a. *Cuando él trabaja, Juan no bebe.*
 when he works Juan not drink
 ‘When he works, Juan does not drink.’
 b. [CP [CP [NP *él*₁] [CP *cuando t*₁ *trabaja*]], [CP *Juan no bebe*]].
 *he*₁ when *t*₁ works Juan not drink

Since the pronoun in (93b) c-commands the matrix subject *Juan*, their coreference is not permitted in Spanish. The analysis of LF movement of Larson & Luján 1989 also corresponds to that of Lust et al. 1996, who investigate Chinese backward anaphora by using the theory of quantifier raising in LF. Thus, their analysis stands as an alternative way to explain why Chinese

has a distinct interpretation for forward and backward anaphora with overt pronouns.

Lust et al. 1996 firstly questioned if the proposal of cyclic c-command of Huang 1982 should be included into the Universal Grammar (UG). Based on an experimental study concerning L1 acquisition of Chinese, the authors used a picture truth value task²⁸ to test Chinese children and adults' interpretation of null and overt subjects in both forward and backward anaphora conditions²⁹.

The results show that the adult native Chinese speakers accept the coreference between the matrix subject and the subordinate subject (either null or overt) in forward anaphora. In backward anaphora, on the other hand, they accept the coreference interpretation between the matrix subject and the subordinate null subject, and the disjoint interpretation when having an overt subordinate subject, which corroborates the proposal of Huang 1982. However, for the Chinese children, they accept the coreferential reading in every condition of the test, including the backward anaphora condition with overt pronoun subjects.

Lust et al. 1996 analyzed the case in the following way: if cyclic c-command is part of UG, then the Chinese children should acquire this property from early on, since the adult speakers should always offer them input which precludes the coreference interpretation between the matrix subject and the overt pronoun in backward anaphora situations. But the Chinese children in their test did show a preference for coreference between the two constituents, which indicates that their grammar does not analyze the subordinate overt pronoun as in a c-command relation with the matrix subject. As a result, Lust et al. 1996 argued that there is no need to include the cyclic c-command into the UG. Besides, they proposed that the disjoint interpretation between overt backward anaphoric pronoun and the matrix subject made by the adult speakers should be acquired in later stage of acquisition.

In the second part of Lust et al.'s 1996 study, the authors discussed why there is a difference between forward and backward anaphora with overt pronouns in Chinese. Firstly, they proposed that, in Chinese, the full NP, overt pronoun, reflexive and null subject all have the structure of

²⁸ The picture truth value task proceeds in the following way: the participant reads or listens to a sentence and then is asked to tell if a given picture correctly describes the scenario indicated in the sentence.

²⁹ The sentences tested in Lust et al. 1996 had exactly the same structure as in (92), where the matrix clause, either in forward anaphora or in backward anaphora, only has one potential antecedent (namely the subject) for the embedded null or overt pronoun.

an NP. For the full NP, the head is clearly occupied by a noun form. For the overt pronoun, the NP head is null, while the pronoun occupies the SpecNP position. For the reflexive, *ziji* occupies the NP head position, while SpecNP is null. Finally, for the null subject, both the NP head and the SpecNP are empty. These structures are shown in (94):

- (94) [NP [-][N noun]] full NP
 [NP [-] [N *ziji*]] reflexive
 [NP *ta* [N [-]]] overt pronoun
 [NP [-] [N [-]]] null subject (adapted from Lust et al. 1996: 29)

It should be noticed that, in their analysis, the overt pronoun does not stand as a head, but is only located in a Spec position. Lust et al. 1996 then argued that, in all kinds of these NP structures, the whole NP should receive a phi-feature and a r(eferential)-feature, which can be achieved internally by some part of the NP or through external information, such as pragmatics. For example, in the case of the overt pronoun, the SpecNP *ta* can assign a phi-feature to the NP, while the r-feature of the NP should be achieved by external information. For reflexives and null subjects, both the phi-feature and the r-feature are achieved by external information.

The authors then argued that the overt pronoun in Chinese involves an obligatory [+focus] feature, which results in quantifier raising (QR) at LF; this analysis was first proposed by Larson & Luján 1991 for Spanish. In this way, Lust et al. 1996 propose that the Chinese overt pronoun *ta*, with the structure of [*ta* [-]] undergoes QR in LF, while the null subject, with the structure of [[-][-]], does not. Therefore, the structures involving forward and backward overt pronominal resolution in Chinese could be analyzed as in (95) and (96):

- (95) LF for backward anaphora
 [CP [CP [*ta* [-]] [CP [*t* *qi chezi de-shihou*]]] [CP *Milaoshu bei zhe yige shubao.*]]
 he t ride bike time-Rel Mickey Mouse carry Prg one-CL backpack
 ‘When he is riding a bike, Mickey Mouse is carrying a backpack.’

- (96) LF for forward anaphora
 [CP [*ta* [-]] [CP [CP *Tanglaoya kan dianshi de-shihou*] [CP [*t* *bao zhe yige piqiu.*]]]]
 he Donald Duck watch TV time-Rel t hold Prg one-CL ball
 ‘When Donald Duck is watching TV, he is holding a ball.’

In both (95) and (96), the pronoun *ta* ‘he’ undergoes QR in LF, leaving a trace in its original

position. In these structures, the raised pronoun is located in a higher CP position, which c-commands (thus, has scope over) the left dislocated adverbial adjunct. However, the matrix subject ‘Mickey Mouse’ is not included in the scope of the raised NP pronoun. In (96), on the other side, the raised pronoun is located in the highest CP tier, which takes the scope of the whole sentence. The authors then argued that, in (95), the interpretation of the trace (the determination of phi- and r-feature of the trace) is determined by the raised pronoun, and should not be related to the matrix subject, since it is outside of the quantifier scope. As a result, the trace should have a deictic reading inherited by the raised pronoun.

In (96), on the other hand, the NP ‘Donald Duck’ is also included in the quantifier scope. As a consequence, the interpretation of the trace can be either determined by the raised pronoun or by the NP ‘Donald Duck’, which legitimates the coindexation between the trace and the NP ‘Donald Duck’.

In summary, the analysis of Lust et al. 1996 is able to explain the asymmetry between forward and backward anaphora with overt pronouns in Chinese, without requiring the specialized device of cyclic c-command. For sentences with null subjects, the authors argued that there is no QR in this case³⁰, so those sentences can be analyzed in a normal way. Since there is no c-command relation between the subordinate null subject and the matrix subject, their coindexation is allowed by the Binding Theory.

This kind of analysis is able to explain the interpretative difference between forward and backward anaphora with overt pronouns in Chinese without changing the basic notion of c-command, which is believed by the authors to be superior to the notion of cyclic c-command proposed by Huang 1982. However, there is still some aspects which are not clear under Lust et al.’s 1996 proposal.

First, the authors propose that the pronoun is in fact located in the specifier position of an NP, instead of functioning as a head. Though this kind of analysis has its own advantages, it is not the normal way to analyze pronouns. For example, Déchaine & Wiltschko 2003 also consider the Japanese pronoun *Kare* ‘he’ as presenting an NP nature, but in their study, the pronoun is considered as an NP head.

³⁰ In fact, the empty categories behave differently from the quantifiers, since they cannot be focalized.

Second, though Lust et al. 1996 argued that the overt pronoun in Chinese has a [+focus] feature, this language in fact uses intonation to mark the emphatic pronoun. In other words, it is not possible to consider the unstressed overt pronoun *ta* ‘he’ in Chinese to be equivalent to the Spanish pronoun *él* ‘he’, since the Chinese *ta* ‘he’, if not stressed, does not imply any sense of emphasis or topic switch, in contrast with the Spanish overt pronoun³¹. This point of view may be further strengthened by the fact that, in forward anaphora structures, the null subject and the overt pronoun may convey the same meaning, with the coreference of the matrix subject and the embedded subject, see (97).³² So, it is questionable to consider that the unstressed Chinese pronoun must undergo QR.

- (97) Zhangsan₁ renwei [-]_{1,2} / ta_{1,2} neng de jiang.
 Zhangsan thinks he can win prize
 ‘Zhangsan₁ thinks that [-]_{1,2} / he_{1,2} can win the prize.’

Third, the way that Lust et al 1996 analyze the interpretation of the trace is also not canonical. As described above, the authors argued that the interpretation of a trace can be decided either by the original constituent which occupies the position of the trace, or by another constituent which is inside the quantifier scope of the raised constituent. However, in generative grammar, a trace is a gap left by movement of an element, and should be closely linked to the moved constituent. So, it is unclear why a constituent other than the original moved element (for example, the NP ‘Donald Duck’ in (96)) can serve as an antecedent of the trace.

Last, even if the interpretation of a trace can be decided by a constituent other than the original moved element, it is still unclear why in (95) the trace cannot refer to the matrix subject. In Lust et al.’s 1996 analysis, the main difference between (95) and (96) is the fact that the matrix subject of (96) is inside the quantifier scope of the raised pronoun, while in (95) it is not. So, the authors argued that in (95) the trace must have a deictic reading with the phi-feature determined by the raised pronoun, while in (96) the trace may also take the matrix subject as its antecedent. However, if the trace is determined by the raised pronoun, it can only determine the phi-feature of the trace, while its r-feature must depend on the pragmatic context. In this

³¹ In EP, on the contrary, it is the stressed overt pronoun that implies a topic maintenance interpretation, as shown in Costa & Matos 2012.

³² However, both the null and overt embedded subject can refer to an extra-linguistic referent.

sense, in (95), there is no pragmatic rule to prevent the trace from sharing its r-feature with the matrix subject ‘Mickey Mouse’, unless a rule regarding linear order is taken into consideration. But it is clear that the authors did not intend to lead the analysis in this direction. As a result, being inside or outside the quantifier scope should not count as an explanation for the less favored coreferential reading in backward anaphora with overt pronouns.

In fact, in my pilot Chinese test concerning the interpretation of this language³³, it has been found that native speakers of Chinese still accept the coreferential interpretation in the backward anaphora situation with overt pronouns, though their acceptance rate is only around 25%³⁴. This result also indicates that the interpretation of the pronoun should not be constrained only by syntactic factors. In fact, Chomsky 1981, 1986 considers that the Avoid Pronoun Principle should involve both syntax and pragmatics, which is corroborated by studies such as Costa et al. 1998. Later studies concerning L2 acquisition (Sorace & Filiaci 2006) or bilingual acquisition (Serratrice 2007) also consider the interpretation of null and overt pronouns as a discourse-pragmatic issue, and not as a merely syntactic issue. In this sense, I will propose a pragmatic analysis for this question in Chapter 4.

In summary, this chapter presents a review of the literature on some of the key factors concerning anaphoric resolution, namely Binding Theory, the Null Subject Parameter and the Avoid Pronoun Principle. Section 2.5 also argues that the structures discussed in this thesis do not violate the Binding Principles, which implies that anaphoric resolution should be influenced by other factors. In the next chapter, I will make a review of the studies that analyze anaphoric resolution under hypotheses about language processing.

³³ The pilot Chinese test was administered to a group of native speakers of Chinese and consisted of four conditions of backward anaphora in this language. Two conditions had prepositions in the adverbial adjunct (one with null embedded subject, the other with overt embedded subject pronoun) and the other two conditions did not have preposition (also one with null embedded subject, the other with overt embedded subject pronoun).

³⁴ The acceptance rate for the condition with preposition was 31.8%, while for that without preposition the percentage was 22.7%.

Chapter 3 Processing of anaphoric pronominals

In this chapter, I will present some of the theories and experimental studies which analyzed anaphoric resolution on the basis of language processing. The Accessibility Theory (in section 3.1), the PAH (in section 3.2), the Centering Theory (in section 3.3), the Advantage of First-mention Hypothesis (in section 3.4) and the active search mechanism (in section 3.5) will be discussed in this chapter.

3.1 The Accessibility Theory

Ariel 1990, 2001 aims to investigate referential dependence in terms of the processing costs, proposing the Accessibility Theory. She proposes that there is a relationship between the form of an anaphoric expression and its accessibility level, which is similar to the notion of givenness of Prince 1981, who describes givenness as *'the speaker assumes that the hearer can predict or could have predicted that a particular linguistic item will or would occur in a particular position within a sentence'* (Prince 1981: 228). Ariel 1990 then tried to prove that the notion of accessibility is superior to givenness when explaining anaphoric dependence.

Ariel 1990 firstly explained that the context could be divided into three types: the General or Encyclopedic Knowledge, the Physical Environment of the speech event, and the Linguistic Context. The Encyclopedic Knowledge refers to the information previously stored in our brain, which should not be changed according to the context. The Physical Environment of the speech event is related to the entities which are exactly involved in a physical circumstance of a speech or a dialogue. The Linguistic Context refers to the information that can be found intrinsically within the discourse of a dialogue, which is not recovered via the physical environment or the previous knowledge. As a result, in terms of memory structure, the information relating to those three contexts may also be stored in different memory structures. For example, the encyclopedic knowledge is stored in our long-term memory, while the linguistic context and the physical context are stored in our short-term memory. Then it is reasonable to consider that different anaphoric expressions may correspond to different memory structure(s).

Studies like Clark & Marshall 1981 noticed that pronouns normally refer to Linguistic Context, demonstratives refer to the Physical Environment, while proper names refer to Encyclopedic

Knowledge. However, Ariel 1990 found that this kind of correspondence could be too strong, as one anaphoric expression may correspond to different contexts. Thus, a more precise theory concerning the relationship between anaphoric expression and memory structure should be developed.

Ariel 1990 then suggested that there exists a degree of accessibility of the antecedent of an anaphoric expression. Ariel 1990: 16 pointed out that *it is the specific degree of Accessibility of mental entities attributed by the speaker to the addressee which is the crucial criterion determining the forms of retrieval marking*. Different anaphoric expressions correspond to entities of different degree of accessibility.

Ariel 1990 argued that there are four factors that may influence the accessibility of an antecedent. Distance: an antecedent gains high accessibility with a shorter distance between the antecedent and the anaphor, while it gains low accessibility with a longer distance; Competition: an antecedent gains high accessibility when it has fewer competitors to serve as the antecedent of an anaphor; Salience: an antecedent gains high accessibility if it is located in a salient position, such as topic position; Unity: an antecedent gains high accessibility if it is within the same frame, world, point of view, segment of paragraph of the anaphor.

She then proposes that there are three different groups of accessibility markers. Pronouns, for instance, serve as high accessibility markers, demonstratives serve as intermediate accessibility markers, while definite expressions serve as low accessibility markers.

For low accessibility markers, Ariel 1990 considers that these markers are normally used to refer to encyclopedic materials. Proper names, definite descriptions, etc. are believed to belong to this group of accessibility markers. However, there is also a scale inside the group of low accessibility markers, which indicates that those four factors (Distance, Competition, Salience and Unity) may also influence the degree of accessibility of the antecedent that should be retrieved by low accessibility markers.

Ariel 1990 then argued that the intermediate accessibility markers normally have a deictic or indexical value, which includes presentatives, demonstrative pronouns, 1st and 2nd person pronouns, and so on, which indicates the physical situation of a speech or dialogue. Like the low accessibility markers, there is also a scale inside the intermediate accessibility markers.

The high accessibility markers, in Ariel 1990's theory, are those which refer to text-dependent

entities. Null elements and third person pronouns are among the high accessibility markers, which also form a scale according to the degree of accessibility. Ariel 1990 also assumed that the high accessibility markers are the most common ones and are often used in subsequent contexts.

As a consequence, it is then possible to establish the general scale of all of the accessibility markers, which can be listed as following (Ariel 1990: 73, 2001:51):

(1) Full name+modifier > full name > long definite description > short definite descriptions > last name > first name > distal demonstrative+modifier > proximate demonstrative+modifier > distal demonstrative + NP > proximate demonstrative + NP > stressed pronoun+gesture > stressed pronoun > unstressed pronoun > cliticized pronoun > verbal person inflections > Extremely high accessibility markers (including empty categories)

It should be noticed that the scale of accessibility is highly related to some properties of the markers. A fuller or more complex anaphoric expression (such as proper names) often serves as a low accessibility marker, while an emptier or simpler anaphoric expression (such as empty categories) often serves as a high accessibility marker. Generally speaking, an emptier anaphor is less informative, less rigid and more attenuated, while a fuller anaphor is more informative, more rigid and less attenuated.

Thus, it is possible to explain the correspondence between the accessibility markers (anaphors) and their antecedent. Ariel 1990 believes that antecedents with high accessibility are often stored in short-term memory, while antecedents with low accessibility are often stored in long-term memory. The basic idea concerning the correspondence between an anaphor and its antecedent can be explained in the following way: if the speaker uses a less informative anaphor to refer to an antecedent, and the hearer can easily retrieve that antecedent, then the speaker has no need to use an anaphor with more information. For example, if the speaker intends to refer to a highly accessible antecedent, which is stored in the short-term memory, then there is no need for him/her to use a more informative anaphoric form, because an emptier anaphoric expression is already enough for the hearer to retrieve the antecedent. On the other hand, if the speaker intends to refer to a less accessible antecedent, then he should use an anaphor with more information, because, with a less informative anaphor, the hearer may not be able to retrieve that less accessible antecedent. This rule then implies that the high accessibility markers are

used to refer to antecedents with high accessibility, the intermediate accessibility markers are used to refer to antecedents with intermediate accessibility, while low accessibility markers are used to refer to antecedents with low accessibility. This stands as the nuclear part of the Accessibility Theory.

In this sense, Ariel 1990 believes that the Accessibility Theory and the scale of accessibility is universal³⁵ and may explain a series of linguistic phenomena, including Binding Theory and anaphoric resolution.

For *Principle A* of the Binding Theory, which requires that the antecedent of a reflexive must be within its local domain, Ariel 1990 considers that a reflexive serves as a high accessibility marker, so that it must refer to a high accessibility antecedent. An antecedent inside the local domain of the reflexive qualifies as a high accessibility antecedent since it satisfies two criteria, namely Distance and Unity, which was explained before.

For *Principle B* of the Binding Theory, which requires that a pronoun must be free in its local domain, Ariel 1990 considers that a pronoun should be a lower accessibility marker, compared with a reflexive. As a result, a pronoun should not refer to a very local antecedent, which is expected to be highly accessible, according to the criteria of Distance and Unity. For an antecedent outside the local domain of the pronoun, its accessibility is reduced so it is not problematic for it to serve as the antecedent of the pronoun.

For Principle C, which requires that a referential expression must be free everywhere, Ariel 1990 considers that an R-expression, usually a name, should not typically serve as an anaphor, due to its nature as a low accessibility marker.

Ariel 1990 also argues that the Accessibility Theory can be extended to sentence-level anaphoric resolution. She believes that the Avoid Pronoun Principle, which stipulates that a pronoun should be avoided whenever possible, should be replaced by the Accessibility Theory. Ariel 1990 proposes that, in null subject languages, a null pronoun, or more specifically, *pro*, serves as a high accessibility marker. When the speaker intends to retrieve a high accessibility entity, it is natural that a high accessibility marker should be used, according to the Accessibility Theory. This explains why a null pronoun is used to refer to a subject antecedent in Romance

³⁵ Of course, Ariel 2001 also pointed out that the Accessibility Scale should be adapted for each language.

null subject languages.

Taking EP as an example, in a sentence like (2), the matrix subject qualifies as a high accessibility entity, since it occupies the subject and topic position, which is supposed to be salient. As a consequence, the subordinate null subject (which stands as a high accessibility marker) should preferentially refer to a high accessibility antecedent, namely the matrix subject. On the other hand, the matrix object, which does not occupy a salient position, is supposed to be a less accessible antecedent, and so it should be preferentially recovered by a lower accessibility marker compared with the null pronoun, namely the overt pronoun. That explains why an overt pronoun should be avoided when referring to a salient entity.

- (2) O João disse ao Rui que [-]/ele ia receber o prémio.
'John told Rui that [-]/he would receive the prize.'

The example above involves intra-sentential structures. For inter-sentential structures, the interpretation is a little bit different. As reported by Morgado 2011, in EP, an overt pronoun may also refer to a subject antecedent if it is located in a previous sentence. For example, in juxtaposition structures, as in (3), coreference between the subject of the first sentence and the overt pronoun of the second sentence is possible³⁶. This phenomenon is in fact in line with the Accessibility Theory, as in the current structure the distance between the two entities is longer, and the two entities also do not belong to the same domain. As a result, the subject of the first sentence is less accessible compared with that of (2). Consequently, it is possible for a lower accessibility marker, namely the overt pronoun, to retrieve the subject.

- (3) No sábado passado, a Luísa reconheceu a Eunice no café. Durante a conversa, ela encomendou um bolo com chocolate.
'Last Saturday, Luísa recognized Eunice in the coffee shop. During the conversation, she recommended a chocolate cake.'
(Morgado 2011)

Concerning the case of backward anaphora, Ariel 1990 considers that there are two types of backward anaphora. One involves introduction of a new referent, while the other does not. Ariel argued that only the first case is a real backward anaphora, while the other one is in fact a case

³⁶ Of course, the overt pronoun of the second sentence may also take the object of the first sentence as its antecedent.

of forward anaphora. For example, in a sentence like (4), if the child is a newly introduced entity and ‘she’ refers to the child, then the structure stands as backward anaphora.

(4) When *she* was five years old, *the child of my acquaintance* announced a theory that she was inhabited by rabbits.

However, if the ‘child of my acquaintance’ had already been referred in the previous context, then the structure should not be counted as a backward anaphora, because the real antecedent of ‘she’ is not the ‘child of my acquaintance’ in the current sentence, but is located in the previous context. This phenomenon in fact implies that even in a sentence which apparently contains a backward anaphora structure, there is still a possibility that the anaphoric expression takes a forward anaphora interpretation.

Ariel 1990 also noticed that, for languages which allow null subjects, it is preferentially the null subject, and not the overt one, that allows the backward anaphora interpretation. Ariel only cited studies on Malayalam (Mohanam 1983) and Chinese (Xu 1986), but this kind of interpretation is also true for Romance null subject languages, as reported by Sorace & Filiaci 2006 and Serratrice 2007 for Italian, and Lobo & Silva 2016 and Lobo et al. 2017 for EP, see (5)³⁷:

(5) Quando [-]₁ chegou a casa, o avô₁ cumprimentou o menino₂.
when arrived to home the grandfather greeted the boy
‘When (he) arrived home, the grandfather greeted the boy.’

(adapted from Lobo & Silva 2016)

Here, Ariel 1990 also mentioned Biller-Lappin’s 1983 proposal that the use of a backward anaphora structure is a conventional way to introduce a new referent. I will discuss this claim when analyzing the Chinese backward anaphora in section 4.6 of Chapter 4.

Some phenomena related to the Accessibility Theory can be explained by the Informational Load Hypothesis of Almor 1996, 1999 and 2000, which considers the processing cost. This hypothesis tried to answer why the salient element should be retrieved by a less informative anaphoric expression (for example, a null pronoun).

Almor defined the informational load in terms of conceptual representation. He used C-

³⁷ Of course, in Romance null subject languages such as EP, it is still possible for an overt pronoun to appear in a backward anaphora structure.

difference to measure the informational load of an anaphor. For a pair consisting of an anaphoric expression and its antecedent, if the anaphor is more generic than the antecedent, then the C-difference is negative. The larger the semantic distance between them, the more negative the C-difference will be; if the antecedent is more generic than the anaphor, then the C-difference is positive. The larger the semantic distance between them, the larger the C-difference will be. As a result, the C-difference will always be larger when the anaphor is more specific than when it is more generic.

For example, just as Almor 1999 has shown, the C-difference of a pair of anaphora and its antecedent can be as follows: CD (the thing, a bird) < CD (the creature, a bird) < CD (the bird, a bird) < CD (the robin, a bird) < CD (the crippled robin, a bird)³⁸.

The informational load is directly related to the C-difference: a higher C-difference implies higher informational load. This kind of definition implies the following: it is not adequate to retrieve an antecedent for an anaphor which is more specific than the antecedent. For example, it is not possible to let *bird* to serve as the antecedent of *robin*, because *bird* is more generic than *robin*, which may cause an increased informational load. This hypothesis also indicates that different anaphoric expressions possess different processing costs. For a more generic anaphoric expression, the semantic distance between the expression and its antecedent also increases, which results in a lower informational load. In this sense, it is believed that the null pronoun is the most generic expression, and therefore, has the least informational load.

Almor proposes that it is only possible to use an anaphoric expression with higher informational load when it adds some new information or helps to identify the antecedent. This serves to explain why in null subject languages it is preferential to use a null pronoun to refer to a salient element, since the null pronoun has a lower informational load when retrieving the subject. The overt pronoun, which contains higher informational load than the null pronoun, must indicate some new information, which in many cases implies topic change.

3.2 The Position of Antecedent Hypothesis

Carminati 2002 investigated the processing of pronouns in Italian, a Romance null subject

³⁸ Here, CD equals C-difference. Inside each bracket, the first word is the anaphor and the second one is the antecedent.

language. She verifies that there could be a division of labor between the null and overt pronouns in this language. Inspired by the Accessibility theory, Carminati 2002 proposes the PAH, which is initially defined in the following way:

(6) The Position of Antecedent Hypothesis for the Italian null and overt pronouns in intra-sentential anaphora: the null pronoun prefers an antecedent which is in the Spec IP position, while the overt pronoun prefers an antecedent which is not in the Spec IP position.

(Carminati 2002: 57)

This hypothesis is based on the fact that the key factor that determines the preference of use between the null and overt pronoun is the prominence level of the antecedent. Carminati 2002 assumes that there is a scale of prominence between different antecedents. The null pronouns are preferentially interpreted as coreferent with the most prominent antecedent, while the overt pronouns are preferentially interpreted as referring to a less prominent antecedent. For intra-sentential structures in Italian, Carminati 2002 proposes that the prominence level of an antecedent is determined by its syntactic position, as the one in the subject position is considered as more prominent than those in lower positions. Since the canonical subjects appear in the SpecIP position, SpecIP becomes the most prominent position. As a consequence, a null pronoun should preferentially refer to an antecedent in the SpecIP position, while an overt pronoun should preferentially refer to an antecedent in other positions.

Carminati applied several on-line and off-line experiments to test the PAH. In most experiments, Carminati used sentences with a subordinate adverbial clause preceding the matrix clause. She explains that the preposed subordinate clause may help to maintain the structure information in the participants' memory. As for sentences with matrix-subordinate order, the matrix clause (which contains the antecedents) can be processed independently, therefore, when the subordinate clause (which contains the pronoun) is processed, the structure information of the matrix clause may have been already eliminated from the participants' memory. But for the subordinate-matrix order, when the embedded clause (which in turn contains the antecedents) is processed, the parser knows that a matrix clause would follow, thus the information structure in the subordinate clause can be kept until the processing of the matrix clause (which contains the pronoun) is completed. In such a way, Carminati considered that it is more adequate to use

the second order to test the PAH³⁹.

Experiment 1 was a self-paced reading test⁴⁰. The participants were asked to read sentences divided in 4 conditions, as shown in (7) and (8), which were disambiguated by pragmatic factors. After reading the sentence, the participants were asked to answer a question about the interpretation of the sentence.

(7) a. Dopo che Giovanni ha messo in imbarazzo Giorgio di fronte a tutti, [-] si è scusato ripetutamente.

‘After G. embarrassed G. in front of everyone, [-] apologized repeatedly.’

b. Dopo che Giovanni ha messo in imbarazzo Giorgio di fronte a tutti, lui si è scusato ripetutamente.

‘After G. embarrassed G. in front of everyone, he apologized repeatedly.’

(8) a. Dopo che Giovanni ha messo in imbarazzo Giorgio di fronte a tutti, [-] si è offeso tremendamente.

‘After G. embarrassed G. in front of everyone, [-] was very offended.’

b. Dopo che Giovanni ha messo in imbarazzo Giorgio di fronte a tutti, lui si è offeso tremendamente.

‘After G. embarrassed G. in front of everyone, he was very offended.’

(Carminati 2002: 69)

For the sentences (7a) and (7b), the pragmatics favors the reading that the subject of the second clause refers to the subject of the preposed subordinate clause, while for the sentences (8a) and (8b), the co-reference between the subordinate object and the subject of the second clause is pragmatically favored.

The results showed differences between reading times on the matrix clause among conditions. Considering the response time, it could be found that the reading time for (8a) was higher than that for (7a), which also resulted in a lower accuracy rate for the former. For the conditions (7b) and (8b), it took longer to read condition (7b) than in (8b), which also resulted in a lower accuracy rate for the former.

The results indicate that in Italian the speakers prefer to interpret the null pronoun as referring to a subject antecedent. When the pragmatic context does not favor this reading (8a), it takes longer to read the sentence and, therefore, this means that the reading costs are higher. On the

³⁹ It should be noticed that in the subordinate-main clause order, there is no c-command relation between the subordinate subject and the main subject, thus the anaphoric relation between them is not constraint by syntactic principles such as the Binding Theory.

⁴⁰ The self-paced reading paradigm will be explained in Chapter 7.

other hand, for the interpretation of the overt pronoun, the speakers prefer to interpret it as coreferent with a non-subject antecedent. When the pragmatics does not favor this reading (7b), the reading time also increases. The same phenomenon was also shown for the response time and the accuracy rating, as conditions (8a) and (7b) need more response time and also have a lower accuracy rating.

Experiment 2 was an off-line test which tested structures with main-subordinate order. The participants were asked to read sentences like (9a) and (9b) and then choose the preferred interpretation for the embedded pronoun. Afterwards, they were also asked to evaluate the chosen interpretation by using a rating scale from 1 to 5 (with 1 very unsure and 5 absolutely sure).

(9) a. Marta scriveva frequentemente a Piera quando [-] era negli Stati Uniti.

‘M. wrote frequently to P. when [-] was in the United States.’

b. Marta scriveva frequentemente a Piera quando lei era negli Stati Uniti.

‘M. wrote frequently to P. when she was in the United States.’

(Carminati 2002: 83)

In both (9a) and (9b), the matrix clause contains a subject and an object, while the subordinate clause contains a null (9a) or an overt (9b) pronoun. The tested subordinate adverbial clauses were if-clauses or temporal clauses.

The results showed a clear preference for the coreferential reading between the matrix subject and the subordinate null pronoun (9a) and for the coreferential reading between the matrix object and the subordinate overt pronoun (9b). The test also revealed that this preference was more relevant for the if-clauses than the temporal clauses.

Experiment 3 was also an off-line task, which tested complement clauses with the order of main-subordinate. Participants were asked to read sentences like (10a) and (10b) and then to choose the preferred interpretation for the embedded pronoun (10c). Afterwards, they were also asked to evaluate the chosen interpretation by using a rating scale from 1 to 5 (with 1 very unsure and 5 absolutely sure).

(10) a. Gregorio ha detto che [-] sarà presente al matrimonio di Maria.

‘G. has said that [-] will be present at the wedding of M.’

- b. Gregorio ha detto che lui sarà presente al matrimonio di Maria.
'G. has said that he will be present at the wedding of M.'
- c. A. Gregorio stesso sarà presente al matrimonio.
'G. himself will be present at the wedding.'
- B. Una persona diversa da Gregorio sarà presente al matrimonio.
'A person different from G. will be present at the wedding.'

(Carminati 2002: 91-92)

As can be seen, the matrix clause only contains one antecedent, namely the subject. The subordinate complement clause can be either a null or an overt pronoun.

The results show that, in both conditions, the speakers prefer to accept the coreferential reading between the matrix subject and the subordinate pronoun (null (97%) or overt (86%)). At first glance, it seems that the data go against the PAH, as the overt pronoun is also preferentially interpreted as referring to the matrix subject. However, there is a statistical difference between the two conditions, with condition (10a) exhibiting higher preference for the co-reference between the matrix subject and the pronoun. Carminati 2002 considered that the speakers would rather choose an antecedent which is referred in the utterance than a non-referred antecedent in the context, which results in the higher acceptance rate for the coreference reading between the overt pronoun and the matrix subject. However, the acceptance rate for this reading is still significantly lower than the coreferential reading between the null pronoun and the matrix subject, which continues to favor the PAH.

The first 3 experiments made by Carminati 2002 preliminarily prove the PAH for Italian. In Experiment 4, Carminati 2002 tested structures with dative subjects and the results proved that it is also easier for the null pronoun, rather than the overt pronoun, to retrieve the dative subject as its antecedent.

Carminati 2002 then began to test sentences that contain two SpecIPs, and reformulated the PAH:

(11) *The Revised Position of Antecedent Hypothesis for intra-sentential anaphora: the null pronoun finds an antecedent in the highest Spec IP position, while the overt pronoun prefers an antecedent elsewhere.*

(Carminati, 2002:109)

The revised version of PAH indicates that, for sentences with two or more SpecIPs, only the

highest one is the preferred antecedent for null pronouns.

In Experiment 5, Carminati tested sentences with expletive subjects (null and overt), comparing inter-sentential and extra-sentential structures. There are 3 groups of sentences in Experiment 5, which contain null expletive subjects, raising *seem* and existential-*there* respectively, see (12) to (14):

- (12) a. Quando fu confermato che Maria era completamente guarita, [-]/lei riprese a lavorare.
 ‘When it was confirmed that M. had completely recovered, [-]/she went back to work.’
 b. Quando Maria fu certa che era completamente guarita, [-]/lei riprese a lavorare.
 ‘When M. was certain that [-] had completely recovered, [-]/she went back to work.’
- (13) a. Siccome sembra che Alda sia brava in matematica, [-]/lei è stata scelta come tesoriere.
 ‘Since it seems that A. is clever at math, [-]/she has been chosen as a treasurer.’
 b. Siccome Alda sembra essere brava in matematica, [-]/lei è stata scelta come tesoriere.
 ‘Since A. seems to be clever at math, [-]/she has been chosen as a treasurer.’
- (14) a. Adesso che c’è Gianna disoccupata, [-]/lei non fa altro che lamentarsi.
 ‘Now that there is G. unemployed, [-]/she does nothing but complain.’
 b. Adesso che Gianna è disoccupata, [-]/lei non fa altro che lamentarsi.
 ‘Now that G. is unemployed, [-]/she does nothing but complain.’

(Carminati 2002: 132-134)

As can be seen from the examples, each sentence may have two variations, with the true subject in the higher SpecIP position or in the lower SpecIP position. Participants were asked to rate the acceptability of the sentences with null or overt pronoun in the second clause, with 1=very nature and 5=very awkward.

The results showed that, for group 1 and group 3, the null pronoun received a higher score when the true subject antecedent is located in the lower SpecIP position. Carminati considered that the expletive in the higher SpecIP may cause an interference effect for processing, as the antecedent of a null pronoun should have appeared in the place of the expletive. However, in this case, the sentences with an overt pronoun in the second clause were also not favored by the participants. Carminati 2002 then considered that the lower SpecIP is still too prominent to serve as the antecedent of the overt pronouns. As for group 2, with the raising *seem* structures, there are no significant differences between the two variations, which is a question that she left for future work. Experiment 5 also tested extra-sentential structures, where it is easier for the subject in lower SpecIP to be interpreted as the antecedent of the null pronoun in the second clause.

In Experiment 6, Carminati 2002 tested sentences with inversion between the subject and the verb in a self-paced reading test. In syntax, it is widely assumed that the inverted subject is in the SpecvP position, and not in the SpecIP position. As a consequence, it should not be considered as a potential antecedent for a null pronoun, which seeks an antecedent in the SpecIP position, which is occupied by an expletive.

The results also confirm the PAH, as it is harder for the participants to accept the coreference between a null pronoun and a post-verbal subject in the intra-sentential structures. Carminati 2002 claimed that, in both experiments 5 and 6, the parser firstly considered the expletive in the highest SpecIP position as the antecedent of the null pronoun. However, the semantic incompatibility of the expletive forced the parser to accept the subject in the lower SpecIP position as the antecedent of the null pronoun, which caused a ‘penalty’ in the processing. However, in the extra-sentential structures, this kind of penalty was not as obvious as in the intra-sentential structures, as the parser accepted much more easily the post-verbal subject as the null pronoun’s antecedent.

Then, in Experiment 7, Carminati 2002 tested the acceptability of continuations of predications about an entity or an event. In an off-line test, participants were asked to rate sentences starting with a clause which contained a pre-verbal or post-verbal subject. The following clause either contained an individual predication for the pre-verbal or post-verbal subject, or an event predication. When it was an individual predication, it contained a null subject. The results showed that sentences with a post-verbal subject and an individual predication had the worst rating, which can be interpreted as showing that the parser does not favor the post-verbal subject to be the antecedent of a null pronoun.

As a result, Experiments 5 to 7 in general proved the revised PAH, as the subject in a lower position does not serve as the preferred antecedent for a null pronoun, which seeks an antecedent in the highest SpecIP position.

From Experiment 8 to 11, Carminati 2002 began to investigate the hierarchy among phi-features and discussed how gender, number and person interact with the PAH. In Experiment 8, she tested sentences which can be disambiguated by the gender feature, taking (15) and (16) as an example.

- (15) a. Quando Mario chiama Liliana, [-] è contento.
 ‘When Mario calls Liliana, [-] is happy (masc).’
 b. Quando Mario chiama Liliana, lui è contento.
 ‘When Mario calls Liliana, he is happy (masc).’
 (16) a. Quando Mario chiama Liliana, [-] è contenta.
 ‘When Mario calls Liliana, [-] is happy (fem).’
 b. Quando Mario chiama Liliana, lei è contenta.
 ‘When Mario calls Liliana, she is happy (fem).’
 (17) a. Quando Mario canta, [-] è contento.
 ‘When Mario sings, [-] is happy (masc).’
 b. Quando Mario canta, lui è contento.
 ‘When Mario sings, he is happy (masc).’

(Carminati 2002: 204-205)

Sentences (15) and (16) are disambiguated by the gender agreement between the antecedent and the past participle. While (15) implies a subject antecedent for the pronoun (null or overt), (16) implies an object antecedent. Sentences with only one potential antecedent were also tested, which favor the subject antecedent reading, as shown in (17). Participants were asked to read the sentences in a self-paced reading test, and answer a question about the interpretation of the sentences afterwards.

The results showed that, even when the sentences can be disambiguated by gender, the parser still has a preference which favors the PAH. For example, for condition (15b), the reading time is much higher than that for (15a), which means that the parser does not prefer the overt pronoun to be retrieved by the matrix subject, even if it is forced to do so by the gender agreement. The same is true for (16a), when the null pronoun is forced to refer to the matrix object. For sentences with only one potential antecedent, the reading time is also shorter for the null pronoun condition. These findings suggest that even for the sentences which can be disambiguated by gender, the preference predicted by PAH was still detected in the processing. Then, from Experiments 9 to 11, Carminati tested sentences which can be disambiguated by number and person, and discussed if there is a hierarchy between the three types of phi-features. I will put these experiments aside as this question is outside of the scope of this study. Carminati also discussed the Overt Pronoun Constraint of Montalbetti 1984 and compared *PRO* and *pro* in the Experiments 12 to 14. However, these issues are not the object of the current study.

Generally speaking, the experiments in Carminati 2002 proved the PAH, which claims that the

processing of null and overt pronouns in Italian is primarily determined by the syntactic position of the antecedent. In addition to the first three experiments, which tested the canonical structures, Experiments 5 to 7 also proved that even an expletive in the highest SpecIP may be considered by the parser as the antecedent of a null pronoun, which makes it more difficult for the parser to accept the post-verbal subject antecedent. Thus, it causes a processing penalty. Experiment 8 then proved that the processing preference indicated by the PAH remains even in contexts where gender information forces an interpretation which runs counter to the PAH.

Carminati does not believe this strategy is a purely processing strategy, nor a purely syntactic constraint. Considering Grice 1975's maxims of quantity theory, Carminati 2002 assumed that PAH is a strategy at the interface level between the grammar and the discourse.

The PAH was also confirmed by other studies on other null subject languages, such as Spanish and Portuguese, though Carminati does not make this prediction. I will present the related studies in the following sections.

Generally speaking, the PAH corroborates the findings of the previous studies concerning the division of labor between null and overt pronouns in null subject languages, not only those of theoretical studies such as Calabrese 1986 and Brito 1991, but also those of experimental studies such as Costa et al. 1998, 1999, and is able to explain the related phenomenon in terms of processing costs and connect the antecedent resolution to the syntactic position.

The following experimental studies adopted the central ideas of PAH, but also considered some other factors. For example, Morgado 2011 additionally considered some semantic aspects, such as the thematic role of the antecedent, while Luegi 2012 and Fukumura & van Gompel 2015 also considered some aspects concerning the linear order, namely the order of reference, which will be discussed in section 3.4, when considering the Hypothesis of Advantage of First-mention.

3.3 The Centering Theory

Based on the theories of computational linguistics, the Centering Theory of Grosz et al. 1995 aims to analyze nominal and pronominal anaphora in terms of discourse coherence, which comprises two levels: the global coherence and the local coherence (Grosz & Sidner 1986). While the global coherence concerns the relationship between one discourse segment and other

segments, the local one concerns the coherence between the utterances of a single discursive segment. Grosz et al. 1995 focus on the latter case.

Grosz et al. 1995 use the term *center of an utterance* to refer to those entities that serve to link that utterance to other utterances in the discourse segment which contains it (p. 8). They also claim that it is an utterance and not a sentence in isolation that has centers. Thus, the same sentence uttered in different discourse situations may have different centers (p. 8).

Grosz et al. 1995 propose that there are two types of centers: *forward-looking centers* and *backward-looking centers*. Each utterance of a discursive segment may contain various *forward-looking centers*, which include all of the entities introduced into the utterance and may be retrieved in the following utterances. The utterances (except the first one) of a discourse segment also contain a *backward-looking center*, which serves as the most salient entity. There is a hierarchy between the *forward-looking centers*, according to their salience level. The most salient *forward-looking center* then serves as the *backward-looking center* of the next utterance. Each utterance can only have one *backward-looking center*, which also serves as a *forward-looking center* of the same utterance. It is also possible that the *backward-looking center* of an utterance becomes the *backward-looking center* of the next utterance once again, due to its high salience.

Grosz et al. 1995 defined three types of transition relations between utterances. The first one is *center continuation*, when the *backward-looking center* of an utterance is the *backward-looking center* of the previous utterance. At the same time, it is also the most salient *forward-looking center* of the current utterance when compared with other *forward-looking centers* of the same utterance. As a result, in this case, it may become the *backward-looking center* again in the next utterance. The second type is defined as *center retaining*, where the *backward-looking center* of an utterance is the *backward-looking center* of the previous utterance. However, in this case, the current *backward-looking center* is no longer the most salient center among the *forward-looking centers* of the current utterance. As a consequence, it may not become the *backward-looking center* again in the next utterance. The third type concerns the *center shift*, in which the *backward-looking center* of an utterance is not the *backward-looking center* of the previous utterance. See the examples below:

- (18) a. John has been having a lot of trouble arranging his vacation.
 b. He cannot find anyone to take over his responsibilities. (he = John)
 Backward-looking center = John; Forward-looking center = {John}
 c. He called up Mike yesterday to work out a plan. (he = John)
 Backward-looking center = John; Forward-looking center = {John, Mike} (continuation)
 d. Mike had annoyed him a lot recently. (him = John)
 Backward-looking center = John; Forward-looking center = {John, Mike} (retaining)
 e. He called John at 5 am on Friday last week. (he = Mike)
 Backward-looking center = Mike; Forward-looking center = {John, Mike} (shift)
 (Grosz et al. 1995: 19)

In sentences (a) and (b), there is only one *forward-looking center*, ‘John’, which also serves as the *backward-looking center* of (b). In (c), another *forward-looking center*, ‘Mike’, is introduced, while ‘John’ continues to serve as the *backward-looking center*. In this case, there is a *center continuation*, as ‘John’ is still the most salient center when compared with ‘Mike’. In (d), ‘John’ continues to be the *backward-looking center*, however, it is no longer the most salient one, when compared with ‘Mike’, hence, there is a *center retaining*. In (e), ‘Mike’ becomes the *backward-looking center*, which constitutes the *center shift*.

Grosz et al. 1995 then propose two centering rules. The Rule 1 states that no element in an utterance can be realized as a pronoun unless the *backward-looking center* of the utterance is realized as a pronoun also (p. 15), which means that the use of a pronoun to realize a *backward-looking center* signals to the hearer that the speaker is continuing to talk about the same thing (p. 15). See example (70), which violates this rule, according to Grosz et al. 1995.

- (19) a. He has been acting quite odd. (backward-looking center = John = he)
 b. He called up Mike yesterday. (backward-looking center = John = he)
 c. John wanted to meet him urgently. (backward-looking center = John; him = Mike)
 (Grosz et al 1995: 16-17)

Grosz et al 1995 consider that (19c) is not coherent, since ‘Mike’, which is not the *backward-looking center*, is realized by a pronoun, while the real *backward-looking center*, ‘John’, is not realized by a pronoun. The coherence of (19c) becomes much better if the sentence-initial ‘John’ is replaced by a pronoun ‘He’. Example (19c) clearly shows that if a pronoun is used to refer to an entity in the previous utterance, then that entity must be the most salient one, unless another pronoun has been used to refer to the most salient entity of the previous utterance. Rule 1 in

fact corroborates the Accessibility Theory: as predicted, it is not possible to use a less informative expression (pronoun) to refer to a less salient entity, while using a more informative expression (NP) to refer to a more salient entity inside the same utterance. Rule 1 also suggests that the use of a pronoun occurs in the case of continuation of referring to the same entity, in other words, it also implies that if a pronoun has been used to refer to an entity, then the most coherent way to continue to refer to the same entity is to use the same pronoun.

Rule 2 specifies that *center continuation* is relatively preferred to *center retaining*, while *center retaining* is relatively preferred to *center shift*. This rule in fact indicates that the most coherent form to construct a discourse segment is to always maintain the same *backward-looking center*. In other words, to keep the coherence of discourse, it is not preferred to change too often the topic which is talked about.

Generally speaking, the Centering Theory uses a different way from the Accessibility Theory or the PAH to explain nominal and pronominal anaphora. As described in the previous paragraphs, the Centering Theory is focused on the relationship between utterances inside a discourse segment, and the most important part of the theory relies on the possibility of *center continuation* or *center shift*.

It should be noted that the proposal of Frascarelli 2007 is, in some sense, similar to the Centering Theory. The aboutness-shift topic of Frascarelli 2007 is compatible with (but not equal to) the *backward-looking center* of Grosz et al. 1995. However, Grosz et al. 1995 mainly analyze examples from English, a non-null subject language. Concerning the null subject languages, there is a conflict between null and overt pronouns regarding their pragmatic uses. It should also be noticed that the notion of center is different from that of topic, though they share some similarities (since the *backward-looking center* refers to the most salient entity, which is often considered as a topic).

For example, in (18), the sentence in (c) is considered as *centering continuation*, which can be defined also as topic continuation in the theory of topic. However, (d) is definitely a topic shift situation, but, in the Centering Theory, it is only considered as *centering retaining*. And for (e), it is a topic continuation situation, but a *center shift* situation in Centering Theory. These differences are caused by the definition of centers and topics.

Nevertheless, in terms of topic change or center shift, the three types of transition relationships

of Grosz et al. 1995 are still consistent with the theory of Frascarelli 2007 in some sense, though they are not equivalent. Frascarelli 2007 proposes that the null subject is used to refer to the closest A-topic, while an overt pronoun is used to refer to a non-topic referent or implies a topic shift. Though a (aboutness-shift) topic may be related to a certain syntactic position (subject position), the key point of Frascarelli 2007 focuses on how to decide an A-topic through phonetic verifications and if there is topic change. In the same way, the Centering Theory focuses on how to decide a *backward-looking center* and if there is *center shift*. In this sense, both the Centering Theory and the theory of Frascarelli 2007 take the *center shift* or topic change as their main focus, which have an intimate link to pragmatic features. Though there may be some relationship between topic/center and syntactic position, these aspects should not be the only focus when analyzing anaphoric resolutions in general.

3.4 The Advantage of First-mention Hypothesis

Gernsbacher & Hargreaves 1988 and Gernsbacher 1989 aim to analyze the effect of order of reference when concerning the salience level of an entity. They proposed the Advantage of First-mention Hypothesis, which argues that the constituents which are mentioned first are more salient than those which are mentioned later. As a result, the first-mentioned constituents are more likely to be retrieved in anaphoric resolution. According to this hypothesis, *First-mentioned participants form the foundation of sentence-level representations, and therefore the remainder of the sentences is represented vis-à-vis those initial participants* (Gernsbacher & Hargreaves, 1988:701).

Gernsbacher & Hargreaves 1988 applied a probe-word task, in which the participants were asked to read sentences like (20). In (a), the first mentioned NP is inside a PP complement, while in (b) the first mentioned NP is the subject. After reading the sentences, participants were required to indicate if some of the constituents had appeared in the sentences. The results showed that it was easier for the participants to remember the first nominal constituents of the sentences than other constituents.

- (20) a. Because of Lisa, Tina was evicted from the apartment.
b. Tina was evicted from the apartment because of Lisa

Though Gernsbacher & Hargreaves 1988 argued that the Advantage of first-mention is a universal rule, some following studies have shown that there may be some interaction between advantage of first-mention and other factors and the effect of first-mention is only valid in certain situations.

For example, Luegi 2012 analyzed the forward anaphoric resolution in EP, considering the PAH and the advantage of first-mention. In several off-line and on-line tasks, the author concludes that both the syntactic function (which originates the PAH) and the order of reference (which is related to advantage of first-mention) may influence the pronominal (null and overt) interpretation in EP.

For example, Luegi tested sentences like the following: (21) has the canonical word order, where the oblique follows the subject, while (22) has the inverse word order, where the oblique is left-dislocated.

(21) O bombeiro perguntou pelo militar no quartel quando [-] recebeu a medalha de condecoração.

‘The fireman asked for the military in the barracks when [-] received the medal.’

(22) Pelo militar perguntou o bombeiro no quartel quando [-] recebeu a medalha de condecoração.

‘For the military the fireman asked in the barracks when [-] received the medal.’

(Luegi 2012: 143)

According to the PAH and the Advantage of First-mention Hypothesis, in (21), the subject *o bombeiro* ‘the fireman’ is in the SpecIP position and is the first entity to be processed, so it should be the most salient constituent and be recovered by the null subject of the adverbial adjunct. In (22), however, *o bombeiro* ‘the fireman’ is no longer the first-mentioned entity, though it remains in the subject position. In this case, the most salient constituent should be the oblique, as it is the first-mentioned entity, assuming the Advantage of First-mention Hypothesis. Thus, this involves a conflict between the PAH and the advantage of first-mention.

Luegi 2012 tested structures like (21) and (22) through questionnaires, the self-paced reading paradigm, the visual word paradigm and eye-tracking during reading, as well as sentences with an overt pronoun in the adverbial adjunct in some of the tests, such as (23) and (24). Though the tests with different methodologies may vary from each other, the basic results can be summarized as follows. In sentences like (21) and (23), when the subject is simultaneously the

first-mentioned entity, the participants prefer to accept the coreference between the subject and the null pronoun, and the coreference between the oblique and the overt pronoun. However, in sentences like (22) and (24), when the subject is not the first-mentioned entity, the participants may also accept the coreference between the null pronoun and the oblique or the coreference between the overt pronoun and the subject.

(23) O bombeiro perguntou pelo militar no quartel quando ele recebeu a medalha de condecoração.

‘The fireman asked for the military in the barracks when he received the medal.’

(24) Pelo militar perguntou o bombeiro no quartel quando ele recebeu a medalha de condecoração.

‘For the military the fireman asked in the barracks when he received the medal.’

(Luegi 2012: 143)

In this way, Luegi 2012 concluded that the syntactic function is not the only factor to decide the pronominal resolution in forward anaphora of EP, since there is an interaction between the syntactic function and the order or reference.

A similar study was made in English by Fukumura & van Gompel 2015, who aimed to analyze the processing of pronouns and repeated names in this language. In an eye-tracking task, the authors tested sentences like the followings:

(25) a. Barry was in debt like Sally. He had always struggled with the control of finances.

b. Like Barry Sally was in debt. He had always struggled with the control of finances.

c. Like Sally Barry was in debt. He had always struggled with the control of finances.

d. Sally was in debt like Barry. He had always struggled with the control of finances.

(Fukumura & van Gompel 2015: 503)

In (a) and (d), the first-mentioned entity is also the subject of the first sentence, while in (b) and (c) it is not. The results of the test showed that the participants accept more easily the coreference between the pronoun (of the second sentence) and the subject of the first sentence, regardless of whether there is a PP left dislocation. These results indicate that the processing of the structure in question is merely determined by the syntactic function of the antecedent, and the advantage of first-mention does not influence the pronoun resolution.

In another eye-tracking task, Fukumura & van Gompel 2015 tested the following sentences:

- (26) a. Barry was in debt like Sally. Barry had always struggled with the control of finances.
b. Like Barry Sally was in debt. Barry had always struggled with the control of finances.
c. Like Sally Barry was in debt. Barry had always struggled with the control of finances.
d. Sally was in debt like Barry. Barry had always struggled with the control of finances
(Fukumura & van Gompel 2015: 511)

In (a) and (d), the first-mentioned entity is also the subject of the first sentence, while in (b) and (c) it is not. The results of the test showed that the participants did not prefer the coreferential reading between the first-mentioned entity (either subject or PP) and the subject of the second sentence, while the effect of syntactic position is relatively weaker than in the test on pronouns. These results suggested that the advantage of first-mention has more influence on processing of repeated names. In summary, Fukumura & van Gompel 2015 have shown that the syntactic function and order of reference may influence the processing of pronouns and repeated names differently in English.

3.5 The active search mechanism

Continuing to consider the order in which information is perceived, studies like Kazanina 2005 and Kazanina et al. 2007 aimed to analyze a special case of anaphoric resolution, namely the backward anaphora, where the processing may be more complex than that in forward anaphora, because the cataphoric expression precedes its potential antecedents.

Based on Kazanina 2005, Kazanina et al. 2007 proposed that the processing of backward anaphora is constrained by the language universal active search mechanism, which is very similar to the processing of Wh-movement, where the strategy of filler-gap is applied (see earlier studies such as Crain & Fodor 1985, Frazier et al. 1983, Frazier & Clifton 1989 and Stowe 1986). For example, for a language with Wh-movement, when the parser encounters a Wh-constituent, a mechanism is activated to search for a gap, which is the base position of the Wh-constituent. In this process, due to memory limitations, the parser will preferentially choose the first potential gap position as the initial position of the Wh-constituent. As a result, it is predictable that, in (27), reading 1 is preferred over reading 2, since the gap in 1 is closer to the Wh-constituent than that of 2, which indicates that the Wh-constituent prefers to take scope over the matrix clause, rather than over the subordinate clause. In other words, readers may prefer to interpret the sentence as ‘when did John say that sentence’, rather than ‘when did Peter

buy the car, according to what John had said.’

(27) When did John say [-]₁ that Peter bought the car [-]₂?

Hence, Kazanina et al. 2007 consider backward anaphora resolution to be similar to Wh-interpretation in (27). In the case of left-dislocated adverbial adjuncts, the parser first encounters the subordinate pronoun, which is a referentially dependent form. Then the same mechanism as in (27) will be activated, as the parser tries to find the closest nominal element as the antecedent of the embedded pronoun, because of memory limitations. A self-paced reading test confirms this proposal, see (28).

(28) a. Because last semester while she was taking classes full-time Kathryn was working two jobs to pay the bills, Russell never got to see her.

b. Because last semester while she was taking classes full-time Russell was working two jobs to pay the bills, Erica never got to see her.

(adapted from Kazanina et al. 2007: 390)

Participants (native speakers of English) read sentences like (28a) and (28b), where a pronoun functions as the subject of the subordinate clause, while there are two names (distinguished by gender) which function as the subjects of the larger subordinate clause and of the matrix clause, respectively. Kazanina et al. 2007 found that the reading time of the first name (which is closer to the pronoun) was increased when there was a mismatch of gender between the pronoun and the name (28b). These results reveal that the readers prefer the coreferential interpretation between the cataphoric pronoun and the first nominal element (which is the closest) coming after it. Similar results were also found in van Gompel & Liversedge’s 2003 study, which also analyzed the backward anaphora of English and tested only overt pronouns in sentences with two referents.

Kazanina et al. 2007 also tested structures that involve *Binding Principle C*, such as in (29), where the pronoun of the first clause c-commands the subject of the subordinate clause (initiated by ‘while’). According to the binding theory, the pronoun should not have a coreferential interpretation with the embedded subject, and the self-paced reading test shows that the participants did not show a gender mismatch effect in this condition. The authors considered that the participants never thought the embedded subject to be the antecedent of the pronoun,

which resulted in the absence of effect. Thus, the authors argued that the active search mechanism is not applied in this case, since the embedded subject position is not available due to syntactic constraints. As a consequence, the active search mechanism proposes that, in backward anaphora conditions, the left-dislocated pronoun should take the first available nominal entity as its antecedent.

(29) Because last semester she was taking classes full-time while Kathryn/Russell was working two jobs to pay the bills, Erica felt guilty.

(adapted from Kazanina et al. 2007: 390)

Based on the active search mechanism, Kazanina & Philips 2010 analyzed the backward anaphora processing of Russian, which has a special structure initiated by the connector *poka* ‘while’. Kazanina & Philips 2010 stated that due to an idiosyncratic constraint of Russian, in backward anaphora the pronoun in a clause initiated by *poka* ‘while’ cannot refer to the matrix subject that appears after it. The authors then tested the processing of the *poka* structure in a self-paced reading task.

Three conditions were tested in this study, which are exemplified in (30). The first condition corresponds to the *Principle C* condition. The second condition corresponds to a backward anaphora condition with *poka* structure. The third condition corresponds to a normal backward anaphora condition, which was designated as the no-constraint condition. There may be gender match or mismatch between the sentence initial pronoun and the subject of the second clause in all of the three conditions.

(30)

Conditions	Russian	English translation
Principle C conditions, gender match/gender mismatch	Poskol’ku pered efirom onai prosmatrivala teksty soobsč’enij, poka Marina/Daniil grimmirovalas’/ grimmirovalsja k nac’alu s”emok, Zojai pervoj uznala sensacionnuju novost’.	“Since before the broadcast she looked through the news texts while Marina/Daniel put on makeup for the shoot, Zoja was the first one to learn about the sensational news.”

Poka conditions, gender match/ gender mismatch	Poskol'ku pered efirom poka onai prosmatrivala teksty soobs'č'enij, Marina/Daniil grimmirovalas'/ grimmirovalsja k nac'alu s'emok, Zojai sama opredelila porjadok reportaz'ej v vypuske.	“Since before the broadcast while she looked through the news texts Marina/Daniel put on make-up for the shoot, Zoja figured out the order of the reports in the program by herself.”
No constraint, gender match/gender mismatch	Xotja posle togo kak onai napisala zakazannuju stat'ju, Natas'ai/Mixail pravila/pravil tekst neskol'ko raz, Mixail/Natas'ai bol's'e vsego gordilsja svojim pervonac'al'nym variantom.	“Although after she wrote the commissioned article Natasha/Michael edited the text several times, Michael/Natasha was most proud of the original version.”

(Kazanina & Philips 2010: 377)

The results of Kazanina & Philips 2010 showed that there is no gender effect (Marina vs. Daniel) in the *Principle C* condition, since the parser never considers that the sentence initial pronoun may refer to the subject of the second clause. In the no-constraint condition, there is a gender-mismatch effect in the second clause (more time after Daniel), two words after the subject. The authors considered that this is due to the active search mechanism, indicating that the participants prefer to interpret the subject of the second clause as the antecedent of the sentence initial pronoun. For the *poka* condition, there is a gender-match (more time in Natasha vs. Michael) effect in the second clause, three words after the subject. The authors considered that this may also be caused by the active search mechanism. That is to say, the parser first considers the subject of the second clause as the antecedent of the pronoun when the gender information allows this interpretation, due to the active search mechanism. Afterwards, the idiosyncratic constraint on *poka* refuses such an interpretation, thus increasing the reading time and resulting in the gender-match effect. In this way, Kazanina & Philips 2010 concluded that the *Principle C* condition and the *poka* condition were processed in different ways, though both conditions do not allow the coreferential reading.

Kazanina et al. 2007 and Kazanina & Philips 2010 only investigate languages without null subjects (English and Russian), so it is unclear whether the active search mechanism is also extensible to languages with two pronominal forms, namely the null subject languages. Nevertheless, it is indeed possible to assume that their proposal can be extended to null subject

languages, since the embedded pronoun in dislocated adverbial adjuncts, either null or overt, is referentially dependent, and must find its antecedent as soon as possible. If this is true, then in null subject languages, even the overt backward anaphoric pronoun should choose the matrix subject as its antecedent, which should be different from the case of forward anaphora.

In fact, the universal active search mechanism of Kazanina et al. 2007 predicts that the key factor for backward anaphora is the proximity between the pronoun and its antecedent. A similar proposal, namely the active parser proposal of Sorace & Filiaci 2006, based on Kazanina 2005, also predicts that second language learners of Italian tend to accept the coreference reading between the embedded overt pronoun and the matrix subject in backward anaphora. However, the native speakers of Italian in Sorace & Filiaci's study did not show this preference, which was also corroborated by Serratrice 2007 and Belletti et al. 2007, see (31).

- (31) *Mentre lui₃ versa il vino nel bicchiere,*
while he pours the wine in+the glass
il cliente₁ paga il conto al cameriere₂.
the client pays the bill to+the waiter
'While he pours wine in the glass, the client pays the bill to the waiter.'
(Serratrice 2007: 230)

In a picture verification task⁴¹, Serratrice 2007 tested sentences like (31) with three groups of Italian speakers: monolingual adults, monolingual children and bilingual children who also speak English. The results showed that, in backward anaphora, the monolingual adults have a dominant preference for the subject antecedent in the case of the null pronoun, while such preference was reduced dramatically in the case of the overt pronoun. Serratrice 2007 also analyzed the individual performance of the participants, and the results showed that 75% of the monolingual adults chose more than three (out of five) times the extra-linguistic antecedent for the overt pronoun. Belletti et al. 2007 also used a picture verification task and had a similar result in regard to the interpretation of the overt pronoun in backward anaphora by native speakers of Italian.⁴²

⁴¹ the picture verification task presented in Chapters 3 and 5, proceeds in the following way: the participant reads or listens to a sentence and then chooses the picture (out of two or three) that best describes the scenario indicated in the sentence.

⁴² Sorace & Filiaci 2006, Serratrice 2007 and Belletti et al. 2007 also tested forward anaphora and compared the interpretation between null and overt pronouns. Some aspects of these tests will be presented in chapter 5.

The incompatibility between the interpretation chosen for the overt pronoun by adult Italian native speakers and the active search mechanism may be caused by the fact that Italian has two pronominal forms⁴³. For languages that only have one form, it seems the native speakers' interpretation corroborates the active search mechanism. For example, as discussed before, van Gompel & Liversedge 2003 tested backward anaphora by native speakers of English by using an eye-tracking task. The test sentences are exemplified in (32).

(32) When **he/she** was fed up, **the boy** visited **the girl** very often.

(adapted from van Gompel & Liversedge 2003:8)

As can be seen from (32), the embedded pronoun may be matched in gender with the matrix subject or with the matrix object. The results showed a longer reading time for the matrix verb when there is gender mismatch between the embedded pronoun and the matrix subject than when there is gender match. This finding reveals that the readers prefer to establish an anaphoric relation between the embedded pronoun and the matrix subject, rather than an anaphoric relation between the embedded pronoun and the matrix object. Van Gompel & Liversedge 2003 also administered two similar tests to corroborate the finding and concluded that, in real time processing, the computation of coreference relations is established earlier than the use of morphological information.

Another similar study has been made for L2 German acquisition. Drummer & Felser 2018 applied an eye-tracking task to test backward anaphora structures in German by native speakers and L2 speakers who speak Russian as L1. The test sentences are exemplified in (33).

(33) Als **er** im Krankenhaus war, fragte **Joseph/Sandra** gleich den Arzt nach Schmerzmitteln...
'When **he** was in hospital, **Joseph/Sandra** asked the doctor for pain killers straight away...'

The results reveal that both the L1 and the L2 speakers have longer reading times when there is gender mismatch between the embedded pronoun and the matrix subject than when there is gender match, which indicates that both the L1 and the L2 speakers prefer a subject antecedent for the embedded pronoun, which yields a backward anaphora interpretation. However, unlike

⁴³ However, some studies do not have the same results as those found in Serratrice 2007, for example, Fedele & Kaiser 2014 reveal that the Italian native speakers in their study chose subject antecedents for the overt pronoun in backward anaphora more than in forward anaphora, which may be influenced by the active search mechanism.

Kazanina et al. 2007, Drummer & Felser 2018 found in a later test that for the L1 and L2 speakers of German, the *Principle C* of Binding Theory is only applied as a late filter on potential coreference assignments. Since *Principle C* is not the focus of the current thesis, I will not discuss this aspect here.

In summary, this chapter reviews some studies on anaphoric resolution in terms of language processing. Some of the hypotheses presented in this chapter, such as the PAH, are designed to explain the interpretation of null and overt pronouns in forward anaphora of Romance null subject languages, while others, such as the active search mechanism, are set to explain the processing of backward anaphora. However, the proposals presented in this chapter do not address the question as to why there is an asymmetry between forward and backward anaphora in Chinese. So, in the next chapter, I will reanalyze Chinese anaphoric resolution on the basis of syntactic structure and topic chain theories.

Chapter 4 Re-analysis of anaphoric resolution of Chinese

In this chapter we come back to the question of forward and backward anaphor in Chinese. In order to figure out the reasons why Chinese displays some interpretative differences between forward and backward anaphora, it is necessary to make a detailed analysis of the possible structures of Chinese adverbial adjunct clauses. I will be based on the recent work of Pan & Paul 2018, which analyzed these structures under the MP. In the second part of this chapter, I will propose a topic chain analysis for the Chinese forward and backward anaphora, on the basis of the theory of Pu & Pu 2014.

4.1 Analysis of Pan & Paul 2018

Pan & Paul 2018 analyzed several adverbial adjunct structures of Chinese, including conditionals, causal clauses, inferential clauses, concessive clauses and temporal clauses. The authors made a clear claim that the default order of all these structures is subordinate-main order, where the adjunct clause is base generated in topic position, and the main-subordinate order is not generated by movement. Some of these structures may also allow the adverbial adjunct clause to be placed inside the main clause, namely between the main subject and the main verb. They also propose that the adverbial adjunct clause can be analyzed either as a specifier of a CP, or as a complement of a head final CP.

For instance, taking the conditionals as an example, Pan & Paul 2018 claim that they can have the following two structures, when using sentence final particles.

- (1) [TopP [cond.cl. Rúguǒ tā bù lái] [Top' [Top^o ne],⁴⁴
 if he not come top
 [TP wǒ jiù zìjǐ qù]].
 I then self go
 ‘If he doesn’t come, then I’ll go on my own.’

- (2) [TP_{main cl.} [CP_{cond.cl.} [TP Rúguǒ tā bù lái] [C ne]],
 if he not come sfp

⁴⁴ As can be seen from some of the examples in this chapter, the sentences cited from Pan & Paul 2018 contain tones for Chinese words, while the other Chinese sentences used in this thesis do not. Since the addition or not of the tones does not influence the comprehension of the structures in discussion, I did not delete the tones in the examples cited from Pan & Paul 2018, in order to maintain the consistency with the original sentences.

[TP_{main cl.} wǒ jiù zìjǐ qù]].

I then self go

‘If he doesn’t come, then I’ll go on my own.’

(Pan & Paul 2018: 64)

Here, the particle *ne* is a common word which often appears after a topic structure in Chinese. In either (1) and (2), the adverbial adjunct is located inside the topic structure of the main clause, which is base generated. In (1), the particle *ne* occupies the Top₀ position of the TopP, which takes the main clause as its complement. Thus, the adverbial clause is located in the Specifier position of the TopP. In (2), on the other hand, the particle *ne* occupies the C₀ position of a head final CP, which takes the conditional as its complement. Thus, the whole adverbial clause can be analyzed as adjoined to the main TP. The authors then claimed that both analyses are available and did not assume that one of them is superior to the other.

Pan & Paul 2018 argued that the adverbial-main order and the main-adverbial order are both independently generated, and none of them can be analyzed as derived from the other by movement. Based on previous studies, Pan & Paul 2018 claim that there are some meaning differences between the sentence initial and the sentence final adverbials, since sentence final adverbials often convey an afterthought meaning (Chao 1968) or are used to draw the attention of the hearer (Xing 2001).

Moreover, in an analysis of some functional words, it has been found that some functional words are only possible in sentence initial adverbial structures, but not in final ones. For example, in the matrix clause of conditional structures, the functional words *cai* ‘only then’ and *fouze* ‘otherwise’ can only appear in the sentence initial adjunct structures (cf. 3a, 4a), and are excluded from the sentence final adjunct structures (cf. 3b, 4b). Such differences also lead the authors to believe that the sentence initial adverbial structures are the default order in Chinese for the conditional clauses, inferential clauses, concessive clauses, causal clauses and temporal clauses, since the same functional word test holds for these adverbial structures. For ease of exposition, I will not list the same test for the other structures.

(3) a. Chúfēi nǐ qù, tā cái huìqù.

only if you go he only.then will go

‘He will only go, if you go.’

(Pan & Paul 2018: 84, cited from Lǚ 2000: 125)

b. Tā (*cái) huì qù, chūfēi nǐ qù.
 he only.then will go unless you go
 ‘He will go only if you go.’ (Pan & Paul 2018: 87)

(4) a. Chūfēi xià yǔ, fǒuzé tā shàng bān
 only.if fall rain otherwise he go work
 cónglái bù zuò chē.
 ever not sit bus
 ‘Unless it rains, he never takes the bus to go to work.’
 (Pan & Paul 2018: 84, cited from Lǚ 2000: 125)

b. (*Fǒuzé) tā shàng bān cónglái
 otherwise he go work ever
 bù zuò chē, chūfēi xià yǔ.
 neg sit bus unless fall rain
 ‘He never takes the bus to go to work, unless it rains.’ (Pan & Paul 2018: 87)

Pan & Paul 2018 also analyzed the temporal adverbial clauses in Chinese. There are two types of such structures: one is headed by *de shihou*, while the other is headed by postpositions such as *yiqian* ‘before’, *yihou* ‘after’. For the case of *de shihou*, the authors claim that such a structure can be considered as a relative clause, which can be translated as ‘the time in which’.

(5) Tā dào Běijīng de shíhou, tiānqì bù tài hǎo.
 he arrive Beijing_{sub} time weather not too good
 ‘When he arrived at Beijing, the weather was not good.’
 (= ‘The time when he arrived at Beijing, ...’) (Pan & Paul 2018: 103)

Since Chinese relative clauses are generated by the functional word *de*, and *shihou* in Chinese means ‘time’, ‘moment’, sentence (5) can be literally translated as ‘the time in which he arrived Beijing’, which suggests that the temporal adverbial adjunct has in fact an NP or DP nature. In this case, the subordinate clause *ta dao Beijing* ‘he arrived Beijing’ can be analyzed as a TP that serves as the complement of *de*, a head-final C. It is also possible to add a preposition before the NP/DP, making the temporal adverbial adjunct a PP, where the NP/DP temporal clause becomes the complement of the preposition *zai* ‘at’. (see 5’).

(5’) Zai ta dao Beijing de shíhou, tiānqì bù tài hǎo.
 prep. he arrive Beijing_{sub} time weather not too good
 ‘When he arrived at Beijing, the weather was not good.’
 (= ‘At the time when he arrived at Beijing, ...’)

It should be noticed that *shihou* may also be used in a simple adverbial adjunct which does not contain a clause, such as *zai zhege shihou* ‘at this moment’. It is also worth noting that another preposition *dang* may be used to generate a temporal adverbial adjunct. In this case, its complement must be a complex DP which contains a clause, but not a simple NP/DP.

- (6) a. *Dang ta dao Beijing de shihou.*
 prep. he arrive Beijing_{sub} time
 ‘When he arrived at Beijing’ (= at the time when he arrived Beijing)
- b. **Dang zhege/nage shihou.*
 prep. this that time
 ‘at this/that time’ (Pan & Paul 2018: 104, cited from Lü 2000:149)

Another type of temporal adverbial adjunct is headed directly by a postposition, such as *yiqian* ‘before’, *yihou* ‘after’. In this case, the subordinate clause serves as the complement of such postpositions, which clearly generates a head final structure, as illustrated by its definition. It is also possible to have a pure NP/DP as the complement of the temporal postpositions.

- (7) a. *Ta du daxue yihou*
 he study university after
 ‘After he started to study at university.’
- b. *Chuntian Yiqian*
 spring before
 ‘Before spring.’

It is also possible to add a preposition before the whole postposition structure, with the exception of some postpositions, such as *yilai* ‘since’.

- (8) a. *Zai ta du daxue yihou*
 prep. he study university after
 ‘After he started to study at university’
- b. *Zai chuntian yiqian*
 prep. spring before
 ‘before spring’

Pan & Paul 2018 point out that the adverbial clause in Chinese can actually be placed in three positions: in sentence initial position, inside the sentence (between subject and verb) and in sentence final position. They propose that the sentence initial adverbial clauses can either be analyzed as a base generated Topic, or as adjunct to a TP structure, depending on how

topicalization structures are analyzed (see examples 1 and 2). For inside sentence adverbial clauses and sentence final adverbial clauses, they claim that these sentences are adjuncts to the matrix TP (see 9).

- (9) [TP subject [TP adjunct [vP main verb ...]]. (for inside sentence adverbial adjunct)
 [TP subject [vP main verb ...][TP adjunct]]. (for sentence final adverbial adjunct)

The authors also observed that the temporal adverbial adjunct clauses may also be placed between the matrix subject and the matrix verb, either above or below the auxiliaries⁴⁵. However, in this case, a preposition is obligatory if the adjunct has an overt subject, otherwise the sentence is ungrammatical.

- (10) Tā [PP *(zài) [DP wǒ shàng bān de shíhòu]]
 he at I attend work DE time
 qù-le yóujú.
 go-Perf post office
 ‘He went to the post office while I was working.’ (Pan & Paul 2018: 133)

Pan & Paul 2018 also discussed the nature of the functional words which can initiate an adverbial clause. After considering several parsing possibilities, they propose that functional words which initiate an adverbial clause, such as *ruguo* ‘if’ (for conditional), *suiran* ‘although’ (for concessive clause) should be analyzed as a C-head, rather than an adverb. See the examples below (11), where a constituent cannot be extracted from a conditional or concessive clause headed by *ruguo* ‘if’, *suiran* ‘although’, which indicates that such clauses should be considered as adjunct islands. Thus, the functional words that head these clauses should be analyzed as a C head, rather than an adverb. Otherwise, the island conditions cannot be explained.

- (11) a. *[TopP Zhè-ge xuéshēng [CP [C° rúguo]
 this-CL student if
 [TP Zhāngsān dǎ zhè-ge xuéshēng]]
 Zhangsan beat this-CL student
 [Main cl. Xiǎohóng jiù huì hěn bùgāoxìng]. (conditional)
 Xiaohong then will very unhappy
 (Intended: ‘This student, if Zhangsan beats (him), Xiaohong will be very unhappy.’)

⁴⁵ Pan & Paul 2018 observed that conditional and causal clause can be placed between the matrix subject and the verb, while concessive clauses cannot. I will not discuss these structures here, since they are not the focus of the thesis.

(Pan & Paul 2018: 142)

- b. *Mǎli [CP suīrán Zhāngsān piàn Mǎli],
 Mary although Zhangsan cheat Mary
 [Matrix-TP dàjiā háishì rènwéi Zhāngsān shì ge hǎo rén.
 everyone still think Zhangsan be CL good person

(‘Although Mary, Zhangsan cheats on (her), everybody still thinks that Zhangsan is a good person.’)
 (Pan & Paul 2018: 145)

Pan & Paul 2018 made a detailed analysis of the structures of sentence initial and inside sentence adverbial clauses, focusing on the nature of the sentence initial subject. They firstly illustrated some parsing possibilities of the structures under evaluation.

- (12) Structure A: DP conjunction Ø..., [main clause DPpron]
 Structure B: DP conjunction Ø..., [main clause Ø]
 Structure C: DP conjunction DPpron..., [main clause Ø]
 Structure D: DP conjunction DPpron..., [main clause DPpron....]

Then they analyzed all the parsing possibilities for the sentence initial subject, taking examples from conditional clauses⁴⁶.

For structure A, the sentence initial subject can be analyzed as a matrix topic or as a subordinate subject which undergoes extraction.

- (13) DP conjunction Ø..., [main clause DPpron]
 Zhangsan ruguo e le, ta hui qu mai dongxi de.
 Zhangsan if hungry Perf. he will go buy thing DE
 ‘If Zhangsan is hungry, he will go and buy something.’ (Pan & Paul 2018: 120)

a. [matrixTopP Zhāngsān₁ [[adv.cl. [C° rúguǒ] [TP *pro*₁....]] [main cl.TP tā₁]]
 Zhangsan if he

b. [matrix cl [adv.cl. Zhāngsān₁ [rúguǒ [TP *pro*₁]]] [main cl.TP tā₁]
 Zhangsan if he

c. [matrix cl [adv.cl. Zhāngsān₁ [rúguǒ [TP-Zhāngsān₁]]] [main cl.TP tā₁]
 Zhangsan if he

In (13a) and (13b), the main topic *Zhangsan* or subordinate topic *Zhangsan* can control the subordinate null subject via GCR, while, in (13c), the subordinate subject *Zhangsan* must be

⁴⁶ When analyzing these structures, the authors still maintained the possibility that *ruguo* ‘if’ is an adverb, which was denied in the later discussion. For ease of exposition, I will only present the parsing possibilities which are compatible with the analysis of *ruguo* ‘if’ as a C head.

extracted from its original position to SpecCP of the subordinate clause, by assuming there is exemption of island violation under specific conditions, namely in case of non-episodic eventuality, as proposed by Zhang 2002 and Pan 2014⁴⁷.

For structure B, the sentence initial subject can be analyzed in four different ways: (i) a base generated matrix topic, (ii) a matrix or subordinate topic generated by movement out of the subordinate clause, (iii) a matrix topic generated by movement out of the matrix clause, (iv) the matrix subject.

(14) DP conjunction \emptyset ..., [main clause \emptyset ...]

Zhāngsān rúguǒ è -le, [huì qù mǎi dōngxi de].

Zhangsan if be.hungry-perf will go buy thing DE

'If Zhangsan is hungry, he will go and buy something.'

(Pan & Paul 2018: 124)

a. [matrixTopP Zhāngsān₁ (,)[[adv.cl. rúguǒ *pro*₁.....] [main cl.TP *pro*₁]]]

Zhangsan if

b. [MatrixTopP DP₁ [Adv.cl. Conj \overline{DP} ₁ + vP], [Main cl. *pro*₁ + vP]]

c. [Adv.cl. [TopP DP₁ [Adv.cl. Conj \overline{DP} ₁ + vP]]], [Main cl. *pro*₁ + vP]

d. [Matrix TopP DP₁ [Adv.cl. Conj *pro*₁ + vP], [Main cl. \overline{DP} ₁ + vP]]

e. [TP DP₁ [vP [Adv.cl. Conj *pro*₁ + vP] vP]]

In (14a), *Zhangsan* is the matrix topic, which binds the subordinate and matrix subjects via GCR and shares its index with the two null subjects. In (14b) and (14c), the DP (*Zhangsan*) can be considered as extracted from the subordinate clause, either to the matrix topic position or to the subordinate topic position, which can be legitimated by the non-episodic eventuality conditions, as discussed above⁴⁸.

In (14d), the DP (*Zhangsan*) is also located in the matrix topic position, generated by movement from the matrix clause and, in (14e), the DP (*Zhangsan*) is simply the matrix subject, which indicates that the subordinate adverbial clause is in fact an inside sentence adjunct clause.

For structure C, the sentence initial subject can be analyzed as a matrix topic, either base generated or generated by movement from the matrix clause, or as the matrix subject, which indicates again that the subordinate clause is in fact an inside sentence adjunct.

⁴⁷ The proposals of non-episodic eventuality will be discussed in section 4.3.

⁴⁸ Another way to analyze (14c) is to consider that the subordinate topic is base-generated and controls the subordinate subject. In this case, the subordinate subject is again *pro*, as in the case of (13b).

(15) DP conjunction DPpron..., [main clause Ø]

Zhāngsān rúguǒ tā zhēnde xiǎng qù dehuà,
 Zhangsan if he really want go C(-root)
 [Ø kěndìng huì qù de].
 certainly will go DE

‘If Zhangsan really wants to go, then he will go for sure.’ (Pan & Paul 2018: 126)

a. [Matrix TopP DP₁ [[Adv.cl. Conj tā₁ + vP](,) [Main cl. *pro*₁ + vP]]]

b. [Matrix TopP DP₁ [[Adv.cl. Conj tā₁ + vP](,) [Main cl. $\overline{\text{DP}}_1$ + vP]]]

c. [TP DP₁ [vP [Adv. cl. Conj tā₁ + vP] vP]]

In (15a), the DP (*Zhangsan*) is base generated in the matrix topic position, which shares the index with the subordinate pronoun and the matrix null subject. In (15b), the DP (*Zhangsan*) is extracted from the matrix subject position to the matrix topic position. In (15c), the DP (*Zhangsan*) simply serves as the matrix subject, indicating that the subordinate clause is also an inside sentence adjunct.

Finally, for structure D, the sentence initial subject can be analyzed as a matrix or subordinate topic.

(16) DP conjunction DPpron..., [main clause DPpron....]

Zhāngsān rúguǒ tā è -le, [ta huì qù mǎi dōngxi de].
 Zhangsan if he be.hungry-perf he will go buy thing DE

‘If Zhangsan is hungry, he will go and buy something.’ (Pan & Paul 2018: 127)

a. [MatrixTopP DP₁ [[Adv.cl. Conj tā₁ + vP], [Main cl. tā₁ vP]]]

b. [Adv.cl. [TopP DP₁ [Adv.cl. Conj tā₁ + vP]]], [Main cl. tā₁ vP]

In (16a), the DP (*Zhangsan*) is a base generated matrix topic, which shares its index with the following two pronouns. In (16b), the DP (*Zhangsan*) is a base-generated subordinate topic.

As can be seen from (13) to (16), there are several ways to parse the Chinese sentence initial adverbial clauses. Since Pan & Paul 2018 only took the conditional clause examples, and only analyzed structures with a forward anaphora interpretation, I will make a similar analysis for temporal adverbial clauses, which are the topic of this thesis, and also taking into account the contrast between forward and backward anaphora.

Before starting the analysis, I would like to compare the inside sentence temporal adverbial adjunct clause and the right dislocated temporal adverbial adjunct clause regarding its status as an integrated/central or a non-integrated/peripheral constituent (following Lobo 2003,

Haegeman 2012 and Canceiro 2016, among others), concerning the matrix sentence negation scope. See example (17a), which yields an inside sentence adjunct, which is below the scope of the negative word *meiyou* ‘have not’. Since this sentence allows the interpretation that *Zhangsan* ‘went to the school at another moment’, the adjunct can be analyzed as an integrated adjunct.

- (17) a. Zhangsan meiyou zai Lisi chifan de shihou qu xuexiao.
 Zhangsan not prep. Lisi eat DE time go school
 ‘Zhangsan did not go to school when Lisi was having meal.’
 b. Zhangsan meiyou qu xuexiao, zai Lisi chifan de shihou.
 Zhangsan not go school prep. Lisi eat DE time
 ‘Zhangsan did not go to school, when Lisi was having meal.’

However, in (17b), with a right dislocated temporal adverbial adjunct, the sentence does not necessarily transmit this information, which indicates that the structure in (17b) is a non-integrated adverbial adjunct structure. Thus, it also proves that there is no reconstruction effect here, hence (17b) is not derived by movement from (17a). The same is also true for the sentence initial temporal adverbial adjunct, which also does not transmit the information that *Zhangsan* ‘went to the school at another moment’ (cf. 18).

- (18) Zai Lisi chifan de shihou, Zhangsan meiyou qu xuexiao.
 prep. Lisi eat DE time Zhangsan not go school
 ‘When Lisi was having meal, Zhangsan did not go to school.’

However, the inside sentence adjunct may also be located above the negative word (Pan & Paul 2018). In this case, sentence (19) also cannot transmit the information that ‘*Zhangsan* went to the school at another moment’. However, this is only due to the question of scope range, because the negation word *meiyou* is located below the inner sentence adjunct. But this does not indicate that the adverbial adjunct is not integrated.

- (19) Zhangsan zai Lisi chifan de shihou meiyou qu xuexiao.
 Zhangsan prep. Lisi eat DE time not go school
 ‘Zhangsan did not go to school when Lisi was having meal.’

4.2 Re-analyzing Chinese temporal adverbial clauses

Now I begin the analysis of the parsing possibilities of temporal adverbial clauses. As discussed

above, the temporal adverbial clauses in Chinese can be generated with and without a preposition, which may also influence their parsing. For ease of exposition, I will only analyze sentences headed by *de shihou*, which have a relative clause nature. I will also compare the structures with and without preposition. The structures that I analyze display subordinate-main order, but in some cases the inside sentence adjunct clause and main-subordinate order will also be considered.

I will first analyze sentences initiated by a full DP, which implies a forward anaphora interpretation, with the following structures to be analyzed:

- (20) Structure A: DP (preposition) Ø..., [main clause DP/pron]
 Structure B: DP (preposition) Ø..., [main clause Ø]
 Structure C: DP preposition DP/pron..., [main clause Ø]
 Structure D: DP preposition DP/pron..., [main clause DP/pron....]

Structure A can be exemplified as follows:

- (21) DP prep [-]..., [main clause pron ...]
 Zhangsan, zai [-] chifan de shihou, ta chang le yishou ge.
 Zhangsan prep. eat DE time he sing Pst one.CL song
- a. [Matrix TopP DP₁ (,)[[Adv.cl. prep. *pro*₁ +vP] [Main cl.TP pron. +vP]]]
 ‘Zhangsan, when (he) was having meal, he sang a song.’

First of all, it is crucial to decide the nature of the sentence initial DP. As has been argued by Pan & Paul 2018, the sentence initial DP can serve as matrix topic, matrix subject, subordinate topic and subordinate subject. When analyzing this structure and all of the remainder structures, I will start by deciding the nature of the sentence initial constituent (when it is possible).

In (21), since the main clause already contains a subject, the sentence initial DP *Zhangsan* cannot be the matrix subject. So, it may serve as the matrix topic. It should be noticed that *Zhangsan* in this structure cannot be the subordinate subject or topic, due to the presence of the preposition *zai*⁴⁹. As discussed by Pan & Paul 2018, the whole relative clause (with an NP/DP

⁴⁹ It is worth noting that *Zhangsan* in (21) can indeed be analyzed as the subordinate subject; this is because the preposition *zai* in Chinese can also function as a verb, indicating progressive tense. So, the subordinate clause can be interpreted as ‘when Zhangsan was having a meal’. However, I will not consider this parsing possibility and will only treat *zai* as a preposition.

nature) serves as the complement of the preposition *zai*. Since *Zhangsan* is to the left of the preposition, it cannot be considered as the subject or topic of the temporal relative clause. As a result, *Zhangsan* can only be parsed as the matrix topic, while the adverbial clause is believed to have a null subject. Since (21) involves a complex NP island, *Zhangsan* should be analyzed as a base-generated topic which controls the subordinate subject via the GCR. It is also necessary to have a pause between *Zhangsan* and *zai*, underlining the topic nature of the former. Structure B can be exemplified as follows:

- (22) DP prep. [-]..., [main clause [-] ...]
 Zhangsan zai [-] chifan de shihou [-] chang le yishou ge.
 Zhangsan prep. eat DE time sing Pst one.CL song
- a. [TP DP₁ [vP [Adv.cl. prep. *pro*₁ + vP] vP]]
 ‘(i) Zhangsan sang a song when (he) was having meal.’
- b. [Matrix TopP DP₁ (,)[[Adv.cl. prep. *pro*₁ + vP] [Main cl.TP *pro*₁ +vP]]]
- c. [Matrix TopP DP₁ [Adv.cl. prep. *pro*₁ + vP], [Main cl. DP₁ + vP]]
 ‘(ii) Zhangsan, when (he) was having meal, (he) sang a song.’

In this structure, since there is no matrix subject, the sentence initial DP *Zhangsan* can serve as the matrix subject, indicating that the adverbial clause is in fact an inside sentence adjunct (cf. 22a). It is also possible to consider *Zhangsan* as the matrix topic, either base-generated or derived by A'-movement from the matrix subject position (cf. 22b-c). However, *Zhangsan* cannot be considered as the subordinate subject or topic, since it is to the left of the preposition, which is the same case as in (21). The subordinate subject is also null and controlled by the matrix topic or subject *Zhangsan*.

Now we consider the structures without preposition. Structure A without preposition can be exemplified as in (23).

- (23) DP [-] ..., [main clause pron ...]
 Zhangsan [-] chifan de shihou, ta chang le yishou ge.
 Zhangsan eat DE time he sing Pst one.CL song
- a. [Matrix TopP DP₁ (,)[[Adv.cl. *pro*₁ +vP] [Main cl.TP pron. +vP]]]
 ‘(i) Zhangsan, when (he) was having meal, he sang a song.’

- b. [Adv.cl. DP₁ + vP]], [Main cl. pron. + vP]
 c. [Adv.cl. [TopP DP₁ [Adv.cl. *pro*₁ + vP]], [Main cl. pron. + vP]
 ‘(ii) When Zhangsan was having meal, he sang a song.’

In this case, the parsing possibilities can be more complicated than the structures with preposition. In the presence of the matrix subject, *Zhangsan* cannot serve as the matrix subject, but can still be considered as a base generated matrix topic (cf. 23a), as in (21) and (22). In (23), *Zhangsan* can also be considered as the subordinate subject or topic⁵⁰, since there is no PP projection here (cf. 23b-c).

In the situation shown in (24), i.e., Structure B without preposition, the sentence initial DP *Zhangsan* can have all the four parsing possibilities, namely (i) matrix topic (cf. 24a-b), (ii) matrix subject (cf. 24c), (iii) subordinate topic (cf. 24d) and (iv) subordinate subject (cf. 24e):

- (24) DP ([-]) ..., [main clause [-] ...]
 Zhangsan [-] chifan de shihou, [-] chang le yishou ge.
 Zhangsan eat DE time sing Pst one.CL song
- a. [Matrix TopP DP₁ [Adv.cl. *pro*₁ + vP], [Main cl. *pro*₁ + vP]]
 b. [Matrix TopP DP₁ [Adv.cl. *pro*₁ + vP], [Main cl. DP₁ + vP]]
 ‘(i) Zhangsan, when (he) was having meal, (he) sang a song.’
- c. [TP DP₁ [vP [Adv.cl. *pro*₁ + vP] vP]]
 ‘(ii) Zhangsan sang a song when (he) was having meal.’
- d. [Adv.cl. [TopP DP₁ [Adv.cl. *pro*₁ + vP]], [Main cl. *pro*₁ + vP]
 e. [Adv.cl. DP₁ + vP]], [Main cl. *pro*₁ + vP]
 ‘When Zhangsan was having meal, (he) sang a song.’

When it is matrix topic, it can also be base-generated (cf. 24a) or generated by movement from the matrix subject position (cf. 24b).

Now we see Structure D, which can be exemplified as follows:

- (25) DP prep. pron ..., [main clause pron ...]
 Zhangsan, zai ta chifan de shihou, ta chang le yishou ge.
 Zhangsan prep. he eat DE time he sing Pst one.CL song

⁵⁰ Unlike Pan & Paul 2018, I will not consider that the subordinate topic is generated by movement, which is proved to be available out of an island when there is no episodic eventuality, since the GCR analysis accommodates all of the cases, while the former is only available in some circumstances.

- a. [Matrix TopP DP₁ (,)[[Adv.cl. prep. pron. +vP] [Main cl.TP pron. +vP]]]
 ‘Zhangsan, when he was having meal, he sang a song.’

In this case, *Zhangsan* cannot serve as the matrix or subordinate subject, due to the presence of the two pronouns in the corresponding positions. With the presence of the preposition *zai*, *Zhangsan* should not serve as the subordinate topic, so its only possibility is to serve as the matrix topic, which is supposed to share an index with at least one of the two pronouns.

Finally, for structure C, the example is presented below:

- (26) DP prep. pron ..., [main clause [-] ...]
 Zhangsan zai ta chifan de shihou [-] chang le yishou ge.
 Zhangsan prep. he eat DE time sing Pst one.CL song

- a. [TP DP₁ [vP [Adv.cl. prep. pron. + vP] vP]]
 ‘(i) Zhangsan sang a song when he was having meal.’
- b. [Matrix TopP DP₁ (,)[[Adv.cl. prep. pron. +vP] [Main cl.TP *pro*₁ +vP]]]
 c. [Matrix TopP DP₁ (,)[Adv.cl. prep. pron. + vP], [Main cl. ~~DP~~₁ + vP]]
 ‘(ii) Zhangsan, when he was having meal, (he) sang a song.’

In (26), since there is no matrix subject, *Zhangsan* can serve as the matrix subject (cf. 26a), though it is also possible for it to serve as the matrix topic, either base-generated or generated by A'-movement from the matrix subject position (cf. 26b-c). The presence of *zai* and *ta* ‘he’ in the subordinate clause eliminates the possibility for *Zhangsan* to become the subordinate topic or subject, respectively. The subordinate subject *ta* ‘he’ may or may not share the same index with the topic *Zhangsan*.

Here I will not discuss the cases without preposition when the subordinate subject is overt, which are considered to be odd by Pan & Paul 2018. So, I will pass to the cases where there is no sentence initial DP, which yields a backward anaphora situation. The following structures will be considered:

- (27) Structure A [-] (prep.) [-] ..., [main clause pron/DP ...]
 Structure B [-] (prep) [-] ..., [main clause [-] ...]
 Structure C [-] prep. DP/pron ..., [main clause [-] ...]
 Structure D [-] prep. DP/pron ..., [main clause DP/pron ...]

Structure A can be exemplified as follows:

- (28) [-] (prep.) [-] ..., [main clause pron/DP ...]
 (Zai) Chifan de shihou, ta/Lisi chang le yishou ge.
 prep. eat DE time he/Lisi sing Pst one.CL song
- a. [Adv.cl. (prep.) *pro*₁ + vP]], [Main cl. pron/DP₁ + vP]
 ‘When (he) was having meal, he/Lisi sang a song.’

In this condition, since there is no major difference between structures with and without preposition, I will discuss them all together. It is clear that the subordinate subject position is occupied by a null element, which is preferentially co-indexed with the matrix subject (pronoun or DP).

Here is an example of structure B:

- (29) [-] (prep.) [-] ..., [main clause [-] ...]
 (Zai) Chifan de shihou, [-] chang le yishou ge.
 prep. eat DE time sing Pst one.CL song
- a. [Adv.cl. (prep.) *pro*₁ + vP]], [Main cl. *pro*₁ + vP]
 ‘When (he) was having meal, (he) sang a song.’

In this condition, both the subordinate and the matrix subject are null, and they are preferentially co-indexed.

Now let us consider structure C:

- (30) [-] prep. DP/pron ..., [main clause [-] ...]
 Zai Zhangsan/ta chifan de shiou, [-] chang le yishou ge.
 prep. Zhangsan/he eat DE time, sing Pst one.CL song
- a. [Adv.cl. prep. DP/pron.₁ + vP]], [Main cl. *pro*₁ + vP]
 b. [Adv.cl. prep. [TopP DP/pron.₁ [Adv.cl. *pro*₁ + vP]], [Main cl. *pro*₁ + vP]
 ‘When Zhangsan/he was having meal, (he) sang a song.’

In this condition, there is no matrix topic, and the subordinate subject is a DP or a pronoun (cf. 30a). Since the DP or pronoun is inside the PP, it is possible to consider that they are in the topic position of the subordinate clause (cf. 30b), though hard to parse. The matrix subject is null and may be co-indexed with the subordinate subject.

Finally, structure D can be presented in the following way:

(31) [-] prep. DP/pron ..., [main clause DP/pron ...]

Zhai Zhangsan/ta chifan de shiyou, Lisi/ta chang le yishou ge.
 prep. Zhangsan/he eat DE time, Lisi/he sing Pst one.CL song

a. [Adv.cl. prep. DP/pron.₁ + vP]], [Main cl. DP₂/pron._{1,2} + vP]

b. [Adv.cl. prep. [TopP DP/pron.₁ [Adv.cl. *pro*₁ + vP]], [Main cl. DP₂/pron._{1,2} + vP]
 ‘When Zhangsan/he was having meal, Lisi/he sang a song.’

In this situation, there is still no matrix topic, while a DP or pronoun occupies the subordinate subject position (cf. 31a), or subordinate topic position (cf. 31b, hard to interpret). Concerning the correlation between the subordinate and the matrix overt subjects’ interpretations, the results are quite interesting. If both subjects are pronouns, their coreference is acceptable and even preferential. If the subordinate subject is DP while the matrix subject is a pronoun, both the coreferential and the disjoint readings are available. It is worth noting that the previous two conditions are in a forward anaphora situation. If the subordinate subject is a pronoun and the matrix subject is a DP, which yields a backward anaphora situation, the disjoint reading will be the preferential one. These properties are one of the focus of this thesis. The reasons why there are such interpretative preferences in Chinese will be discussed in section 4.6.

Here, I will not analyze examples without preposition, like the following ones (cf. 32 and 33), because in the absence of a preposition, these cases can be confused with those with a DP or a pronoun in sentence initial position, which are discussed in the conditions with DP (which have been already analyzed in the previous part) or pronoun at the sentence initial position (which will be analyzed in the next part).

(32) [-] DP/pron ..., [main clause DP/pron ...]

Zhangsan/ta chifan de shiyou, Lisi/ta chang le yishou ge.
 Zhangsan/he eat DE time Lisi/he sing Pst one.CL song
 ‘When Zhangsan/he was having meal, Lisi/he sang a song.’

(33) [-] DP/pron ..., [main clause [-] ...]

Zhangsan/ta chifan de shiyou, [-] chang le yishou ge.
 Zhangsan/he eat DE time sing Pst one.CL song
 ‘When Zhangsan/he was having meal, (he) sang a song.’

Finally, we turn to the structures with a pronoun at the sentence initial position. The following

structures will be discussed (34). Like the cases in which the sentence is initiated by a DP, the preposition is obligatory in structures C and D.

- (34) Structure A: pron (prep.) [-] ..., [main clause [-] ...]
 Structure B: pron, (prep.) [-] ..., [main clause pron/DP ...]
 Structure C: pron prep. pron/DP ..., [main clause [-] ...]
 Structure D: pron prep. pron/DP ..., [main clause pron/DP ...]

Structure A with preposition can be exemplified as follows:

- (35) pron prep. [-] ..., [main clause [-] ...]
 Ta zai [-] chifan de shihou [-] chang le yishou ge.
 he prep. eat DE time sing Pst one.CL song
- a. [Matrix TopP pron.₁ (,)[[Adv.cl. prep. *pro*₁ + vP] [Main cl.TP *pro*₁ + vP]]]
 b. [Matrix TopP pron.₁ (,)[Adv.cl. prep. *pro*₁ + vP], [Main cl. ~~pron~~₁ + vP]]
 ‘(i) He, when (he) was having meal, (he) sang a song.’
- c. [TP pron.₁ [vP [Adv.cl. prep. *pro*₁ + vP] vP]]
 ‘(ii) He sang a song when (he) was having meal.’

This condition is equivalent to example (22), while the pronoun *ta* ‘he’ can be analyzed as the matrix topic (base-generated or generated by movement from matrix subject position, cf. 35a-b) or the matrix subject (cf. 35c), but cannot be analyzed as the subordinate topic or subject, because it is to the left of the preposition. The subordinate null subject may share the same index with the pronoun. Now let us consider the case without preposition.

- (36) pron [-] ..., [main clause [-] ...]
 Ta [-] chifan de shihou [-] chang le yishou ge.
 he eat DE time sing Pst one.CL song
- a. [Matrix TopP pron.₁ [Adv.cl. *pro*₁ + vP], [Main cl. *pro*₁ + vP]]
 b. [Matrix TopP pron.₁ [Adv.cl. *pro*₁ + vP], [Main cl. ~~pron~~₁ + vP]]
 ‘(i) He, when (he) was having meal, (he) sang a song.’
- c. [TP pron.₁ [vP [Adv.cl. *pro*₁ + vP] vP]]
 ‘(ii) He sang a song when (he) was having meal.’
- d. [Adv.cl. [TopP pron.₁ [Adv.cl. *pro*₁ + vP]]], [Main cl. *pro*₁ + vP]
 e. [Adv.cl. pron.₁ + vP]], [Main cl. *pro*₁ + vP]

‘(iii) When he was having meal, (he) sang a song.’

Here, as in example (24), the pronoun *ta* ‘he’ can have all the four parsing possibilities. When it is the matrix topic, it can either be base-generated or generated by movement (cf. 36a-b).

The following example corresponds to Structure B with preposition:

(37) pron, prep. [-] ..., [main clause pron/DP ...]

Ta, zai chifan de shihou, ta/Zhangsan chang le yishou ge.
he prep. eat DE time he/Zhangsan sing Pst one.CL song

a. [Matrix TopP pron.₁ (,)[[Adv.cl. prep. *pro*₁ +vP] [Main cl.TP pron._{1,2}/DP₂ +vP]]]
‘He, when (he) was having meal, he/Zhangsan sang a song.’

Like in example (21), the sentence initial pronoun *ta* ‘he’ can be analyzed as a base-generated matrix topic, but not a matrix subject, or a subordinate subject or topic. Thus, a pause is needed between the first *ta* and *zai* to underline the topic nature of the sentence initial pronoun. The subordinate null subject can be controlled by *ta* via the GCR. For the matrix subject, if it is also a pronoun, coreference between the two pronouns is possible; if it is a DP, coreference is not possible, due to *Principle C* of the Binding theory.

Structure B without preposition can be exemplified as in (38):

(38) pron [-] ..., [main clause pron/DP ...]

Ta chifan de shihou Zhangsan/ta chang le yishou ge.
he eat DE time Zhangsan/he sing Pst one.CL song

a. [Matrix TopP pron.₁ (,)[[Adv.cl. *pro*₁ +vP] [Main cl.TP DP₂/pron.₁ +vP]]]
‘(i) He, when (he) was having meal, Zhangsan/he sang a song.’

b. [Adv.cl. pron.₁ + vP]], [Main cl. DP₂/pron.₁ + vP]

c. [Adv.cl. [TopP pron.₁ [Adv.cl. *pro*₁ + vP]]], [Main cl. DP₂/pron.₁ + vP]
‘(ii) When he was having meal, Zhangsan/he sang a song.’

Like in (23), the situation in (38) is quite complex. The sentence initial pronoun can be analyzed as a base-generated matrix topic (cf. 38a), or as a subordinate subject or topic (cf. 38b-c). However, it cannot be the matrix subject, since this position is already occupied. Concerning the referential possibilities, if the matrix sentence has a subject pronoun, coreference between the two pronouns is possible. If the matrix subject is a DP, coreference between the DP and the

pronoun is less favored, a phenomenon which is one of the focuses of this thesis.

Now let us look at an example of structure C:

(39) pron prep. pron/DP ..., [main clause [-] ...]

Ta zai Zhangsan/ta chifan de shihou chang le yishou ge.
he prep. Zhangsan/he eat DE time sing Pst one.CL song

a. [TP pron.₁ [vP [Adv.cl. prep. DP₂/pron.₁ + vP] vP]]

‘(i) He sang a song when Zhangsan/he was having a meal.’

b. [Matrix TopP pron.₁ (,)[[Adv.cl. prep. DP₂/pron.₁ +vP] [Main cl.TP *pro*₁ +vP]]]

c. [Matrix TopP pron.₁ (,)[Adv.cl. prep. DP₂/pron.₁ + vP], [Main cl. ~~pron~~₁ + vP]]]

‘(ii) He, when Zhangsan/he was having meal, (he) sang a song.’

As is (26), the sentence initial pronoun can serve as the matrix subject (cf. 39a) or the matrix topic, either base-generated or generated by movement (cf. 39b-c). It cannot be considered as the subordinate subject or topic. Concerning the referential possibilities, like in (38), if the subordinate clause has a subject pronoun, coreference between the two pronouns is possible. If the subordinate subject is a DP, the coreferential reading is not possible, due to *Principle C* of the Binding Theory.

Finally, structure D can be exemplified as follows:

(40) pron prep. pron/DP ..., [main clause pron/DP ...]

Ta, zai ta/Zhangsan chifan de shihou, ta/Zhangsan chang le yishou ge.
he prep. he/Zhangsan eat DE time he/Zhangsan sing Pst one.CL song

a. [Matrix TopP pron.₁ (,)[[Adv.cl. prep. DP₂/pron.₁ +vP] [Main cl.TP DP₂/pron.₁ +vP]]]

‘He, when he/Zhangsan was having meal, he/Zhangsan sang a song.’

Considering (40), which is similar to (25), the sentence initial pronoun cannot be considered as a matrix or subordinate subject, since these positions are already occupied. It also cannot be analyzed as a subordinate topic, due to the presence of *zai*. So, the first *ta* can only be considered as a matrix topic, and a pause is needed between *ta* and *zai*. Considering the referential possibilities, like in the previous conditions, if one of the overt subjects (subordinate or matrix) is a DP, it cannot be bound by *ta* ‘he’, due to *Principle C* of the Binding Theory. However, if the subordinate or matrix subject is a pronoun, the coreference reading is still possible. Some

further discussion will be made in section 4.6.

Now we recheck examples (31) and (40). It is interesting to observe that for Chinese temporal adverbial adjuncts with the subordinate-main order, a DP cannot be co-indexed (or at least this is not the preferred interpretation) with an overt pronoun on its left-hand side in the linear order. This observation is also true for the structures with main-subordinate order. Recall that, in section 2.5, we have already shown that this structure in Chinese is also non-integrated, taking into account the scope of matrix clause negation. Pan & Paul 2018 also claim that the Chinese temporal adverbial adjunct with main-subordinate order indicates that the information from the matrix sentence is more prominent, and the preposition *zai* is obligatory in this structure.

The parsing possibility for this order is relatively simpler than the reverse order, as the sentence initial DP should always be interpreted as the matrix subject, or as an unmarked matrix topic. Concerning the referential interpretative properties, a null or overt pronoun may take the matrix subject as its antecedent⁵¹, see (41)

- (41) a. DP ... [subordinate clause [-] ...]
 Zhangsan chang le yishou ge, zai [-] chifan de shihou.
 Zhangsan sing Pst one.CL song prep. eat DE time
 ‘Zhangsan sang a song, when (he) was having meal.’
 b. DP ... [subordinate clause pron ...]
 Zhangsan chang le yishou ge, zai ta chifan de shihou.
 Zhangsan sing Pst one.CL song prep. he eat DE time
 ‘Zhangsan sang a song, when he was having meal.’

When both the matrix and the embedded subjects are pronouns, the coreferential reading is also acceptable.

- (42) Pron ... [subordinate clause pron ...]
 Ta chang le yishou ge, zai ta chifan de shihou.
 he sing Pst one.CL song prep. he eat DE time
 ‘He sang a song, when he was having meal.’

However, if the subordinate subject is a DP while the matrix subject is a pronoun, their coreference is not possible, still due to *Principle C* of the Binding Theory.

⁵¹ However, the overt embedded pronoun, but not the null one, can also have an antecedent other than the matrix subject.

(43) Pron ... [subordinate clause DP ...]

Ta chang le yishou ge, zai Zhangsan chifan de shihou.
 he sing Pst one.CL song prep. Zhangsan eat DE time
 ‘He sang a song, when Zhangsan was having a meal.’

Differently from what happens with the subordinate-main order, the matrix null subject cannot take the subordinate subject DP as its antecedent, which is also due to *Principle C* of the Binding Theory.

(44) [-] ... [subordinate clause DP ...]

[-] Chang le yishou ge, zai Zhangsan chifan de shihou.
 sing Pst one.CL song prep. Zhangsan eat DE time
 ‘(He) sang a song, when Zhangsan was having meal.’

Now let us consider the last situation. If there is a marked topic in the matrix clause, it cannot refer to the subordinate null subject, which indicates that there is an island effect (45a). However, the left dislocated topic can refer to an overt subordinate subject pronoun, indicating the absence of the island condition in this situation (45b). This property is also presented in Pan 2016, 2017 and will be discussed in the following section.

(45) Topic DP ... [subordinate clause [-]/pron ...]

- a. *Lisi₁ (a), Zhangsan₂ chang le yishou ge, zai [-]₁ chifan de shihou.
 Lisi Sfa. Zhangsan sing Pst. one.CL song prep. eat DE time
 ‘As for Lisi, Zhangsan sang a song when (he) was having meal.’
- b. Lisi₁ (a), Zhangsan₂ chang le yishou ge, zai ta₁ chifan de shihou.
 Lisi Sfa. Zhangsan sing Pst one.CL song prep. he eat DE time
 ‘As for Lisi, Zhangsan sang a song when he was having meal.’

Now we have seen the parsing possibilities of the structures with subordinate-main and main-subordinate orders. It should be noticed that those parsing possibilities are quite complex, and some structures can also be confused with others. This fact may also lead us to revise some previous studies which involve such structures.

Zhao 2014 claims that, in the backward anaphora structures of Chinese, a subordinate pronoun cannot share the same index with the matrix subject DP, due to the cyclic c-command of Huang 1982. However, if we check the sentences tested in Zhao 2014, it seems that the interpretative judgement found by Zhao 2014 may result from an ambiguity of parsing, since all the

subordinate-main order sentences tested in Zhao 2014 did not have a preposition in its subordinate clause and were basically similar to (38), here repeated as (46):

- (46) Ta chifan de shihou, Zhangsan chang le yishou ge.
 he eat DE time Zhangsan sing Pst one.CL song
 ‘When he was having meal, Zhangsan sang a song.’

As I have already shown before, this structure may have at least two different parsing possibilities. One is to consider the sentence initial *ta* ‘he’ as the subordinate subject or topic, and the other is to consider *ta* as the matrix topic. In Zhao 2014’s analysis, she clearly considered *ta* as the subordinate subject, and claimed that the less favored (or impossible) coreferential interpretation between *ta* and *Zhangsan* is caused by the cyclic c-command relation between them. However, if we use the second parsing possibility, considering *ta* as the matrix topic, then it should be found that the matrix subject *Zhangsan* is c-commanded (but not necessarily cyclic c-commanded) by the matrix topic *ta*. Thus, it is clear that their coreference is not possible.

As a result, in structures like (46), what the cyclic c-command apparently predicts may be caused only by the ambiguity of parsing. That is to say, since (46) has two parsing possibilities, native speakers may choose the second parsing possibility and consider *ta* as the matrix topic, but not as the subordinate subject. Since sentence interpretation is always from left to right, the speakers will only recognize that the first clause in (46) can be a subordinate clause when they read the final part of the clause *de shihou*. This fact may also increase the possibility that the speakers choose to analyze *ta* as a matrix topic. Thus, it is obvious that they will not accept the coreference interpretation. In this way, it seems that there is no need to analyze the interpretation in (46) applying the cyclic c-command proposal.

So, it seems clear that the Chinese temporal adverbial adjunct without preposition may cause some parsing ambiguities, which no longer exist in the structures which have a preposition. For example, in (47), since there is a preposition *zai* at the start of the sentence, the first clause must be a subordinate clause, and the pronoun that follows *zai* can only be analyzed as the subordinate subject, but not as a matrix topic or subject. The pilot Chinese test of the current thesis also suggests that, when there is a preposition, speakers are more likely to accept the

coreferential reading between the pronoun and the matrix subject, though still at a lower acceptance level.

- (47) Zai ta chifan de shihou, Zhangsan chang le yishou ge.
 prep. he eat DE time Zhangsan sing Pst one.CL song
 ‘When he was having meal, Zhangsan sang a song.’

Considering these facts, I will only analyze temporal adverbial adjuncts with preposition in the remainder of the thesis. Furthermore, in the chapter 6, in which I present the experimental study of the current thesis, I will only use sentences with preposition in the Chinese test, thus eliminating the possible parsing ambiguities. Concerning the subordinate-main order sentences, I will only use sentences starting with a preposition and I do not include those with a DP preceding the preposition. So, the sentences with overt embedded subject will be like (47), with the subordinate subject inside the PP. In this way, the participants can recognize from the start that the first clause of the sentence is a subordinate clause, which may reduce or eliminate the influence from the parsing ambiguities discussed in this section.

For sentences with main-subordinate order, it seems that Chinese sentences like (48) are not equivalent to their Portuguese counterparts (49), since considering the negation scope test, the Chinese sentence (48) should be considered as a non-integrated structure, while the EP sentence (49) has an integrated structure. At this point, it seems that the inside sentence temporal clauses in Chinese are closer to the EP sentence in (49), since they share the same result in the negation scope test.

- (48) Zhangsan chang le yishou ge, zai ta chifan de shihou.
 Zhangsan sing Pst one.CL song prep. he eat DE time
 ‘Zhangsan sang a song, when he was having meal.’
 (49) O João cantou uma canção quando almoçava.
 the John sang one song when had lunch
 ‘John sang a song when he was having lunch.’

However, considering the similarity of linear order between (48) and (49), I will still test sentences like (48) in the experimental test of Chinese. It should be noticed that in Chinese the right dislocated temporal adverbial adjunct clause shares the same referential interpretative preferences with the inside sentence adjunct as far as the subjects are concerned. This fact also

guarantees the feasibility of testing sentences like (48) in my experimental test.

There is another point relating to the analysis of the parsing possibilities of the temporal adverbial clauses to which attention should be drawn, namely the nature of the empty categories involved. In Pan & Paul's 2018 analysis of conditionals, the subordinate null subjects, especially those located inside an island, are easily considered as a *pro* which can be controlled and identified by the closest nominal element via GCR. In my analysis of temporal adverbial adjuncts (22) to (41), the same holds for the subordinate null subjects. However, my analysis also involves matrix null subjects such as those in (30), here repeated as (50):

(50) Zai Zhangsan/ta chifan de shiou, [-] chang le yishou ge.
 prep. Zhangsan/he eat DE time sing Pst one.CL song
 'When Zhangsan/he was having meal, (he) sang a song.'

Then the question will be the following: should the matrix null subject also be analyzed as a *pro* controlled by some nominal element, or it is generated by A'-movement? In the next section, I will make a summary of the studies concerning Chinese null subjects and propose an analysis to explain several phenomena related to these empty categories.

4.3 Re-analyzing Chinese null subjects

The studies about Chinese null subjects (and objects) in the GB framework can be dated back to Huang 1982, 1984, who argued that the null subject in this language can be a *pro* controlled by the closest nominal element via GCR. Since GCR has been discussed several times in the current thesis (see Chapter 1 and section 2.2 of Chapter 2), I will not repeat it here. However, GCR was only set to explain subordinate null subjects in Chinese. As for the null subject in matrix sentences, Huang 1984 argued that it can be analyzed as a variable generated by A'-movement, and thus it should be bound by a topic element, which can be overt or null.

For example, in (51), it is assumed that the speaker was waiting for someone, and suddenly he saw that person and uttered the sentence in (51). In Huang's 1984 theory, it is believed that a null topic is located in the matrix topic position, which can be identified as the person that the speaker was waiting for via the context. The null topic is also believed to be moved from the matrix subject position via A'-movement. Huang 1984 believed that all of the matrix null

explain Chinese null subjects in matrix clauses, while Li's 2007 theory only concerns the identification of subordinate *pro*.⁵²

If we adopt the proposal of Liu 2014, then it is possible to analyze any matrix null subject in Chinese as a *pro* which is in an Agree relation with a previous aboutness-shift topic. However, such a generalization may be too strong, since there is another possibility to generate the matrix null subject, namely the ATB movement of the subject in sentence coordinate structures, which has been tested in various languages (see examples for EP, Costa & Matos 2012: 213).

In fact, Liu 2014 himself also considered this kind of analysis for Chinese null subjects, which he abandoned at the end. Liu 2014 observed that an analysis of ATB movement cannot explain the null subject inside subordinate clauses, especially inside islands. For example, (54) clearly allows the embedded null subject to refer to *John*. Since this null subject is inside an island, it is absolutely impossible to consider that it is generated by ATB movement. Because of this, Liu 2014 rejected the proposal of ATB movement and proposed the agreement analysis.

- (54) Yuehan_i zuotian xiawu hen e, suoyi [yi e_i hui-dao jia
 John yesterday afternoon very hungry so once return-arrive home
 hou], e_i jiu xian chi-le yi-ge sanmingzhi.
 after then first eat-ASP one-CL sandwich
 'John was very hungry in the afternoon yesterday, so once he got home, he immediately ate a sandwich.'
 (Liu 2014: 174)

However, in Chinese, there are some cases in which several clauses have a null subject sharing the reference of the overt subject of the first clause, and it is not possible to replace all of the null subjects with an overt pronoun or DP. In these cases, an ATB movement analysis should still be favored.

As a result, I propose that it should be plausible to maintain the ATB movement analysis for some matrix null subjects in Chinese, in the cases where the clauses can be analyzed as coordinate structures. Since coordinate structures are not the focus of this thesis, I will not make a detailed analysis of these cases. Of course, it should not be possible to analyze all kinds of clauses as coordinate clauses. So, in these cases, the matrix null subject should still be analyzed

⁵² Li's 2007 theory is based on the GCR of Huang 1984, who believes that the null subject in Chinese matrix clauses should be analyzed as a variable left by A'-movement. In this sense, Li's 2007 uD identification may only apply to subordinate null subjects.

as *pro*⁵³ or gap generated by A'-movement.

There is also another alternative proposal concerning the legitimation and identification of *pro* in Chinese, using the theory of Agree. Zhao 2012 proposes that the phenomenon described by GCR can be analyzed as a null self-reflexive (null *ziji*) agreeing with its antecedent. For example, in (55), the embedded null subject can refer to the matrix subject *Zhangsan*.

- (55) *Zhangsan shuo e renshi Lisi.*
Zhangsan say know Lisi
 'Zhangsan says that he knows Lisi.' (Zhao 2012: 172)

Zhao 2012 proposes that the embedded null subject is in fact a null reflexive *ziji*, which is in an Agree relation with the matrix subject *Zhangsan*, thus, the null *ziji* receives the same referential value as *Zhangsan*. This claim can be further supported by the fact that an overt reflexive *ziji* may also be placed in the embedded subject position, conveying the same meaning.

- (56) *Zhangsan shuo ziji renshi Lisi.*
Zhangsan say self know Lisi
 'Zhangsan says that he himself knows Lisi.'

However, there are some problems if all of the cases of *pro* in Chinese are analyzed as a null reflexive. For example, in (57), the reflexive *ziji* in the deepest embedded clause may still refer to the matrix subject *Zhangsan*, since Chinese allows long-distance binding of reflexives. (Cole & Sung 1994).

- (57) *Zhangsan₁ yiwei Lisi₂ zhidao ziji_{1,2} yao qu xiu che.*
Zhangsan think Lisi know self will go repair car
 'Zhangsan thinks that Lisi knows that he will go to repair the car.'

Of course, the reflexive *ziji* in (57) can also refer to the embedded subject *Lisi*, making the sentence ambiguous. However, if *ziji* in (57) is replaced by a null subject, then the preferred interpretation is to consider the null subject to refer to the embedded subject *Lisi*, and not to the matrix subject *Zhangsan*.

⁵³ This is also true for EP, see Costa & Matos 2012.

- (58) Zhangsan₁ yiwei Lisi₂ zhidao [-]??_{1/2} yao qu xiu che.
 Zhangsan think Lisi know will go repair car
 ‘Zhangsan thinks that Lisi knows that he will go to repair the car.’

As a result, if the null subject in (58) is analyzed as a null *ziji*, then it is problematic that it cannot refer to the matrix subject, when the overt form permits this interpretation. Considering this fact, I will not adopt the idea of null *ziji* of Zhao 2012 in the current thesis.

In a recent study also involving agreement, Pan 2016 analyzed some Chinese null categories and pronouns, concerning their internal structures. Adopting proposals from Adger & Ramchand 2001, 2005, and Rouveret 2002, 2008, 2019, Pan 2016, 2017 proposes that the Chinese relative clauses and the topicalization structures with gap or with resumptive pronoun can be analyzed as involving an Agree or Match operation.

Pan 2016, 2017 first suggests that the classic movement analysis for these structures can be replaced by operations like Agree and Match in the MP, while Match may be separated from Agree and serve as an independent operation to legitimate some topicalization structures. In Pan’s 2016, 2017 analysis, Agree involves the valuation of an uninterpretable feature by a matched interpretable feature, while Match does not require feature valuation. As a result, two uninterpretable features can be matched, but do not undergo Agree, in Pan’s analysis. For relative clauses, there is an Agree operation between the NP head and the gap or the resumptive pronoun. Since there is an uninterpretable [variable] feature in the C of the relative clause, which makes it a Probe, the gap or the resumptive pronoun, with an interpretable [variable] feature, which allows it to serve as a Goal, is triggered to agree with the relative C. Pan 2016, 2017 considers that an alternative way to analyze Agree is to consider that the C-Rel bears an interpretable [λ] feature, while the resumptive pronoun bears an uninterpretable [λ] feature. (59) shows an example of a relative clause with a resumptive pronoun.

- (59) [NP NP₁ [CP [C° C-Rel] [TP RP₁]]]
 either u-[var] i-[var]
 or i- $[\lambda]$ u- $[\lambda]$
 $\sqrt{\text{Match}}$
 $\sqrt{\text{Agree}}$

(Pan 2016: 111)

For left dislocation structures, Pan 2016, 2017 proposes that the Agree operation applies

between the topicalized constituent and the gap, since the former bears an uninterpretable [variable] feature and the latter bears an interpretable [variable] feature. However, for left dislocation structures with a resumptive pronoun, there is no Agree operation, since both the topicalized structure and the resumptive pronoun bear an uninterpretable [variable] feature. Nevertheless, the Match operation is still possible in this case, since in the theory of Pan 2016, 2017 Match can be separated from Agree. In this case, since both the topicalized structure and the resumptive pronoun bear a [variable] feature, a Match operation can be accomplished, with no feature valuing taking place. (60) shows an example of left dislocation with a resumptive pronoun.

(60) [TopP NP₁ [Top° Top°], [TP RP₁]]
 either u-[var] u-[var]
 or u-[λ] u-[λ]
 √ Match
 *Agree

(Pan 2016: 111)

Pan 2016, 2017 proposes that only the Agree operation, but not the Match operation, is subject to the island condition. This explains why in Chinese relativization is not possible inside a strong island, no matter whether a resumptive pronoun is used (61a). Topicalization structures with a gap, where the Agree operation applies, are subject to the island condition (61b). However, in left dislocation structures with a resumptive pronoun, since only Match operates, not Agree, the island condition can be violated (61c).

(61) a. *Zhe shi [wo jiandao-guo [tanlun-guo ta₁ / [-]₁
 this be I meet-Exp talk-Exp he
 de] na-ge nütongxue de] zuojia₁.
 C that-Cl female.student C writer
 ‘This is the writer₁ [whom I met the student [who talked about him₁/[-]₁].’
 b. Na-wei Faguo yingxing₁, wo pengdao-le
 that-Cl French movie.star I meet-Perf
 [Xiaoqian renshi [yongbao-guo *[-]₁]]
 Xiaoqian know embrace-Exp
 de na-wei nüsheng.
 C that-Cl female.student
 ‘As for that French movie star₁, I met the female student [that Xiaoqian knows [who embraced [-]₁].’

- c. Na-wei Faguo yingxing₁, wo pengdao-le
 that-Cl French movie.star I meet-Perf
 [Xiaoqian renshi [yongbao-guo ta₁]]
 Xiaoqian know embrace-Exp he
 de na-wei nüsheng.
 C that-Cl female.student

‘As for that French movie star₁, I met the female student [that Xiaoqian knows [who embraced him₁]].’
 (Pan 2016: 115 and 119)

As a result, under the analysis of Agree and Match, Pan’s 2016, 2017 theory is able to explain some properties of A’-movement in relation to island conditions. He then analyzed the internal structure of the gap and the resumptive pronoun on the basis of some tests of reconstruction. He proposes that, in a relative clause, both the gap and the resumptive pronoun have an extended internal structure [[D- ϕ] NP], which contains an NP structure and gives rise to the reconstruction effect. The same is also true for the gap in left dislocation structures, which also presents the reconstruction effect. However, Pan 2016, 2017 proposes that the resumptive pronoun in the left dislocation structure only has a simple structure [D- ϕ], without the NP structure, which does not give rise to the reconstruction effect.

Pan 2017 also analyzed another case, namely the null object inside an island. Recall the example (50) from Chapter 2, here repeated as (62), where an A’-bound null object obviously violates the island condition.

- (62) (Naxie ren₂), wo yinwei [Lisi hai bu renshi e₂] hen danxin.
 those people I because Lisi still not know very concerned
 (Those people), I am very concerned because Lisi still does not know *e*.’

(Li 2007: 97)

Li 2007 considers the null object in (62) as a TEP, which is neither a *pro* nor a gap. However, Pan 2017 proposes that this null object can be indeed a *pro*. Zhang 2002 had already noticed that this kind of null object is only available when the verb does not convey an episodic eventuality. Pan 2017 then argues that, when the verb does not convey episodic eventuality, the null object can be analyzed as a *pro*, which serves as a resumptive pronoun and bears an uninterpretable [variable] feature. Since the topicalized structure or the C of the relative clause also bears an uninterpretable [variable] feature, a Match operation will be available. Since Match is not subject to the island condition, those object *pros* can be grammatical inside an

island, either in left dislocation structures or in relative clauses. Pan 2017 also proposes that the *pro* in object position has an extended structure $[[D-\phi] \text{ NP}]$, since it also gives rise to the reconstruction effect.

There are several comparisons that should be made between the proposal of Pan 2016, 2017 and the previous analyses. First, it should be noticed that the majority of examples used in Pan 2016, 2017 are from relative clauses and left dislocation structures, where there is an apparent and obligatory A'-dependence between the relative C or topicalized constituent and the null element or resumptive pronoun. However, the structures and literature discussed in this thesis are mainly about null elements or pronouns which have referential content and may occur in the absence of any antecedent. In addition, even when their reference is fixed by an antecedent, there is no obligatory A'-dependency between the antecedent and the pronoun (null or overt) except in the cases where a null subject is considered as generated by A'-movement, such as (52), here repeated as (63):

- (63) Lisi₂, Zhangsan₁ shuo [-]₂ neng lai.
 Lisi Zhangsan say can come
 'As for Lisi, Zhangsan said that (he) can come.'

In fact, (63) is exactly the structure that Pan 2016, 2017 considered as a left-dislocation structure with a gap. However, in sentences like (64), there is no such A'-dependency applied to the null or overt pronoun. For the null subject in (64a), although it is strongly favored to share the index with the matrix subject, there is no A'-dependency here, since the antecedent of the null subject is in an A-position. And, in (64b), the overt pronoun can refer to *Zhangsan* or to another entity, which means that it cannot be compared to the resumptive pronouns analyzed in Pan 2016, 2017.

- (64) a. Zhangsan₁ shuo [-]₁ bu xiang shangxue.
 Zhangsan say not want go to school
 'Zhangsan says that (he) does not want to go to school.'
 b. Zhangsan₁ shuo ta₁ bu xiang shangxue.
 Zhangsan say he not want go to school
 'Zhangsan says that he does not want to go to school.'

The object *pro* proposed by Pan 2017 should also be distinguished from the subject *pro*, which

is believed to be legitimated by the GCR. Since the GCR predicts that *pro* must be controlled by the closest nominal element, it cannot appear in the object position, as already discussed by Huang 1984 and Li 2007. Under Pan 2017's analysis, the object *pro* is considered as a resumptive pronoun which must be A'-bound. However, it is still possible to consider that the subject *pro* has the same internal structure as the object *pro*.

For instance, Li 2007 claims that the Chinese *pro* must bear a D feature, and, in Pan's 2017 theory, the internal structure of *pro* also contains a D structure, which corroborates Li's 2007 proposal. Concerning the possibility of having reconstruction effects, Pan 2016, 2017 made four tests, namely the quantifier scope reconstruction, the quantificational antecedent, the reconstruction of anaphoric binding and the condition C effect. However, due to the characteristics of these tests⁵⁴, only the quantificational antecedent test can be applied to the subject *pro*, as shown below:

- (65) a. Meige ren₁ dou xiwang [-]₁ neng xingfu.
 every person all hope can happy
 'Everyone hopes that (he) can be happy.' (Xu 1986: 87)
- b. Meige ren₁ zai [-]₁ kan shu de shiou dou yao renzhen zixi.
 every person prep. read book DE time all need serious careful
 'Everyone should be serious and careful when (he) reads the book.'

In (65a-b), the embedded null pronoun can take the quantificational antecedent *meige ren* 'everyone' in complement and adverbial structures, showing that the subject *pro* also has an extended form, which is consistent with the object *pro*. Next let us consider an example with an overt referential pronoun.

- (66) a. ??Meige ren₁ dou xiwang ta₁ neng xingfu.
 every person all hope he can happy
 'Everyone hopes that he can be happy.' (Xu 1986: 87)
- b. ??Meige ren₁ zai ta₁ kan shu de shiou dou yao renzhen zixi.
 every person prep. he read book DE time all need serious careful
 'Everyone should be serious and careful when he reads the book.'

Interestingly, the overt referential pronoun cannot have a quantificational antecedent here,

⁵⁴ The quantifier scope reconstruction, the reconstruction of anaphoric binding and the condition C effect tests require *pro* to be in an object position, which is not applicable to the subject *pro*.

which indicates that it only has a simple internal structure. This is consistent with the resumptive pronoun in the left dislocation structure.

In summary, though there are some differences between the subject and object *pro*, their internal structure may still be the same. The same is true for the resumptive pronoun in left dislocated structure and the referential pronoun. So, at this point, I will propose that Pan's analysis of the internal structure is also true for the subject *pro* and the referential overt pronoun.

Next, we consider the Agree operations proposed by Liu 2014 and Pan 2016, 2017. It is obvious that these two proposals are not the same, since in Pan's theory, Agree must be subject to the island condition, while, in Liu's theory, Agree can be used to explain *pro* inside islands. In Liu's theory, what triggers Agree is the phi-feature, while, in Pan's theory, what triggers Agree can be either the variable feature or a phi-feature. However, if we compare the characterization of Agree in Liu 2014 with the Match version of Pan 2016, 2017, we discover that these two operations share some similarities, since both of them are not subject to island conditions. So, it is reasonable to consider whether the relation between a subject *pro* and its antecedent can be analyzed under the Match theory of Pan 2016, 2017. Nevertheless, before making a connection between the theory of Liu 2014 and Pan 2016, 2017, there is another question about the GCR that should be clarified.

It should be noticed that the previous studies have used the GCR to explain two kinds of anaphoric relations between *pro* and its antecedent, namely *pro* with an A'-antecedent and *pro* with an A-antecedent. In (67), the antecedent of *pro* is in an A-position⁵⁵, while, in (68), the antecedent of *pro* is in an A'-position.

(67) Zhangsan₁ xiwang [-]₁ neng qu meigu.
 Zhangsan hope can go America
 'Zhangsan hopes that (he) can go to America.'

(68) Zhangsan₁, zai [-]₁ chifan de shihou, tingdao le nage xiaoxi.
 Zhangsan prep. eat DE time hear Pst that news
 'Zhangsan, when (he) was having meal, (he) heard that news.'

⁵⁵ One may think that the matrix subject *Zhangsan* can serve as an unmarked topic. However, in the following example, when the matrix topic position is filled with another constituent, the matrix subject *Zhangsan* may still control the embedded *pro*. So, this proves that, in structures like (i), the matrix subject is in an A-position.

(i) 'Lisi₂, Zhangsan₁ zai *pro*₁ chifan de shihou kandao le ta₂. 'As for Lisi₂, Zhangsan₁ saw him₂ when *pro*₁ was having meal'.

In this way, I propose that these two kinds of use of GCR should be distinguished. The case in (67) should be considered as similar to the finite control of BP and Finnish, proposed by Holmberg & Sheehan 2010, since the antecedent of *pro* is always in an A-position. For the case in (68), since the antecedent of *pro* is in an A'-position, it would be interesting to discover if the anaphoric relation in this sentence should also be analyzed under Pan's 2016, 2017 A'-dependency proposals⁵⁶. Since Chinese is a discourse-oriented language (Tsao 1977) where the topic chain plays an important role concerning anaphoric relations, the A'-dependency related to topic should also be analyzed under the topic chain approach. In the following section, I will present a recent study that analyzes the Chinese null and overt pronouns under the topic chain theory.

4.4 Chinese anaphoric resolution under topic chain analyses

Pu & Pu 2014 aim to analyze the occurrence and distribution of zero anaphora and overt pronouns in Chinese, taking topic chain as the focus of the investigation⁵⁷. Based on two important factors, namely memory constraints and discourse functions, the authors propose a general topic chain principle: *Topic chain encodes a referent that is cognitively most accessible at the moment of discourse production, and the referent is kept activated and stays in focus as enhanced by maximum discourse coherence. Specifically, it is used to code a topical referent that persists over a span of maximally coherent discourse. The topic chain is terminated if the maximum discourse continuity is disrupted* (Pu & Pu 2014:32).

Pu & Pu 2014 consider maximum discourse coherence as a two-fold parameter: *TOPICAL CONTINUITY* (c.f. Givón 1983) and *THEMATIC COHERENCE*. *The former specifies topical persistence of a referent across a span of discourse, and the latter describes thematic continuity of the discourse unit in which the referent occurs* (Pu & Pu 2014: 32).

In the theory of topic chain, it is assumed that the topical referent is the focused entity inside

⁵⁶ However, as will be discussed in the next section, such a connection between Liu's and Pan's theory may not be plausible concerning sentences with island structures.

⁵⁷ In this thesis, I consider the topic chain analysis as a complementary analysis of other previous analyses based on topic and A'-movement, especial that of Frascarelli 2007. As can be seen from the discussion in 4.5 and 4.6, even under the topic chain analysis, it is still essential to assume that there is an A'-dependency between the syntactic topic and the null or overt pronouns in lower level structures. In this sense, it is plausible to consider that the topic chain analysis in fact focuses on the identification of topics through discourse information.

one or several subsequent sentences. As long as the topic is maintained, it yields a topic continuity situation. Pu & Pu 2014 claim that in a null subject language like Chinese, when there is topic continuity, the head of a topic chain is usually realized by an overt form, either a full NP or a pronoun, while the subsequent occurrences of the topic are realized by zero anaphora, which are predominantly located in subject positions⁵⁸. For example, in (69), the whole sequences of sentences describe a certain character, namely *Duan Quchen*, which functions as the topic. As all of the sentences and clauses describe the same topic, it yields a topic continuity situation, where only the first occurrence of the topic is realized by a NP, while the remaining occurrences are realized by null subjects.

(69) (a) Duan Quchen wei-ren zhengpai, (b) [-] dui gongzuo yaoqiu yange.

Duan Quchen as-person honorable in work demand strict

(c) [-] dai-ren fangmian you hen suihe, (d) [-] tong shenmeren dou neng

treat-people aspect also very amiable with anyone all can

shuo-shang hua (e) [-] conglai bu bai jiazi.

talk a talk ever no put-on-air

‘(a) Duan Quchen was a good leader, (b) [-] strict when work was concerned. (c) [-] but congenial, when dealing with people, (d) [-] could strike up a conversation with anyone, (e) [-] never appeared arrogant or remote.’ (Pu & Pu 2014: 34)

This kind of topic continuity often occurs in two situations. The first one is when describing a sequence of close-knit actions performed or experienced by the same topical referent, and the second one is when describing a series of closely related past experiences of the same topical referent.

There are also several factors that may cause the termination of a topic chain, which yields Minor Thematic Discontinuity.

The first case is designated as Interruption in Close-Knit Action/Event Sequence and it is caused by time or location change in the action of event sequence. For example, in (70), the topical referent of the first sentence is *Zou Xiping*, which is realized by a null subject in the following sentence, as the sentences (a) and (b) (which describe a close-knit action) form a topic chain. However, in sentences (c) to (f), the location of the actions has been changed, as in sentences

⁵⁸ It should be noticed that this kind of referential dependency between topic and null subjects is not unique to Chinese, as other null subject languages may present the same property, as is the case of EP.

(a) and (b) they occurred outside the ‘tractor’, while in sentences (c) to (f) they took place inside the ‘tractor’. As a result, the change of location of action caused the termination of the topic chain headed by *Zou Xinping* in sentence (a) and (b). Hence, from sentence (c) onwards, a new topic chain is formed, and an overt pronoun *ta* ‘she’ at the start of sentence (c) qualifies as the new topic, which inherits the referential value of the previous topic *Zou Xinping*. The new topic chain lasts until the end of the paragraph, as all of the following actions took place in the same location. As a consequence, all of the subsequent subjects were realized by a null form.

- (70) (a) *Zou Xinping tiaoxia mache*, (b) [-] *xiang tuolaji zouqu*.
 Zou Xinping jump-off horse-carriage toward tractor go
 (c) *ta jinru jiashishi*, (d) [-] *ba tuolaji kaidao lu bian*, (e) [-] *mie le huo*,
 she enter cab BA tractor drive-to road side kill Pst engine
 (f) [-] *que meiyou liji xiala*.
 but not immediately get-off

‘(a) *Zou Xinping* jumped off the horse carriage and (b) [-] walked to the tractor. (c) **She** entered the cab, (d) [-] drove the tractor to the roadside, (e) [-] killed the engine, (f) but [-] didn't climb down right away.’ (Pu & Pu 2014: 37)

The second case is defined as thematic shift, where the narration of physical actions is changed to a state of mind, or the narration of a state of mind is changed to physical actions. For example, in (71), the sentences (a) to (c), which describe a series of physical actions, form a topic chain, where *Zen Huixin* serves as the topic, while the subject of (b) and (c) are realized by null forms. From sentence (d), it began to describe the state of mind of *Zen Huixin*, which causes a thematic shift. As a result, the previous topic chain ends here and a new topic chain is formed, with the pronoun *ta* ‘she’ as the new topic, which inherits the referential value of the previous topic as well.

- (71) (a) *Zen Huixin dingzhe chuanxintougu de hanfeng*, (b) [-] *yongjin-di*
 Zen Huixin brave bone-chilly DE cold wind all-strength
tui zhe xiao che (c) [-] *waiwaixiexie-di zou zai qiqubuping de tulu shang*,
 push PRG small cart wobbly walk in rugged DE dirt-road on
 (d) *ta ganbudao leng...*
 she not-feel cold

‘(a) *Zeng Huixin* braved the cold wind, (b) [-] pushed the cart with all (her) strength, (c) [-] staggering along the rugged dirt road. (d) **She** didn't feel cold ...’

(Pu & Pu 2014: 38)

There are also other two types of Minor Thematic Discontinuity proposed by Pu & Pu 2014, namely Emphatic Effect and Weakened Topicality, which will not be discussed here, due to their irrelevance to the topic of the current study.

From the above examples, the following aspects concerning nominal and pronominal resolution in Chinese can be summarized:

(72) (a) the null subject is not always available (or preferential), even if it is easy to recognize its referential value. As shown by (69c) and (70d), there is no signal of topic change and the subject apparently refers to the previous topic, but the null subject is not preferred in those situations.

(b) the use of a null subject does not depend on whether there is continuity of the same topic; rather, it depends on whether there is continuity of topic chain.

(c) there are various criteria to determine topic chains in Chinese. Those criteria may vary according to different speakers and situations.

(d) several sentences may share the same topic, but are divided into different independent topic chains. In this case, some topic chains may start with an overt pronoun *ta* ‘he/she’ or *tamen* ‘they’, which inherits the referential value of the topic of the previous chain and serves as the new topic.

It is reasonable to assume that the use of an overt pronoun or a full NP marks the beginning of a new topic chain, while the null subject is only used to refer to a topic referent, but does not serve as the head of a topic chain. This is a difference between null and overt subject pronouns in Chinese.

There are also two differences between Chinese and Romance null subject languages, as the latter clearly allow sequences of sentences in which only the first clause has an overt topic and all of the null subjects of the subsequent sentences refer to that topic (see example (69) in section 2.4), while in Chinese this does not always happen. The second difference is that in Romance null subject languages an overt pronoun is preferentially used to refer to a different entity from the previous topic (especially in intra-sentential structures), while in Chinese the overt pronoun may not involve topic switch.

The theory of Pu & Pu 2014 can be compared to the aboutness-shift topic theory of Frascarelli 2007 and its development for Chinese by Liu 2014, who also proposes that the Chinese null subject is related to a topic (and he argues that the topic involved in Chinese is base generated in an A'-position inside CP) and a pronoun may also serve as a topic. Pu & Pu 2014 do not

analyze null subjects in Chinese on a syntactic basis. Their proposal can be complemented by Liu's 2014, who argues that there is an agreement relation between the null subject and the A'-topic in Chinese, and this language does not permit long-distance agreement with null subjects. So, in this case, an overt pronoun must be used, which may inherit the referential value of the previous topic and become the new topic. However, this proposal may not be completely true for Chinese, as (69) clearly shows that this language allows long-distance agreement between the topic and a null subject. Nevertheless, Liu 2014 proposes that the null subject in Chinese is *pro*, which gives us a starting point for analyzing the theory of Pu & Pu 2014 in terms of syntax. There are also some similarities between the proposal of Li 2007 and Liu 2014, as we have discussed in the previous section. One important difference between the two proposals resides in the fact that Liu 2014 used the Agree operation, while Li 2007 only considered the GCR.

4.5 Relations between topic and *pro*

Considering the proposal of Pan 2016, 2017, I will discuss whether *pro* in both matrix clauses and subordinate clauses can undergo an Agree/Match operation with a topic.

First, let us see an example with *pro* inside a subordinate clause⁵⁹. Recall the sentence from Huang 1984, here presented as (73). Huang 1984 considered that *pro* inside the island is controlled by the topic *Zhangsan*, according to the GCR.

- (73) Zhangsan₁, [-]₁ changge de shengying hen haoting.
 Zhangsan sing DE voice very good hear
 'As for Zhangsan, the voice of singing is very nice.'

Now, under the newest theory of MP, it is reasonable to consider that there is some Agree or Match relation between the topic and the subject *pro*. Since *pro* is inside an island, it is not reasonable to consider that there is an Agree operation between *pro* and *Zhangsan*, as Agree is subject to island conditions under Pan's 2016, 2017 theory. So, a better analysis should consider the relation between *pro* and *Zhangsan* as Match, which is not constrained by island conditions. However, the Match analysis may also face one problem. As shown in Huang 1984, if the island is separated from the topic by another nominal constituent, the anaphoric relation is also ruled

⁵⁹ I argued in the previous section that cases with *pro* inside a complement clause (without an island) should be considered as a finite control situation. So, the subordinate clause discussed here only concerns island conditions.

out, see (74).

- (74) *Zhangsan₁, wo xihuan [-]₁ changge de shengyin.
 Zhangsan I like sing DE voice
 ‘As for Zhangsan, I like the voice of singing.’

If the Match operation is possible in (73), then there is no reason for it to be ruled out in (74). In fact, the ungrammaticality of (74) indicates that the null subject must be analyzed as a gap, which is subject to island conditions. Huang 1984 argued that, since there is another nominal element, namely *wo* ‘I’, that intervenes between the topic and *pro*, GCR is not available in this situation.

Considering the difference between (73) and (74), it seems that Match should not be able to explain the anaphoric relation between the topic and *pro* in (73). As a result, the only way to analyze (73) is stick to the GCR, and consider that there is no difference between sentences with a complement structure and those with island structures.

This claim can be further proved by an example with an inside sentence temporal adverbial adjunct. In (75), the embedded *pro* is close to the matrix subject and can receive its referential value thanks to the GCR.

- (75) a. Zhangsan₁ zai [-]₁ chifan de shihou chang le yishou ge.
 Zhangsan prep. eat DE time sing Pst one.CL song
 ‘Zhangsan sang a song when (he) was having meal.’
 b. *Lisi₂, Zhangsan₁ zai [-]₂ chifan de shihou chang le yishou ge.
 Lisi Zhangsan prep. eat DE time sing Pst one.CL song
 ‘As for Lisi, Zhangsan sang a song when (he) was having meal.’

However, in (75), the embedded null subject cannot refer to the left dislocated DP *Lisi*, due to the intervention of the matrix subject *Zhangsan*. In summary, examples (73) to (75) show that the GCR still plays an important role concerning anaphoric resolution in subordinate clauses, even under the newest MP framework.

Now we consider the cases where *pro* is in matrix subject position.

- (76) Zhangsan huidao jia, [-] tuo le yifu. Zhihou, [-] kaishi zuofan.
 Zhangsan arrive home undress Pst cloth after begin cook
 ‘Zhangsan arrived home and undressed. Then (he) began to cook.’

The first null subject should be considered as a gap generated by ATB movement, since the two clauses can be analyzed as in a coordination relation, as discussed in section 4.3. Now consider the second null subject, in a juxtaposition structure.

In Liu's 2014 theory, there is a null topic in the CP structure of the second juxtaposed sentence. This null topic agrees with *Zhangsan*, the topic and subject of the first sentence, and receives its reference, and then agrees again with the null subject of the second sentence, assigning it the reference of *Zhangsan*.

Now we compare the proposals of Pan 2016, 2017 and Liu 2014. There are two questions that should be answered here. First, is the relationship between the null topic of the juxtaposed sentence and the null subject of the preceding sentence an Agree operation or a Match operation? Second, which feature is involved in such an operation?

For the first question, there is no obstacle for the Agree operation, since there are no island structures involved. Furthermore, considering feature interpretability, which will be discussed below, the Agree operation should also be superior to the Match operation.

For the second question, Liu 2014 did not clearly indicate which feature triggered the agreement, but since his theory is based on the A-topic of Frascarelli 2007, it should be reasonable to consider that, in Liu's 2014 theory, that feature is the phi-feature, as proposed by Frascarelli. In Pan's 2016, 2017 theory, the features involved in Agree and Match are [variable] and [phi]. Even in the cases of object *pro*, Pan 2017 still considers that there is variable feature matching between *pro* and its antecedent. However, the sentences analyzed by Pan are different from those analyzed by Liu, since, in Pan's examples, the A'-dependency between a topicalized constituent or a relative head and the gap, *pro* or pronoun is obligatory, while in Liu's examples the reference of *pro* is free (i.e. not bound), though it may be preferentially interpreted as referring to the topic⁶⁰. In this sense, the variable feature used in Pan's analysis should not be used in the case of (76), since the referential subject *pro* may not serve as a bound variable.

However, the phi-feature alone may not indicate the real referential value of a pronominal constituent either (Camacho 2013). Here, I propose that the subject *pro* should also contain an

⁶⁰ In fact, the structures analyzed by Pan 2016, 2017, namely relative clauses and left dislocation structures, are believed to involve movement in the GB framework, while the referential pronoun (overt or null) analyzed by Liu 2014 does not necessarily involve movement.

uninterpretable Definite feature, in the sense of Li 2007. The uD feature of *pro* then agrees with the interpretable D feature of the null topic and *pro* can therefore receive the referential value of the null topic⁶¹. Here, since there is feature valuing between interpretable and uninterpretable features, it is better to consider that such an operation involves Agree, and not merely Match⁶². In summary, we have just shown that, in Chinese, the subordinate null subject is licensed and identified by GCR, while the null subjects in matrix clause of juxtaposition structures can be generated by the Agree operation. However, there are still some problems to be solved regarding some topic properties.

Liu's 2014 analysis is able to explain the relation between the head of a topic chain and the successive null subjects inside a topic chain. However, according to the structure in (77) (copied from (67) in section 2.4), the analysis requires that the head of the topic chain should be in a higher position (CP), which may not always be the case. This is what happens, for example, in a sentence like (78).

(77) [CP John₁ [TP *pro*₁ hen congming]], suoyi [CP Top₁ [TP *pro*₁ chang na
 John very intelligent consequently always get
 diyi-ming.]]
 first-place
 'John is very intelligent, so he always gets the first place.' (Liu 2014: 210)

(78) [CP Jinguan [TP John₁ hen congming]], dan [CP Top₁ [TP [-]₁ rengran buneng na
 although John very intelligent but still cannot get
 diyi-ming.]]
 first-place
 'Although John is very intelligent, (he) still cannot get the first place.'

In (78), it is not possible to consider that the subject of the first clause (*John*) is located in the CP domain, since C is occupied by *jinguan* 'although'⁶³. However, the null subject of the second clause may still refer to *John*. These facts may indicate some incompatibility between the theories of Pu & Pu 2014 and Liu 2014. But, with some adaptations, it should be possible to

⁶¹ Thus, in my proposal, the D feature valuing can occur in matrix clauses, which is different from Li's 2007.

⁶² Interestingly, if we consider Huang's 1984 original proposal, which analyzed the matrix null subject as generated by A'-movement, the Agree analysis may still hold, because, according to Pan 2016, A'-movement in the current framework can be exactly analyzed as an Agree or Match operation.

⁶³ If we adopt the proposal of Pan 2016, who argues that in concessive clauses the functional conjunction *jinguan* 'although' can be analyzed as an adverb, but not a C head, the result still holds in (78). Since the subject is below the adverb, it is clearly not in the topic position.

analyze Chinese anaphoric resolution under the theory of topic chain.

For instance, the main reason why Liu 2014 did not propose that Chinese allows long-distance agreement is that he assumes that every null topic in Chinese must agree with an overt topic. Nevertheless, in Pu & Pu's 2014 theory it is not relevant whether the null topic enters an agreement relation or not, since, in their analysis, each topic chain only has one topic. In that case, a topic chain may last as long as possible since there is no Minor Thematic Discontinuity, which allows the long-distance agreement between the topic and the null subject. As a result, this is not a serious objection to Liu's 2014 theory of Chinese null subjects. Concerning the issue raised by (78), I will propose an analysis to connect the theories of Pu & Pu 2014 and Liu 2014.

In a previous analysis of null subjects in EP, Lobo 1994, 1995 argued that there is an operator located in the CP domain, which makes the connection between discourse and the reference value of *pro*. Inspired by these studies, I propose that, in Chinese, each clause contains an operator in the CP domain, which is a constituent without phonetic realization. To maintain the compatibility with Liu 2014, I will continue to use the term topic to refer to this operator. The referential value of this topic should be marked by the head of the topic chain that contains the topic. The topic (operator) may also agree with a *pro* and assign it a referential value.

Here, I would argue that the topic chain I propose in this thesis does not merely refer to a series of nominal elements that share the same reference, but also refers to a series of sentences/clauses that are closely related in terms of discourse-pragmatics. Recall the examples used in Pu & Pu 2014 (68) to (70): it is clear in Chinese that topic chain is related to the semantic and pragmatic properties of a series of sentences, and not only related to a series of nominal elements.

In this sense, I propose that the head of a topic chain is the core of what a topic chain is about. The head of a topic chain is not related to a certain syntactic position in the sentence, but serves as an abstract concept in the discursive context. That is to say, when several sentences/clauses form a topic chain, it is the context that determines which constituent from the topic chain (or from somewhere outside of the chain) serves as the head. The head of the topic chain may be a constituent in a subject position, object position, or even in a previous sentence outside of the chain, but the determination is only made by the discourse information, and not by the syntactic

position of the constituent.

So, I assume that there are two types of topics. One is the discursive topic, which is the head of a topic chain, determined by the context and stored in speakers' memory as an abstract concept. The other one is the syntactic topic, which is a null form located inside the CP domain of each sentence/clause. The referential value of the null topic is also determined by the head of topic chain.

This proposal may be proved by the following example, which already appeared in (51) above. Imagine that a speaker is waiting for someone, and suddenly he sees the person he waited for and said:

(79) [-] Lai le.
Come Pst
'(He) came.'

In Huang's 1984 original analysis, he proposed that the null subject is generated by A'-movement of a null operator, whose referential value is determined by the context. My proposal functions in a similar way. Despite being formed by a single sentence, (79) presents a topic chain (only has one sentence), whose head is determined by the situational information and receives the reference value of 'the person who the speaker is waiting for'. Then the head of the topic chain will assign this reference value to the null operator (topic) in the CP domain of the sentence⁶⁴. Finally, the null operator agrees with the null subject (here considered as *pro*) and assigns it its referential value.

In this way, sentence (78) can be analyzed in the following way, with the structure presented in (80): (i) The two clauses of (80) form a topic chain. (ii) The context determines that *John* (which is, at the same time, the subject of the first clause) serves as the head of the topic chain. (iii) Both clauses contain a topic in the CP domain. (iv) The head of the topic chain attributes a referential value ('John') to the topic of each clause. (v) The topic of the second clause (receiving the reference value 'John') agrees with the null subject of the second clause and attributes to it the referential value of 'John'.

⁶⁴ I propose that if the topic chain contains more than one sentences/clauses, then the head of topic chain (the abstract topic) will assign its referential value to the operator of all of the sentences/clauses inside the topic chain.

- (80) [_{CP} Top₁ [_C Jinguān] [_{TP} John₁ hén cōngmíng]], dān [_{CP} Top₁ [_{TP} *pro*₁
 although John very intelligent but
 rēngrán bùnéng nà dìyī-míng.]]
 still cannot get first-place
 ‘Although John is very intelligent, (he) still cannot get the first place.’

There are some differences between my proposal and that of Liu 2014. Firstly, in the analysis of Liu 2014, the head of a topic chain is located in the CP domain, while in my proposal, it is always the operator that occupies this position, while the head of a topic chain is an abstract concept. Secondly, Liu 2014 proposed that even the first clause of a topic chain should contain an empty subject, but in my proposal, there is no need to have such an empty subject. Furthermore, in Liu’s 2014 analysis, the feature triggering the agreement is a phi-feature. However, just as what I have argued before, the feature triggering the agreement should be a D feature. With my proposal, it is thus possible to explain the topic chains in Chinese which permit a constituent which is not in the highest position to serve as the head of a chain, and consequently attributes a referential value to the remaining null subjects inside the chain.

4.6 Re-analyzing Chinese forward and backward anaphora

Developing this proposal of topic chain, I will now re-analyze some cases of Chinese anaphora. First of all, I have to claim that we will not discuss null subjects which can be analyzed as gaps. As argued before, the null subject inside an island will be analyzed as a gap generated by A’-movement, if it cannot be controlled by the closest nominal element. Such a gap will cause ungrammaticality since it must be generated by A’-movement, which is excluded from the island structures. I will also not analyze sentences that contain a left dislocated DP, because any empty category generated by the A’-movement of the dislocated DP is also a gap, which is interpreted as a variable. Furthermore, I will not analyze null subjects generated by ATB movement in coordinate structures either. As a result, the null subjects discussed in the current section will be those generated by Agree or by GCR.

I would like to start from some examples of forward anaphora.

- (81) (a) Zhāngsān₁ huídào jiā, (b) [-]₁ kāishǐ zhūnbèi zuòfān.
 Zhāngsān come-back home begin prepare cook

(c) Zai [-]₁ qie cai de shihou, (d) [-]₁ tingdao le yi-zhen zaoyin.
 prep. cut veggies DE time hear Pst one-CL noise

‘(a) Zhangsan₁ arrived home, (b) [-]₁ began to prepare the dinner. (c) when [-]₁ cut the vegetables, (d) [-]₁ heard a noise.’

In these sentences, all of the null subjects (including that in the adverbial adjunct) may refer to *Zhangsan*. In this case, it should be assumed that all of the four clauses form a topic chain, with *Zhangsan* serving as the head of the chain. *Zhangsan* then attributes its referential value to all of the null topics inside the chain, which agree with the respective null subjects, in the cases of (b) and (d)⁶⁵. In the case of (c), since island structure is involved, it is not possible to consider that there is Agree operation. In this case, I propose that the null subject in (c) is *pro* which is controlled by the null topic via GCR⁶⁶. In this way, all of the null subjects in the topic chain receive the reference value of *Zhangsan*. However, it is also possible to use an overt pronoun in the adverbial adjunct, resulting in sentence (82):

(82) (a) Zhangsan₁ huidao jia, (b) [-]₁ kaishi zhunbei zuofan.
 Zhangsan come-back home begin prepare cook
 (c) Zai ta₁ qie cai de shihou, (d) [-]₁ tingdao le yi-zhen zaoyin.
 prep. he cut veggies DE time hear Pst one-CL noise

‘(a) Zhangsan₁ arrived home, (b) [-]₁ began to prepare the dinner. (c) when he₁ cut the vegetables, (d) [-]₁ heard a noise.’

According to the intuitions of native speakers of Chinese, there is no difference between (81) and (82) concerning their meanings, since *ta* ‘he’ in (82) should also refer to *Zhangsan*.⁶⁷ However, since an overt pronoun is used in (82), it may indicate the start of a new topic chain (see the analysis of Pu & Pu 2014). In this sense, (82) may be divided into two topic chains. The first one is (a) and (b), with *Zhangsan* serving as the head of the chain. The second chain is formed by (c) and (d), with *ta* ‘he’ being the head of the chain. Here, the pronoun *ta* ‘he’ itself does not have a referential value (because one can only know the phi-features of the pronoun, but is not able to know the exact identity of the constituent, in the sense of Camacho 2013: 122);

⁶⁵ However, it is possible to analyze (81a)-(81b) (also repeated in (82a)-(82b)) as a coordinate structure. In this case, the null subject could be generated by ATB movement.

⁶⁶ Though Huang & Yang 2013 argued that GCR requires the antecedent of *pro* to be overt, the example (81) shows that a null element may also control a *pro*, as long as it can be recovered by a previously presented overt topic.

⁶⁷ The reason why (81) and (82) are both correct and convey the same meaning can be attributed to the fact that different speakers may apply different criteria to determine the topic chains.

however, this pronoun can inherit the referential value of the previous topic. As in (80), the pronoun first inherits the referential value of *Zhangsan* and then values the topic of the clause in (d)⁶⁸, and, afterwards, the topic of the second clause agrees with the null subject of the sentence and attributes to it the referential value of *Zhangsan*. What should be noted is the fact that a null subject cannot function alone as the head of a topic chain, while an overt pronoun can. As a result, the availability of a null subject depends on whether there is a head for the topic chain.

Another aspect that should be noticed here is the fact that although (81c) and (82c) have the apparent structure of backward anaphora, their subjects have indeed a forward anaphora interpretation, which seems to be preferred whenever it is possible, as suggested by Filik & Sanford 2008, who analyzed similar structures in English.

In an eye-tracking task, which tested sentences like (83)⁶⁹, Filik & Sanford 2008 found that the native speakers of English have longer first-time pass and total time for the pronoun and the verb that comes after the pronoun when there is no previous antecedent (a). This result suggests that the readers prefer a forward anaphoric interpretation for an embedded pronoun in a left-dislocated adverbial adjunct.

(83) a. The final day of the conference had been pretty exhausting. After he returned to the hotel, Brian immediately fell asleep.

b. Brian found that the final day of the conference had been pretty exhausting. After he returned to the hotel, he immediately fell asleep.

(Filik & Sanford 2008: 1116)

Now we turn to the case of backward anaphora. In (84), the start of the sentence is an adverbial adjunct, which contains a null subject.

(84) Zai [-]₁ qie cai de shihou, Lisi₁ chang le yi-shou ge.
 prep. cut veggies DE time Lisi sing Pst one-Cl song
 ‘When [-]₁ cut the vegetables, Lisi₁ sang a song.’

In this sentence, the null subject appears in first position, but it cannot be considered as the head

⁶⁸ Here, according to my proposal, it is the abstract discursive topic *ta* ‘he’, but not the subordinate subject *ta* ‘he’, that values the reference of the null topic (operator) in the CP domain.

⁶⁹ Filik & Sanford 2008 also tested two other conditions. Since they are different from the structures discussed here, I will not present them.

of the topic chain, since in my proposal a null subject may not have this function. Hence the availability of the null subject has to depend on the head of the topic chain where it is located. One way to analyze this structure is to consider that the head of the topic chain that contains this null subject is located in the previous context. However, since this context is not referred in the utterance, it is difficult to recover the head of this topic chain. As a consequence, the alternative analysis is to consider the matrix subject *Lisi* as the head of the topic chain. In fact, the preferred interpretation of (84) is to accept coreference between the null subject and the matrix subject *Lisi*, which indicates that native speakers prefer the backward anaphoric reading, rather than the forward anaphoric reading, which would involve a previously unreferred antecedent for the null subject.

It is also not adequate to think that the preferential interpretation of (84) is caused by the fact that the adverbial adjunct is generated by movement, since Pan & Paul 2018 have shown that subordinate-main order is the default order for Chinese temporal adverbial adjunct, which is not generated by movement. Moreover, if we consider examples with overt pronouns, it is evident that the coreferential reading is only available in forward anaphora, and not in backward anaphora, see (85).

- (85) a. Zhangsan₁ he le yibeī cha, zai ta₁ kan shu de shihou.
 Zhangsan drink Pst one-Cl tea prep. he read book DE time
 ‘Zhangsan had a cup of tea, when he was reading the book.’
 b. Zai ta₂ kan shu de shihou, Zhangsan₁ he le yi bei cha.
 prep. he read book DE time Zhangsan drink Pst one-Cl tea
 ‘When he was reading the book, Zhangsan had a cup of tea.’

As a result, if the left dislocated adverbial clause in (85b) was generated by movement, the coreferential reading of the pronoun should be preferred, just as in (85a). However, the interpretative contrast between (85a) and (85b) proves that the adverbial clause in (85b) is not generated by movement. This case further proves that the subordinate clause in (84) is not generated by movement.

Hence, we can analyze (84) by assuming that the matrix subject *Lisi* serves as the abstract discursive topic, which marks the referential value of the topic of the adverbial clause backwards, which consequently controls the embedded null subject via GCR. I assume that it

is in this way that the backward anaphoric relation is established.

In fact, as already mentioned in section 3.1 of Chapter 3, Biller-Lappin 1983 argues that backward anaphora stands as a conventional way to introduce a new referent. Based on this proposal, Ariel 1990 proposes that the apparent backward anaphora structures (subordinate-matrix) can be divided into two groups: the real backward anaphora structure and structures that should be considered as forward anaphora.

The second case can be exemplified just as (81c-d) and (82c-d), where the antecedent of the anaphoric expression (null or overt pronoun) can be found in the previous sentence. Ariel 1990 considers this situation as a forward anaphora situation, since the referential value of the null subject or pronoun in (c) is not determined by the matrix subject of (d), though they may be co-indexed.

Ariel 1990 then considers that the real backward anaphora structure is the one that introduces a new referent, for example, the case of (84). Although (84) has a subordinate-matrix structure, with the embedded clause containing a referentially dependent entity, the parser, when processing such a structure, already knows that the antecedent of the referentially dependent entity should be in the matrix clause, which will be processed later. So, the apparent subordinate-matrix structure is in fact used to introduce the matrix subject, which serves as the new topic.

Ariel 1990 then argues that, in null subject languages, it is always the emptier form (namely, the null subject) that establishes the backward anaphora relation with the matrix subject, while the overt one normally does not favor the coreference interpretation with the matrix subject. In this sense, I propose that for the apparent backward anaphora structure which introduces a new referent, a mechanism which is similar to the active search of Kazanina et al. 2007 is activated. For example, when the parser encounters a structure like (84), with a null subject in the left-dislocated subordinate clause, it will activate this mechanism and consider that the antecedent of the null subject will be found in the upcoming matrix clause. However, if the subordinate clause has an overt pronoun as its subject, no such mechanism should be activated.

Now we see an example with an overt pronoun. In (86), the adverbial adjunct with an overt pronominal subject is located at the start of a sentence.

- (86) Zai₁ ta₂ qie cai de shihou, Lisi₁ chang le yi-shou ge.
 prep. he cut veggies DE time Lisi sing Pst one-Cl song
 ‘When he₂ cut the vegetables, Lisi₁ sang a song.’

The pronoun *ta* ‘he’ appears in the first place. Since there is no previous context, it is possible to consider the subordinate clause as the start of a topic chain, with *ta* ‘he’ serving as the head of the chain. Here, the referential value of *ta* ‘he’ should be inherited from the topic of the unmentioned previous context, which is not revealed in the utterance. But this does not change the fact that *ta* ‘he’ serves as the head of a topic chain. However, in any case, it is not possible to analyze (86) in the same way as (82). In (82), the subject of the following clause in (d) is null, and it may receive the referential value of the head of the topic chain (the pronoun) through agreement with the topic. But in (86), the subject of the second clause is *Lisi*, a full NP, which should not be included in the topic chain which is headed by *ta* ‘he’. This implies that *Lisi* marks indeed a new topic chain. If this analysis is on the right track, then it is less likely that *Lisi* inherits the referential value of *ta* ‘he’, since, according to Centering Theory, if a referent is introduced in the form of a pronoun, then it cannot be reintroduced by a full NP, for reasons of coherence. This explains why in Chinese coreference between an overt pronoun and the matrix subject is not the preferred interpretation in the condition of backward anaphora.

Now it is natural to ask why (86) cannot be analyzed in the same way as (84), with the matrix subject valuing backwards the reference of the embedded pronoun. This in fact is the difference between null and overt pronominal subjects. The null subject, which lacks phonetic realization, is referentially dependent on other constituents, which facilitates the activation of the active search mechanism. The overt pronoun, on the other hand, is relatively more independent than the null one, since it can function as the head of a topic chain. Thus, it may not depend on the active search mechanism to determine its reference. As a consequence, based on Ariel 1990, I propose that in Chinese only the null subject serves to activate the backward anaphora interpretation. In this way, the overt pronoun in (86) should not activate the same mechanism as in (84), so the coreference between the pronoun and the matrix subject in (86) is not supported by any mechanism, and therefore it is not the preferred interpretation.

Finally, we see the case of forward anaphora with overt pronoun. The example (87), which has the reverse clause order of (86), may permit the coreferential reading between the NP and the

overt pronoun.

- (87) Lisi₁ chang le yi-shou ge, zai ta_{1,2} qie cai de shihou.
 Lisi sing Pst one-Cl song prep. he cut veggies DE time
 ‘Lisi₁ sang a song, when he_{1,2} cut the vegetables.’

I propose that, in (87), the two clauses consist of two different topic chains, with the NP and the overt pronoun as their respective heads. In this case, since the head of the second topic chain is a pronoun, it can inherit the referential value of the first topic chain, which explains the possibility to have a coreferential reading. Note that in (86), the NP cannot inherit the referential value in the same way, as explained before. As a result, the interpretative contrast between forward and backward anaphora with overt pronouns can be explained by the possibility of referential value inheritance.

The same kind of interpretative preference can also be found with conditional clauses, see the examples (88). In (a), when the subordinate subject is null, it can refer to the matrix subject which appears after the subordinate clause. However, when the subordinate subject is overt, the same coreferential reading is not possible (b).

- (88) a. Ruguo [-]₁ neng xiewan zuoye, Zhangsan₁ jiu qu kan zuqiu.
 if can finish homework Zhangsan then go watch football
 ‘If (he) can finish the homework, then Zhangsan will go to watch football.
 b. ??Ruguo ta₁ neng xiewan zuoye, Zhangsan₁ jiu qu kan zuqiu.
 if he can finish homework Zhangsan then go watch football
 ‘If (he) can finish the homework, then Zhangsan will go to watch football.’

As a consequence, it should be noticed that pronominal and null subject resolution is constrained by the theories of topic chain and topic change, while the condition of null subject in backward anaphora structures should be considered as an exception, due to the conventional way of introducing a new referent in the theory of Biller-Lappin 1983.

Compared to other null subject languages, including the Romance null subject languages, it has been found that Chinese does not differ from those languages concerning the interpretation of null subjects, both in forward and backward anaphora. Chinese is also similar to Romance null subject languages with respect to the interpretation of overt pronouns in backward anaphora. The only difference remains in the condition of overt pronoun in forward anaphora, where

Chinese prefers a coreferential reading, while Romance null subject languages prefer a disjoint reading. The similarities and differences between Chinese and Romance null subject languages can be explained by the fact that these two types of languages present different discursive properties concerning pronominal resolution. For Romance null subject languages, the interpretation of pronouns is closely related to the syntactic position of the antecedent, following proposals such as PAH. For Chinese, the current study is able to explain that pronominal resolution is related to the topic chain analysis, which bears a different nature from the syntactic position analysis.⁷⁰ As a result, these two kinds of languages may present some differences when concerning pronominal resolution, though they may show some similarity in the case of null subjects.

In the previous analysis, I propose that anaphoric resolution in Chinese can be explained by the topic chain maintenance and topic switch proposal, which was developed by Pu & Pu 2014. In this language, the null subject can agree with the topic (operator), while an overt pronoun may itself serve as the head of a topic chain. I propose that an overt pronoun can inherit the referential value of a previous topic referent, while an NP cannot, which results in the interpretative contrast between forward and backward anaphora with overt pronouns.

In this way, the analysis based on topic chains is able to neutralize the structural differences between forward and backward anaphora, since, when considering the continuity and change of topics, the linear order should rank higher than the syntactic structure. The proposal presented in this study does not analyze Chinese anaphoric resolution considering the syntactic function of the antecedents, which is the main analysis focus for other null subject languages, such as Romance null subject languages, since Chinese may apply a different mechanism from those languages. The analysis I propose for overt pronouns in backward anaphora contexts is not based on syntactic conditions either, such as cyclic c-command or quantifier raising, which have been shown to have shortcomings. Thus, the current proposal is able to explain Chinese anaphoric resolution under the theory of topic chain and topic change, which is highly related to discourse and pragmatics. In fact, anaphoric resolution is normally considered as an issue in

⁷⁰ Once more, some properties related to the Chinese topic chain analysis may also hold in Romance null subject languages, especially concerning the interpretation of null subjects. However, there are some differences between these two types of languages, since in Romance null subject languages it is easier for the overt pronoun to convey a topic shift interpretation, which can be explained by the influence of PAH.

the domain of discourse and pragmatics (Sorace & Filiaci 2006, etc). As a result, I argue that the proposal of the current study explains Chinese anaphoric resolution more adequately.

In summary, this chapter first reviewed Pan & Paul's 2018 analysis for the Chinese conditional adverbial clauses and proposed a similar analysis for the Chinese temporal adverbial clauses, which implies that the interpretative properties predicted by cyclic-command are a consequence of a parsing ambiguity. The second part of this chapter proposed a topic chain analysis on the basis of the analysis of Pu & Pu 2014, which is able to explain Chinese forward and backward anaphora in terms of discourse-pragmatics.

Chapter 5 Theories and experimental studies on L2 acquisition

This chapter will review some previous studies concerning L2 acquisition. Starting with a brief presentation of the studies concerning acquisition of syntactic properties, the second half of the chapter will concentrate on the acquisition of discourse-pragmatic properties, which is closely related to the IH of Sorace & Filliaci 2006.

5.1 L2 acquisition for syntactic properties

L2 acquisition has been a topic studied by linguists for many years. The earlier studies concerning L2 acquisition concentrated on the acquisition of syntactic properties, assuming that human beings are equipped with the ability to learn language since birth, namely the Faculty of Language, which includes a UG. The UG determines what grammars are possible and provides a list of possible categories and features. According to Chomsky 1981, the UG consists of Principles and Parameters. The Principles are general conditions for all kinds of natural languages spoken in the world, while the Parameters are aspects in which different languages may vary. The grammar of all human languages is constrained by Principles and Parameters.

The initial state of a child is UG. When a child is born, he or she does not have knowledge of the grammar of any particular language, while the UG only provides him/her the principles about how a grammar can be. After having contact with the input of L1, the child will be able to successively reconstruct his/her grammar, valuing the features of his/her grammar. Finally, when reaching the final state of his/her L1, all of the parameters should be properly set.

L2 acquisition may be different from L1 acquisition, since the learners already dominate their L1 grammar. A major question in the investigation of L2 acquisition is whether the learners use UG to construct their L2 grammar, in the same way as in L1 acquisition.

It should be noticed that L2 acquisition by children can be quite different from L2 acquisition by adults (Schwartz 2003). In the current thesis, only adult L2 acquisition will be discussed.

Some authors such as Flynn 1987, White 1985, 1989 argue for the Full Access proposal, which suggests that the L2 grammar is completely constrained by UG, meaning that L2 acquisition should be similar to L1 acquisition. Other authors, namely Bley-Vroman 1990, Clahsen & Muysken 1989 and Schachter 1988, proposed that the L2 grammar is not constrained by UG,

since there are crucial differences between L1 and L2 acquisition. This kind of proposal is often known as No Access or Partial Access.

In L2 acquisition, the initial state is considered as the linguistic knowledge with which the learners begin to construct the grammar. There are different hypotheses concerning the initial state of L2 acquisition.

Schwartz & Sprouse 1994, 1996 proposed the Full Transfer and Full Access hypothesis, which claims that the initial state of L2 is all the properties of the L1 grammar. When the grammar of the L1 does not correspond to the L2 properties (illustrated by L2 input), the learners have access to UG in order to reconstruct their L2 grammar. During the process of acquisition, the learners may experience several intermediate states, where their L2 grammar is different from their L1 and from the final state of L2. Their grammar during the intermediate states is known as interlanguage.

Vainikka & Young-Scholten 1994, 1996 a, b proposed the Minimal Trees Hypothesis, which claims that the functional categories of the L1 cannot be transferred to the initial state of L2, and hence only lexical categories are transferred in the initial state. The functional categories of the L2 only appear throughout the process of acquisition, in a bottom-up fashion. Eubank 1993/94, 1996 proposed the Valueless Features Hypothesis, which admits that the functional categories of L1 can be present in the initial state of L2, but remain unspecified. Throughout the learning process, learners are able to value the features.

Flynn & Martohardjono 1994, Flynn 1996 and Epstein et al. 1996 proposed the Full Access without Transfer Hypothesis, which argues that the initial state of L2 is not a particular grammar, but UG. This hypothesis indicates that L2 acquisition is very similar to L1 acquisition, which is completely constrained by the UG. The learners, as long as they are exposed to the L2 input, are able to successively generate interlanguage grammars via direct access to UG, and finally reach the final state of L2.

There are also some hypotheses about the stages after the initial state. These hypotheses normally concentrate on whether the learner's UG can help them to reset the parameters. For example, authors such as Clahsen & Hong 1995, Neeleman & Weerman 1997 and Beck 1998 propose that there is a breakdown when setting the parameters. According to this proposal, the L2 grammar may contain properties that do not belong to natural language, constituting a wild

language.

Clahsen & Hong 1995 and Neeleman & Weerman 1997 are in favor of the Global Impairment Hypothesis, which suggests that there are no parameters in L2. Rizzi 1982 proposed that one parameter may be related to a series of properties. The L1 learners, after valuing a certain parameter, will simultaneously acquire all of the properties related the parameter. However, in L2 acquisition, the simultaneous acquisition of the related properties of a certain parameter is not always verified, which may support the Global Impairment Hypothesis.

The Local Impairment Hypothesis, which is suggested by Beck 1998, is considered as an extension of the Valueless Feature Hypothesis. According to this hypothesis, the L2 grammar may have features, but the learners may have problems with the parameters associated with functional categories. The difference between the Local Impairment Hypothesis and the Valueless Feature Hypothesis resides in the fact that the latter proposes that the learners are able to reset the features during the acquisition process, while the former claims that the deficit is permanent for L2 learners.

Authors such as Hawkins & Chan 1997 proposed that the parameter values of the L1 will be completely transferred to the L2 and there is no parameter resetting. In this way, the learners are not able to acquire parameter values which are different from their L1.

Hawkins & Chan's hypothesis contrasts with the Full Transfer and Full Access Hypothesis of Schwartz & Sprouse 1994, 1996, according to which the L2 grammar is completely constrained by the UG. All of the parameters related to functional categories are available and the learners are able to set and reset the features in L2.

Since the main focus of the study is the acquisition of properties in the domain of discourse and pragmatics, I will not discuss the hypotheses above in the remainder of the current thesis. In the next section, I will present the IH, which is highly related to discourse-pragmatic properties and was frequently used to explain the results of studies on anaphoric resolution.

5.2 L2 acquisition of Interface properties

The IH was first suggested by Sorace & Filiaci 2006, who noticed that near-native L2 speakers

may be influenced by their L1, showing some undetermined knowledge or optionality⁷¹ when acquiring those aspects that involve the interface between syntax and other cognitive systems, which may not happen when acquiring narrow syntax. As a result, Sorace & Filiaci 2006 propose the IH, which predicts that the properties from narrow syntax can be completely acquired in L2, while those involving interface between syntax and other cognitive domains may not be completely acquired in L2. They also suggest that this hypothesis can be extended to bilingual L1 acquisition and L1 attrition. According to Sorace & Filiaci 2006, the optionality may be caused by the underspecification of the knowledge representation level, or by the insufficient processing resources to integrate information from different domains.

This hypothesis was confirmed by various studies, such as Pérez-Leroux & Glass 1999, Lozano 2002, 2009, Tsimpli et al. 2004, Tsimpli & Sorace 2006, Montrul & Rodriguez-Louro 2006, Margaza & Bel 2006, Rothman 2007, 2008, Belletti, et al. 2007, among others. These studies demonstrate that the properties involving the interface between syntax and discourse-pragmatics are difficult in L2 acquisition.

Sorace 2011 summarizes those studies under the IH. According to Sorace 2011, the interface can be understood as syntactic structures which are sensitive to conditions of varying nature. These conditions must be satisfied in order to guarantee that these structures are grammatical or pragmatically adequate. The interface can involve different language modules, as well as non-linguistic cognitive systems, considering Ramchand & Reiss 2007.

Sorace 2011 also points out that various studies, such as Tsimpli & Sorace 2006 and Ramchand & Reiss 2007, show that the various types of interfaces may not be identical. The interface between syntax and discourse, which involves language and pragmatic processing, is on a higher level comparing to the interface between syntax and semantic, which only involves the language system. The first interface is considered as external, while the latter is internal.

The IH indicates that some syntactic structures involve interfaces, while others do not. However, Sorace 2011 also argues that sometimes it is difficult to determine if a structure involves the interfaces. Furthermore, some structures may involve various types of conditions and it is unclear which type of interface is relevant.

⁷¹ Optionality can be understood as inconsistent performance of the L2 speakers, that is to say, they fluctuate between native-like performance and non-native-like performance.

Sorace 2011 also explains why there is optionality when acquiring structures involving interfaces by using two explanations which had already been presented in Sorace & Filiaci 2006. The first explanation is the underspecification of the interpretable features. The examples that Sorace uses here are the acquisition of null and overt subjects in null subject languages (for example in Italian) by learners whose L1 does not have null subjects. Some studies (cf. Sorace & Filiaci 2006, Belletti et al. 2007) show that L2 learners always have difficulty when using and interpreting null and overt pronouns in null subject languages, especially with respect to the overgeneralization of overt subject pronouns in contexts where the null subject is preferred. Sorace considers that the pragmatic-syntactic interface system is relatively simpler in English, since the overt pronoun may be used either in topic-shift or in topic maintenance contexts, while the null subject is almost impossible. However, the pragmatic-syntactic interface system in Italian is relatively more complicated, due to the fact that the null subject is used in the topic maintenance context, while the overt subject indicates topic shift. In other words, the English overt pronoun may have both the [topic shift] and [topic maintenance] features, while, in Italian, the overt pronoun only has the [topic shift] feature and the null subject has the [topic maintenance] feature, which indicates that English has underspecification of the two features. In this way, the language with the simpler interface system may influence the language with a more complex interface system, which results in influence of the L1 English over the L2 Italian. However, this kind of influence may not happen inversely, because the language with a more complex interface system cannot influence a language with simpler interface system. As a result, Italian cannot influence English in the case of null and overt subjects. However, Sorace also noticed that even L2 learners whose L1 has null subjects may have difficulty when acquiring the L2 null and overt subjects. Sorace argues that one of the explanations is to consider that the null subject scope may not be identical in all of the null subject languages.

Sorace 2011 also proposes that the optionality observed in the L2 acquisition of interface structures may be caused by the processing costs. The author believes that structures which involve two domains (for example, syntax and discourse) have more processing costs than structures which involve only one domain (for example, only syntax). One of the examples that Sorace adopts is the PAH of Carminati 2002 (which was discussed in section 3.2 of Chapter 3), who proposes that, in Italian, the antecedent of a null subject is preferentially a category located

in the position of SpecIP, while the antecedent of an overt pronoun occurs in a lower position. In this way, the processing of a structure with an overt pronoun always depends on discursive information, which is costly for the L2 learners, who eventually show some difficulties when using and interpreting overt pronouns in Italian.

Sorace & Filiaci 2006 believe that the IH only applies to near-native speakers, which is questioned by White 2011, who argues that, if some properties are difficult for near-native speakers, these difficulties should also be present during the process of acquisition. As a result, the IH should be extended to all levels below the near-native level.

There are many studies on L2 acquisition, bilingual acquisition and L1 attrition that corroborate the IH, which cover a variety of null subject languages as target language, for example Italian (Sorace & Filiaci 2006, Belletti et al 2007, Serratrice 2007), Spanish (Rotheman et al 2007, 2008, 2009), EP (Madeira et al 2009, 2010, 2012) and Greek (Tsimpli & Sorace 2006). I will present first Sorace & Filiaci 2006, which first proposes the IH, and then I will present some studies on the L2 acquisition of EP.

Sorace & Filiaci 2006, who originally propose the IH, aim to investigate how near-native speakers of Italian who speak English as L1 acquire the anaphoric resolution strategy in Italian. According to the PAH of Carminati, the native speakers of Italian prefer a subject antecedent for a null subject, and a non-subject antecedent for an overt pronoun.

A Picture Verification Task was used by the authors, which consisted of four conditions: two conditions of forward anaphora (with null and overt pronoun in subordinate clause) and two conditions of backward anaphora (with null and overt pronoun in subordinate clause), see an example in (1). After reading each sentence, the participants were asked to choose the picture (out of three) that better described the sentence presented orally/in writing.

(1) a. Mentre lei_{2/3}/pro₁ si mette il cappotto, la mamma₁ dà un bacio alla figlia₂.
 while she wears the coat, the mother gives a kiss to the daughter
 ‘While she/pro is wearing her coat, the mother kisses her daughter.’

b. La mamma₁ dà un bacio alla figlia₂ mentre lei_{2/3}/pro₁ si mette il cappotto.
 the mother gives a kiss to the daughter, while she wears the coat
 ‘The mother kisses her daughter, while she/pro is wearing her coat.’

(Sorace & Filiaci 2006: 352)

The test results (Table 1) of Sorace & Filiaci 2006 show that the near-native speakers of Italian exhibit a different behavior when compared to the native speakers of Italian. For null subjects, the near-native speakers prefer a subject antecedent, which is still similar to the native speakers' choice. However, for overt subjects, the near-native speakers continue to accept the subject antecedent, while the native speakers prefer the non-subject antecedent (an extra-linguistic entity). The authors also find that the non-target interpretation of the L2 speakers is more obvious in backward anaphora than in forward anaphora.

Group	Forward anaphora					
	Overt pronoun			Null pronoun		
	Other	Comp.	Subject	Other	Comp.	Subject
Native	11	82	8	5	44	51
Near native	13	60	27	11	43	46
Total	12	73	15	8	44	49
Group	Backward anaphora					
	Overt pronoun			Null pronoun		
	Other	Comp.	Subject	Other	Comp.	Subject
Native	64	24	12	4	11	85
Near native	28	25	47	6	9	85
Total	49	24	27	5	10	85

(adapted from Sorace & Filiaci 2006: 354)

Table 1 Results in percentages of the off-line task in Sorace & Filiaci 2006

Sorace & Filiali propose the following explanation for the phenomenon: the near-native speakers of Italian have the null subject in their grammar and they also have knowledge of the PAH. However, they do not have enough processing resources to consistently integrate information from multiple sources. As a result, they show optionality when interpreting structures involving PAH and do not show as clear a preference regarding the interpretation of overt pronouns as the native speakers do. This claim is also supported by the fact that the backward anaphora condition is more problematic than the forward anaphora condition. The authors assume that the processing of backward anaphora is more demanding in relation to forward anaphora. Unlike in forward anaphora, in backward anaphora, the antecedent appears after the processing of a referentially dependent element (null or overt pronoun). In this occasion, according to the authors, when the parser processes the pronoun, an active parser

mechanism will be activated, which requires the antecedent of the pronoun to be encountered as soon as possible, corresponding to a subject antecedent interpretation. It is obvious that the non-native speakers are more likely to rely on such mechanism, which leads them to accept the subject antecedent more easily than the native speakers do. As a result, when the processing is more demanding, the behavior of the near-native speakers also becomes undetermined. In this way, the authors claim that the optionality of the near-native speakers is caused by the lack of processing resources.

These findings are also attested in other studies concerning other null subject languages, and the general findings of these studies can be summarized as follows: the L2 learners have difficulties in the overt pronoun condition, as they tend to accept the coreferential reading between the overt pronoun and a subject antecedent. This kind of non-native behavior may be even found with L2 learners who speak an L1 with null subjects (Bini 1993, Margaza & Bel 2006). However, for null subjects, most L2 learners are able to interpret it as referring to a subject antecedent, with the exception of Margaza & Bel 2006 and Rothman 2009.

The IH was also tested in EP by studies like Madeira et al. 2009, 2010, 2012. Since Madeira et al. 2010 is about inflected infinitives, which is not exactly the topic of the thesis, I will only present Madeira et al. 2009 and 2012.

Madeira et al. 2009 sought to analyze the acquisition of the null subject in L2 EP by learners of L1 Romance and L1 Germanic languages, applying two production tasks (oral and written) and a preference judgement task.

The oral production task was conducted through oral interviews with the participants and the contents were later transcribed by the researchers. The written production task was conducted through a composition written by the participants.

The production tasks (both oral and written) showed that learners from both linguistic groups produce null subjects from a very early level. However, the elementary level learners from both the Romance group and the Germanic group use redundant overt pronouns in the contexts where a null subject is preferred. However, for advanced learners, they significantly produce less redundant overt pronouns. The development is confirmed in both groups and in both tasks.

In the preference judgement task, the participants were given a context sentence and then had to choose a continuation from two options, one with a null subject and the other one with an

overt pronoun in the second conjunct of coordination or in the subordinate clause, see example (2).

- (2) Esta noite os pais da Joana não estão em casa.
a. A Joana faz o jantar e depois ela lava a loiça.
b. A Joana faz o jantar e depois lava a loiça. (Madeira et al. 2009: 173)
(Translation: tonight Joana's parents are not at home.
a. Joana cooked the dinner and then she washed the dishes.
b. Joana cooked the dinner and then [-] washed the dishes.')

In (2), the context sentence indicates that it should be *Joana* that washed the dishes, so option *b* should be the target one, since the null subject in the second conjunct of coordination should refer to the subject of the first conjunct. Thus, option *a* has a redundant use of the overt pronoun. This yields the correferential reading condition. However, there are also items in which the context sentence indicates that, for the continuation sentences, the second conjunct of coordination (or the subordinate clause) should have a subject that is different from that of the first conjunct of coordination (or the matrix clause). In these cases, the option with the overt pronoun in the second conjunct of coordination (or the subordinate clause) should be the target one. This yields the disjoint reading condition. The authors then recorded the percentages with which the participants chose the option with overt pronoun in both conditions.

The results (Table 2) show that lower level learners from both linguistic groups choose sentences with overt pronoun when the intended reading is coreferential, which can be understood as a redundant use of the overt pronoun. Higher level learners from both groups show some improvement in this respect.

In the correferential reading condition, for the Romance group, there is development from the elementary level to the intermediate level, since they chose less sentences with redundant overt pronouns (from 15.2% to 5.3% in coordination structures and from 22.4% to 15% in subordination structures). However, for the Germanic group, this progress is only found from the intermediate level to the advanced level. In other words, development is delayed if the L1 of the participant is a non-null subject language. For the disjoint reading condition, the authors verified an over-generalization of null subjects, since the L2 learners from both language groups did not choose the overt pronoun as much as the native speakers did, which indicates that they

thought that the null subject can also be used in the context with referent change.

	Correferential reading			Disjoint reading		
	Coord.	Subord.	Total	Coord.	Subord.	Total
Germanic elementary	19.7	36.1	30.7	81.8	65.4	70.2
Germanic intermediate	14.7	27.9	23.5	94.1	67.1	74.8
Germanic advanced	8.3	14.6	12.5	95.8	61.7	71.4
Romance elementary	15.2	22.4	22.8	85.9	63.9	70.2
Romance intermediate	5.3	15	11.7	89.5	63.2	70.7
Romance advanced	4.2	9.4	8.3	100	73.3	81
Control group	1.6	7	5.2	100	85.6	89.7

(translated from Madeira et al. 2009: 186)

Table 2 Results in percentages for overt pronouns from the off-line task in Madeira et al. 2009

In summary, Madeira et al. 2009 confirm that the L2 EP learners from both groups acquire null subject from an early stage, so there is no L1 transfer in relation to the null subject parameter. However, the discourse-pragmatic aspects of null and overt subjects are more difficult to acquire, as the elementary learners from both groups show some non-target judgements. In this case, the L1 of the learners may play an important role. Learners whose L1 has null subjects may experience earlier development than learners whose L1 does not have null subjects.

Madeira et al. 2012 aim to investigate null and overt subject acquisition in L2 EP by learners of L1 Italian and Chinese. There are three tasks in this study: a writing production task, a selection task and a comprehension task.

In the production task, due to the reduced relevant occurrences, the authors fail to reach any relevant conclusion. The only finding is that the Italian learners may use a null subject to refer to a non-subject antecedent.

In the selection task, the participants were given a context sentence and a complex sentence, with a gap in the position of the subject of the subordinate clause. They were asked to fill in the

blank with an overt pronoun or a null subject, see example (3). In (3a), the context sentence indicates that the subordinate subject of the complex sentence should be the matrix subject, while, in (3b), the context sentence indicates that the subordinate subject of the complex sentence should be the matrix object. The results can be found in Table 3.

(3) a. A Milena chegou a casa às 7h.

A Milena telefonou à Alexandra quando _____ chegou a casa.

a. ela b. –

b. A Márcia chegou a casa às 7h.

A Mónica telefonou à Márcia quando _____ chegou a casa.

a. ela b. –

(Madeira et al. 2012)

(translation: a. Milena arrived home at 7h.

Milena called Alexandra when _____ arrived home.

a. she b. –

b. Márcia arrived home at 7h.

Mónica called Márcia when _____ arrived home.

a. she b.-)

	Contexts favoring subject antecedent		Contexts favoring object antecedent	
	Overt pronoun	Null pronoun	Overt pronoun	Null pronoun
Native speakers	2.8	97.2	90.3	9.7
Italian elementary	14.3	85.7	90.5	9.5
Italian advanced	8.3	91.7	83.3	16.7
Chinese elementary	19.2	80.8	85	15
Chinese advanced	4.2	95.8	91.7	8.3

(translated and transformed from the bar charts of Madeira et al. 2012)

Table 3 Results in percentages for the selection task in Madeira et al. 2012

The results of the task show that all of the participants prefer the null subject when the context favors the coreferential reading between the matrix and the subordinate subjects, while they prefer the overt pronoun when the context favors the coreferential reading between the matrix object and the subordinate subject. There is no relevant difference between the two L1 groups, nor between different levels (elementary or advanced).

The comprehension task aims to analyze how the participants select the antecedent of null and overt pronouns in adverbial adjunct and coordinate clauses. After reading a test sentence with a null or an overt pronoun in the subordinate clause or the second conjunct of coordination, the participants were asked to select between two sentences the one that better described the test sentence, see example (4) and (5). The results can be found in Table 4.

(4) A Inês vive com a Ana desde que [-]/ela se divorciou. (subordinate clause)

A: A Inês divorciou-se. B: A Ana divorciou-se.

(adapted from Madeira et al. 2012)

(translation: Inês lives with Ana since [-]/she got divorced.

A: Inês got divorced B. Ana got divorced)

(5) A Paula viu a Diana e [-]/ela sorriu. (coordinate clause)

A: A Diana sorriu. B: A Paula sorriu.

(adapted from Madeira et al. 2012)

(translation: Paula saw Diana and [-]/she smiled.

A: Diana smiled B: Paula smiled)

The results show that the Italian learners from both levels prefer a subject antecedent for null subjects in adverbial adjunct and coordination contexts, while they prefer a non-subject antecedent for overt subjects. This behavior is very similar to that of the native speakers.

However, the behavior of the Chinese participants is different from that of the Italian participants. For null subjects, the Chinese learners have a similar interpretation to the Italian learners and native speakers, in subordination and coordination structures. For overt pronouns, the Chinese participants from both levels tend to fluctuate between the subject and non-subject antecedent in the subordination condition. In the coordination structures, those from the elementary level also show such fluctuation, while the advanced learners prefer the non-subject antecedent.

According to the results of the three tasks of Madeira et al. 2012, the behavior of the Italian learners is very similar to that of the native speakers, in subordinate or coordinate structures, with null or overt pronouns. These results go against the idea that L2 learners of a null subject language may present difficulties in acquiring the use and comprehension of overt pronouns, regardless of what kind of L1 they speak.

Subordinate clauses				
	Overt pronouns		Null pronouns	
	Subject ant.	Non-subject ant.	Subject ant.	Non-subject ant.
Native speakers	1.1	98.9	93.3	6.7
Italian elementary	7.4	92.6	94.4	5.6
Italian advanced	7.1	92.9	97.6	2.4
Chinese elementary	41.7	58.3	91.7	8.3
Chinese advanced	43.8	56.2	89.6	10.4
Coordinate clauses				
	Overt pronouns		Null pronouns	
	Subject ant.	Non-subject ant.	Subject ant.	Non-subject ant.
Native speakers	0	100	100	0
Italian elementary	3.7	96.3	100	0
Italian advanced	4.8	95.2	95.2	4.8
Chinese elementary	43.3	56.7	93.3	6.7
Chinese advanced	20.8	79.2	100	0

(translated and transformed from the bar charts of Madeira et al. 2012)

Table 4 Results in percentages for the comprehension task in Madeira et al. 2012

However, the Chinese learners did show some problems when interpreting overt pronouns. As a consequence, the behavior of this group of learners still favors the hypothesis that the L2 learners of a null subject language show optionality when acquiring the use and comprehension of overt pronouns.

Lobo et al. 2017 aim to compare forward and backward anaphora in L2 EP acquisition, considering L1 influence. A picture verification task was administered to two groups of learners: L1 German (non-null subject) speakers and L1 Italian (null subject) speakers. Each group was also divided into three subgroups, according to their L2 proficiency: elementary, intermediate and advanced. There was also a control group which consisted of native speakers of EP.

Four conditions were tested in the experiment, which included forward anaphora with null or overt embedded pronominal subject and backward anaphora with null or overt embedded pronominal subject. Each sentence contained an adverbial adjunct, which was located before or after the matrix clause, depending on whether it was a backward or forward anaphora condition. The subject of the embedded clause was a null or overt pronoun, while the matrix clause contained a subject and a direct object. The sentences are exemplified in (6). After reading each sentence, the participants were asked to choose from two pictures the one that better described the interpretation of the embedded subject of the sentence. The results can be found in Table 5.

- (6) a. O pai fotografou o menino quando *pro* se sentou. (null subject forward)
b. Quando *pro* saiu da garagem, a mãe fotografou a menina. (null subject backward)
c. A avó cumprimentou a menina quando ela chegou a casa. (overt subject forward)
d. Quando ele subiu à árvore, o polícia viu o ladrão. (overt subject backward)
(Lobo et al. 2017)

(translation: a. The father photographed the boy when *pro* sent down.

- b. When *pro* left the garage, the mother photographed the girl.
c. The grandmother greeted the girl when she arrived home.
d. When he climbed the tree, the police saw the thief.)

The results of the test reveal that the native speakers of EP prefer a subject antecedent for the null subject, no matter whether it is forward or backward anaphora. However, they prefer an object antecedent for the overt pronoun, especially in the case of forward anaphora.

Considering the results of the learners, it has been verified in all groups that they prefer the subject antecedent for the null subject, and this interpretation was more marked in the backward condition than in the forward condition. The authors cited Fedele & Kaiser 2014 to explain this phenomenon: “in order to minimize the cognitive cost of maintaining an unresolved dependency”, the parser chooses to encounter the antecedent of the pronoun as soon as possible, thus explaining the highly biased subject antecedent interpretation in backward anaphora.

For overt pronouns in the forward anaphora condition, though learners from all groups have a higher acceptance percentage of object antecedents than that of subject antecedent, their percentages of subject antecedent were higher than those of the native speakers. However, there is a development from the elementary level to the intermediate level of the L1 Italian group, as the intermediate learners have a significant lower percentage for subject antecedent. The same

is also true in the backward anaphora condition with overt pronoun, where the intermediate and advanced learners have the same percentage as the native speakers. However, no such development was found in the L1 German group. The authors consider that there is L1 influence, which results in the faster development of the L1 Italian group.

	Forward anaphora			
	Null subject		Overt pronoun	
	Subject ant.	Object ant.	Subject ant.	Object ant.
German elementary	59	41	43	57
German intermediate	68	32	44	56
German advanced	68	32	41	59
Italian elementary	63	37	40	60
Italian intermediate	71	29	27	73
Italian advanced	71	29	27	73
Native speakers	90	10	15	85
	Backward anaphora			
	Null subject		Overt pronoun	
	Subject ant.	Object ant.	Subject ant.	Object ant.
German elementary	76	24	64	36
German intermediate	81	19	57	43
German advanced	82	18	56	44
Italian elementary	80	20	49	51
Italian intermediate	82	18	38	62
Italian advanced	82	18	38	62
Native speakers	90	10	39	61

(translated and adapted from Lobo et al. 2017)

Table 5 Results in percentages for the picture verification task of Lobo et al. 2017

In general, the authors consider that the results of the test corroborate the IH, since the learners show some non-native interpretations in the overt pronoun condition, which involves semantic factors and discursive adequacy and consequently raises the processing costs.

5.3 Summary of the previous experimental studies

So far, we have reviewed several experimental studies that analyzed the interpretation and processing of forward and backward anaphora structures. In this section, I will make a brief summary of those studies, based on the following features:

- (7) a. Forward or backward anaphora, or both;
- b. Online-test or off-line test;
- c. Language tested;
- d. L1 interpretation or L2 interpretation, or both;
- e. Pronoun forms: null, overt or both;
- f. Number of potential antecedents: one, two or more.

Table 6 lists related studies on Italian anaphoric relation⁷². It has been found that the studies on Italian used similar structures, though there may be some difference between them. For example, the first three studies listed in Table 6 used picture verification tasks, while the last one used a comprehension test. The first three studies allowed the participants to choose among the matrix subject, the matrix complement and an extra-linguistic entity as the antecedent of the pronoun, while in the last study the participants had a fourth option indicating that both the matrix subject and complement could be the antecedent.

The results for native speakers in these studies are similar, though there may be some differences. For example, in the first three studies, the native speakers also accept non-subject antecedents for the null pronoun in forward anaphora, but, in Fedele & Kaiser 2014, they have a dominant preference for the subject antecedent in this condition. Also, in the first three studies, the native speakers prefer the extra-linguistic antecedent for the overt pronoun in backward anaphora, while in the last study they fluctuate between the matrix subject, the matrix object and the extra-linguistic antecedent.

⁷² The studies of Serratrice 2007, Belletti et al. 2007 and Fedele & Kaiser 2014 were presented in section 3.5 of Chapter 3.

	Experimental type	Anaphoric expression	L1 or L2	Type of pronoun	Number of options
Sorace & Filiaci 2006	Off-line: Picture verification	Forward and backward	L1 and L2 adult	Null and overt	Three
Serratrice 2007	Off-line: Picture verification	Forward and backward	L1 adult, L1 children and bilingual children	Null and overt	Three
Belletti et al. 2007	Off-line: Picture verification	Forward and backward	L1 and L2 adult	Null and overt	Three
Fedele & Kaiser 2014	Off-line: Comprehension	Forward and backward	L1 adult	Null and overt	Four

Table 6 Summary of studies on Italian pronouns

Table 7 shows the list of related studies on EP anaphoric resolution. Unlike the studies on Italian, the studies on EP listed here do not allow the participants to choose an extra-linguistic antecedent. While Madeira et al. 2012 used a comprehension task, Lobo et al. 2017 used a picture verification task. For forward anaphora, both studies show that the native speakers prefer the subject antecedent for the null subject and the non-subject antecedent for the overt pronoun. Lobo et al. 2017 also found a similar interpretation for backward anaphora, though the acceptance rate for non-subject antecedent was not as high as in forward anaphora.

	Experimental type	Anaphoric expression	L1 or L2	Type of pronoun	Number of options
Madeira et al. 2012	Off-line: Selection and Comprehension	Forward anaphora	L1 and L2 adult	Null and overt	Two
Lobo et al. 2017	Off-line: Picture verification	Forward and backward	L1 and L2 adult	Null and overt	Two

Table 7 Summary of studies on L2 acquisition of EP pronouns

Now we consider the studies on non-null subject languages, in which on-line tasks were used⁷³ (see table 8):

⁷³ The studies in the table below were discussed in section 3.5 of Chapter 3.

	Experimental type	Anaphoric expression	L1 or L2	Type of pronoun	Number of arguments ⁷⁴
Kazanina et al. 2007	On-line: Self-paced reading	Backward	L1 Adult English	Overt pronoun	Two
Kazanina & Philips 2010	On-line: Self-paced reading	Backward	L1 Adult Russian	Overt pronoun	Two
Van Gompel & Liversedge 2003	On-line: Eye-tracking	Backward	L1 Adult English	Overt pronoun	Two
Drummer & Felser 2018	On-line: Eye-tracking	Backward	L1 and L2 German	Overt pronoun	One

Table 8 Summary of the on-line studies on processing of backward anaphora

It should be noticed that, in these on-line studies on non-null subject languages, only one type of pronoun (the overt one) was tested. These studies also did not compare backward anaphora with forward anaphora. While the first two studies applied a self-paced reading task, the last two applied an eye-tracking task.

Considering the results, Kazanina et al. 2007, van Gompel & Liversedge 2003 and Drummer & Felser 2018 all found that the participants prefer to interpret the backward anaphoric pronoun as referring to the matrix subject, which favors the active search mechanism. Kazanina & Philips 2010 show some different results from the other three, but the authors still consider that the processing of Russian backward anaphora can be influenced by the active search mechanism. The previous paragraphs made a brief summary of the previous studies on backward anaphora. Generally speaking, backward anaphora is not analyzed by many studies, in comparison with forward anaphora, especially when considering on-line tests. None of the studies listed above compared the processing of null and overt pronoun in an on-line task. For studies using on-line tasks, the testing sentences only contain one pronominal form, which is the overt one. In all of the studies listed except Kazanina & Philips 2010, the results always favor the coreferential interpretation between the cataphoric pronoun and the matrix subject. As a result, there are not

⁷⁴ Since the objective of on-line task is not to select an option, here it will count the number of arguments contained in the matrix clause of the testing sentence.

many studies that used on-line task to test a language in which the coreference between cataphoric pronoun and matrix subject is not favored.

Regarding the studies on EP, Madeira et al. 2012 and Lobo et al. 2017 did not allow the participants to choose an extra-linguistic antecedent, which proved to be the favored interpretation in Italian. As a result, it would also be interesting to test whether EP also accepts an extra-linguistic antecedent in backward anaphora. There are also no studies on EP backward anaphora with an on-line test.

Considering the L2 participants, Sorace & Filiaci 2006 and Belletti et al. 2007 only tested participants who speak English, a non-null subject language, as their L1. Lobo et al. 2017 did test L2 learners who speak a null subject L1, but did not include L2 learners who speak Chinese, a language which has different interpretations for forward and backward anaphora. Madeira et al. 2012 tested Chinese learners of L2 EP, but did not test backward anaphora. In this sense, it is worth investigating how Chinese learners of L2 EP interpret and process backward anaphora, considering the similarities and differences between the two languages.

Taking the previous points into consideration, the current study proposes to do the following, addressing aspects which have not been tested in the previous studies:

- (8) a. Use an on-line test to compare the interpretation and processing of null and overt pronouns in backward anaphora of EP.
 - b. Test whether the participants also accept the extra-linguistic antecedent in backward anaphora of EP.
 - c. Test how the Chinese learners of L2 EP interpret and process backward anaphora.

In order to investigate how Chinese learners of L2 EP acquire the properties of anaphoric resolution in this language, as well as to characterize anaphoric resolution in EP native grammars, two experiments were administered in the current thesis. The first experiment is an off-line task, which will be presented in Chapter 6. The second experiment is a self-paced reading test, an on-line task which will be presented in Chapter 7.

Chapter 6 Experiment 1 of the study – off-line task⁷⁵

6.1 Research questions

The main objective of the study is to investigate how Chinese learners of L2 EP interpret null and overt pronouns in forward and backward anaphora in EP, considering the influence from their L1.

Though many of the previous studies (Sorace & Filiaci 2006; Rothman 2008; Madeira et al. 2012) have shown that L2 learners have difficulties in acquiring the interpretative properties of overt subjects in forward anaphoric resolution, not many studies have compared the resolution of forward and backward anaphora with Chinese learners of a null subject language.

According to the discussions in Chapters 2 and 3, it is obvious that the anaphoric resolution in EP is inside the domain of interface between syntactic and discourse-pragmatic properties. As a result, the L2 learners should have difficulties in mastering the anaphoric resolution in EP, which is predicted by the IH. However, if the L1 of the L2 learners share some similarities with the target language regarding the anaphoric resolution, then it is interesting to test whether the L1 influence can facilitate the L2 acquisition.

This is exactly the case of Chinese-speaking learners of L2 EP. Though similar to EP in null pronoun resolution,⁷⁶ Chinese differs from EP in the interpretation of the overt pronoun in forward anaphora. However, the two languages are similar in backward anaphora conditions. If the property of the learners' L1 has an effect on their L2 acquisition, then Chinese L2 learners of EP are predicted to behave differently with overt pronouns in forward and backward anaphora structures. That is to say, they may have a native-like behavior in backward anaphora (but not in forward anaphora), since their L1 is consistent with the target language in this condition.

Moreover, L1 influence may vary depending on the proficiency level of L2 learners, for instance, beginners versus highly proficient learners. Studies like Madeira et al. 2009 and Lobo et al. 2017 have shown that L2 EP learners whose L1 is a Romance null subject language display

⁷⁵ The off-line task presented in this chapter has been published in Zheng et al. 2018.

⁷⁶ It is possible that there might be some variation in the interpretation of null forms in EP and in Chinese, due to the characteristics of the null categories involved in the two languages; however, this analysis is not the focus of the present study.

faster development on the interpretation of pronouns, presenting a native-like interpretation since the beginning levels, than learners whose L1 is a Germanic language, who reach native-like interpretations only at higher levels of proficiency. Therefore, considering that Chinese exhibits different interpretative preferences for forward and backward anaphora, it is crucial to investigate how the Chinese L2 EP learners interpret null and overt subject pronouns in backward and forward anaphora conditions and also how their development is influenced by their L1, comparing the impact of the L1 at different levels of L2 proficiency.

In sum, the research question of Experiment 1 can be formulated as follows: Is the resolution of overt and null subject pronouns, in forward and backward anaphora conditions, by Chinese learners of L2 EP influenced by their L1 preferences? Additionally, I also consider whether this influence is similar across different levels of L2 proficiency.

6.2 Questionnaire study

A questionnaire study was conducted to investigate the research question presented above. In this study I manipulated the order of presentation of overt and null pronouns, in backward and forward anaphora, and tested different groups of Chinese L2 learners of EP, B1 and C1 levels. In order to clearly control for the impact of L1 influence in the current experiment, since, to my knowledge, no previous study had tested forward and backward anaphora resolution in Chinese native speakers, I also conducted a pre-test with Chinese native speakers living in China. This test was a translation of the EP test.

6.2.1 Participants

The Experiment 1 was administered to three groups of participants: a group of native speakers of EP, the control group, and two groups of Chinese learners of L2 EP, who act as the experimental groups. One experimental group is composed of learners from the lower intermediate (B1) level, while the other experimental group is composed of learners from the advanced (C1) level. The inclusion of participants at two different proficiency levels allows us to test the development of learners, which is discussed in the next section.

The control group consists of 31 native speakers of EP, who were attending a bachelor's degree at the School of Arts and Humanities of the University of Lisbon, with an age range from 18 to

24 (M=19; SD=1.84). The B1 group consists of 21 Chinese learners of L2 EP, who were attending the Portuguese Language and Culture course in the School of Arts and Humanities of the University of Lisbon, with an age range from 19 to 44, (M=24; SD=6.42). The C1 group is composed of 23 Chinese learners of L2 EP; some of these were attending the Portuguese Language and Culture course in the University of Coimbra, and some were attending a Portuguese Language course at the School of Business and Economics of Nova University of Lisbon, with an age range from 19 to 22, (M=21; SD=0.92). All participants gave their informed consent and participated voluntarily in the experiment. Concerning their L2 proficiency, both L2 groups completed a placement test before the start of their course, which determined their proficiency level in L2 EP.

The participants also gave their personal information regarding knowledge of foreign languages and information about EP proficiency, use and knowledge, which can be seen in Table 9.

B1	Mean time (in years) of learning (SD)	1.75 (1.71)
(n=21)	Mean time (in years) since arrival (SD)	1.67 (1.28)
	Daily contact	90.00%
	Contact at least 2 times per week	10.00%
	Languages the learner was exposed to before 6 years old	48% none; 52% Engl.
	Other known languages	62% none; 38% Engl.; 5% other ⁷⁷
C1	Mean time (in years) of learning (SD)	2.32 (0.57)
(n=23)	Mean time (in years) since arrival (SD)	1.15 (0.37)
	Daily contact	86%
	Contact at least 2 times per week	14%
	Languages the learner was exposed to before 6 years old ⁷⁸	55% none; 45% Engl. ⁷⁹
	Other known languages ⁸⁰	41% none; 55% Engl.; 18% other ⁸¹

Table 9 L2 learners' characteristics in Experiment 1

⁷⁷ This other, reported only by one participant, refers to Korean.

⁷⁸ In the EP group the results were: 45%, none (no other language), 52%, English, and 7% spoke other languages (French and Dutch, some of them spoken by the same participant).

⁷⁹ Some Chinese participants had English classes (very basic level) in the kindergarten, so they reported that they had exposed to English before 6 years old.

⁸⁰ In the EP group the results were: 28%, none (no other language), 45%, English, and 48% spoke other languages (French, German and Spanish).

⁸¹ This includes Korean, Japanese, Russian and French. Some of them spoken by the same participant.

Both proficiency groups display similar characteristics, although the mean time of EP language learning is higher in the C1 group (with a range from 1 year to 3 years of learning) than in the B1 group (which has a range from 0, that is, less than 12 months, to 7⁸² years of learning). In what concerns their length of residence in Portugal, values are quite similar, although slightly higher in the B1 group which has a range from 1 year to 6 years of residence in Portugal. The C1 Group has a lower range, with a variation between 1 and 2 years of residence in Portugal. Most of the participants of both groups use EP regularly on a daily base when compared to use at least twice a week (B1: 19/21 vs. 2/21; C1: 19/22 vs. 3/22⁸³). None reported using EP “at least once a month” or even “rarely”.

6.2.2 Materials and design

To fulfil the research aims, the Experiment 1 was designed to test four different conditions: forward anaphora with null pronoun; forward anaphora with overt pronoun; backward anaphora with null pronoun and backward anaphora with overt pronoun. Therefore, two independent variables were manipulated, namely Anaphora Type and Pronoun Type, with two levels each: backward vs. forward anaphora and null vs. overt pronoun, respectively. Each test condition is exemplified by the sentences in (1). The full set of the test sentences can be found in Appendix 3.

(1) Condition 1: (Forward overt anaphora)

A Maria apagou a luz enquanto ela comia o bolo.

‘Mary turned off the light while she was eating the cake.’

Condition 2: (Forward null anaphora)

A Maria apagou a luz enquanto [-] comia o bolo.

‘Mary turned off the light while [-] was eating the cake.’

Condition 3: (Backward overt anaphora)

Enquanto ela comia o bolo, a Maria apagou a luz.

‘While she was eating the cake, Mary turned off the light.’

Condition 4: (Backward null anaphora)

Enquanto [-] comia o bolo, a Maria apagou a luz.

⁸² I looked at data from this participant (who arrived in Portugal 1 year ago) and from another one who had started to learn EP 5 years previously and had been in Portugal for the last 6 years and their answers are similar to the ones from the other participants in the group. I also ran the analysis with and without these participants’ data and the results were the same in the statistical models with and without them; therefore, I decided to keep these participants in the analysis.

⁸³ Not all participants in this group gave this information.

‘While [-] was eating the cake, Mary turned off the light.’

As can be seen from the examples, each item consists of a matrix clause plus a temporal adverbial clause. In this experiment, only adverbial subordinate clauses, but not coordinate structures, were considered, and the null or overt pronoun always appears in the subject position of the subordinate clause. Condition 1 yields a forward anaphoric structure, where an overt pronoun occupies the subject position of the subordinate clause. Condition 2 also yields a forward anaphoric structure, with an empty subject in the subordinate clause. Condition 3 yields a backward anaphoric structure, where the adverbial adjunct, with an overt subject pronoun, is left-dislocated. Condition 4 also yields a backward anaphora structure, with an empty subject in the subordinate clause.

In all of the four conditions, the matrix clause contains a subject, which serves as a potential antecedent of the embedded pronoun. The verb of the subordinate clause in all of the four conditions has atelic aspect (process), to ensure that the action occurred in an extendable duration. The verb of the main clause has telic aspect, which guarantees that the action of the main verb can be completely included in the duration of the action described by the subordinate verb.

Twenty-four experimental items, with the structure described above, were distributed by four experimental lists crossing anaphora type and pronoun type according to a Latin Square design: all participants saw all sentences in different conditions. In addition to the experimental sentences, forty-eight filler items were constructed and presented in between the experimental items, in a pseudo-randomized order: experimental items were always preceded by at least one filler item. Sentences were presented in a booklet with four sentences by page.

6.2.3 Procedure

Participants were asked to read the sentences and then to answer a question indicating their interpretation of the embedded pronoun (null or overt), by choosing between two options (an example of question-answer pair is presented in (2)). Each sentence was presented without any previous context and was followed by the question and the two possible answers. One of the answer options corresponds to the matrix subject (always using the referred proper name), while

the other was always “another person”, literally (that is, no proper name was used, since no entity was mentioned before). The option for “another person” was also included in some of the filler sentences (see (3), for instance). Moreover, in filler sentences, places or objects that were not mentioned in the discourse were also included, as exemplified in (4). Order of presentation of each option was counter-balanced: left-right, right-left.

(2) A Maria apagou a luz enquanto comia o bolo. ‘Mary turned off the light while [-] was eating the cake.’

Quem é que comia o bolo? ‘Who was eating the cake?’

A. a Maria ‘Mary’ B. uma outra pessoa ‘another person’

(3) Ela saiu muito cedo para ir às compras, embora o marido consiga fazer todas as encomendas pela Internet. ‘She left very early to go shopping, although her husband can place all the orders online.’

Quem saiu muito cedo? ‘Who left very early?’

A. o marido ‘her husband’ B. uma outra pessoa ‘another person’

(4) A Cláudia recebeu um colar de diamantes que pertencia à avó da Luísa. ‘Cláudia received a diamond necklace which belonged to Luísa’s grandmother.’

O que é que a Cláudia recebeu? ‘What did Cláudia receive?’

A. um anel ‘a ring’ B. um colar ‘a necklace’

It is worth noting that the design of this study is different from that of Lobo & Silva 2016 and of Lobo et al. 2017. Firstly, in the current experiment only one referent was mentioned, since only a single sentence, with just one referent, was presented, while in the referred studies there were always two referents; secondly, while in those two studies, participants were not allowed to choose an antecedent which was not referred in the utterance as a possible answer to the question, in my study this possibility was included.

Studies like Ariel 1990 have argued that, even in structures which strongly imply a cataphoric interpretation, the speakers still maintain the possibility to interpret them as forward anaphoric structures. That is to say, they may always consider that there is a possibility that the apparent cataphoric expression takes as its antecedent an expression mentioned in the previous discourse or an extra-discursive entity, that is, a referent not mentioned in the discourse. Considering this question, I included “another person” as a possible answer for the question so that participants could choose between a referent mentioned in the sentence, the matrix subject, and an

unmentioned one.

In this study I opted for sentences with just one referent because this experience is designed to investigate if L1 Chinese may influence the learners' L2 EP. Since the previous studies on Chinese backward anaphora (Lust et al. 1996 and Zhao 2014) only tested sentences with one referent in the matrix clause, I decided to maintain this structure in the Chinese pre-test of the current study, in order to ensure that the Chinese learners would not have a different performance in comparison with the previous studies. To maintain the compatibility between the Chinese pre-test and the EP test, I decided to test sentences with one referent in the EP test as well.

The questionnaire was presented in a paper booklet and completed in class (in the case of the Chinese learners, it was completed during their Portuguese lesson)⁸⁴. Participants were instructed to carefully read the instructions of the test, which were presented on the first page, and then to fill in their personal information regarding gender, age, knowledge of foreign languages and also information about EP proficiency, use and knowledge⁸⁵. The short questionnaire about participants' personal information can be found in Appendix 1. The experimental task was presented afterwards.

An equivalent Chinese version of the test was also administered to a group of native speakers of Chinese, in China, to test how they interpret null and overt pronouns in their L1. This test serves as the preliminary test of Experiment 1 and contains the same four conditions tested in the Portuguese test⁸⁶. To be distinguished from the 'pilot Chinese test' discussed in Chapters 2 and 3, this test is known as the 'Chinese pre-test' in the remainder of the thesis. Twenty-four native speakers of Chinese participated in this experiment. All participants had a university education and their age range was from 24 years old to 40 years old (M=31; SD=4.06).

⁸⁴ Before the application of the test, a pre-test was administered to a group of Chinese learners of L2 EP at B1 level, in order to test if the vocabulary in the test is appropriate for this group.

⁸⁵ In this task, the introduction, as well as the questions about the participants' profile, were presented only in Portuguese.

⁸⁶ As discussed in Chapter 4, I only tested temporal adverbial adjunct clauses with preposition, in order to reduce the parsing ambiguity. In this case, the subordinate subject cannot be parsed as the matrix subject or topic in the backward anaphora condition, which is different from the case of adjunct clauses without preposition.

6.3 Analysis of the results

For the analysis of the data I implemented a Linear Mixed Model (LMM) with *lme4* package (Bates, Maechler, Bolker, & Walker, 2015) in *R* (R Development Core Team, 2008). Since the dependent variable in the current experiment is binomial, I fitted a mixed logit regression model using *glmer* with “logit” link function (Jaeger, 2008). I included in the model all main effects and interactions. The following main effects were included: Anaphoric expression (Forward vs. Backward), Pronoun (Null vs. Overt), Group (EP native speakers, B1-EP-learners, C1-EP-learners). I also included items and participants as random intercepts as well as by-item and by-participants slopes for all main effects. All independent variables were centered to avoid collinearity effects (Jaeger, 2008; Field, Miles, & Field, 2012).

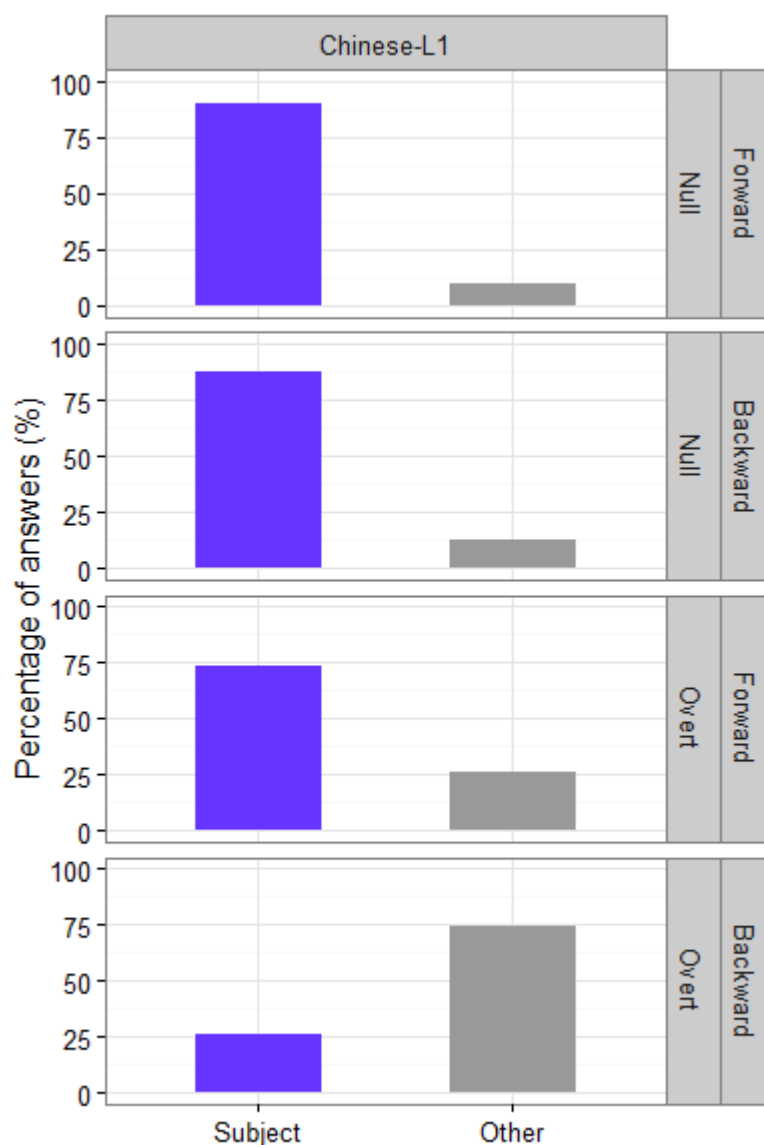
6.3.1. Chinese pre-test

Table 10 presents the LMM effects for the Chinese pre-test and Graphic 1 presents the general percentages of answers. As can be seen, there is a main effect of anaphoric expression and pronoun and also an interaction between these two variables. The main effect of anaphoric expression reveals a preference to interpret the subject as the antecedent of null and overt pronouns in forward anaphora conditions (59%) when compared to backward anaphora conditions (51%). The main effect of pronoun reflects a preference to interpret the null pronoun as referring to the subject antecedent (64%) when compared to the overt pronoun (36%). Effect of interaction of anaphoric expression and pronoun, reveals a difference in the preference to choose the subject as the antecedent in the forward and backward anaphora conditions; moreover, this preference is different for overt and null pronouns, especially in backward anaphora conditions: more subjects with null than with overt.

	Estimate	Std.Error	z-value	Pr(> z)
(Intercept)	-1.881	0.52	-3.64	0.001
Anaphoric Expression	3.495	1.01	3.47	<0.001
Pronoun	4.574	1.03	4.45	<0.001
Anaphoric Expression * Pronoun	2.141	0.95	2.26	0.024

Table 10: LMM effects of Chinese pre-test.⁸⁷

⁸⁷ Statistically significant results are always presented in bold face.



Graphic 1: Results of the Chinese pre-test

The results of the Chinese pre-test show that the native speakers of Chinese prefer the subject antecedent for the null subject in both the forward and backward anaphora structures, obtaining acceptance percentages of 90.3% and 87.5% respectively. For overt pronouns, they also chose the subject antecedent 73.6% of the times when interpreting forward anaphora structures. However, for overt pronouns in backward anaphora, they only chose the subject antecedent 25.7% of the times. It seems that the native speakers of Chinese exhibit distinct interpretations for forward and backward anaphora with respect to the overt pronoun. It should be noticed that the strong preference for the non-subject antecedent for overt pronouns in backward anaphora

may not be constrained by the syntactic factors suggested by Huang 1982 or Lust et al. 1996, as there were still 25.7% of responses favoring the coreferential reading between the matrix subject and the overt pronoun in this condition. In any case, the preference for non-subject antecedent in this condition in Chinese was confirmed by the Chinese pre-test. Overall, the results of the pre-test show that, while there is a general preference to interpret both null and overt pronouns in forward anaphora condition as referring back to the subject of the preceding sentence, in backward anaphora conditions the preference to interpret the null pronoun as referring to the subject of the following sentence is maintained, but the preference for overt pronoun is reversed: participants prefer to interpret the overt pronoun as referring to an extra-discursive referent.

As referred previously, since this is not the focus of the current study, I will not discuss the results of this experiment further. Two aspects, however, are particularly relevant. First, as a null subject language, Chinese does not show the usually described pattern of division of labor between null and overt pronouns in forward anaphora, as the native speakers of this language prefer the subject antecedent for both the null and overt pronouns, which corroborates the theoretical analyses in Chapters 2 and 4. Of course, I admit that this behavior may also be due to the design of my experimental task, since only one referent was presented, and, therefore, participants preferred to interpret the anaphoric pronoun forms as referring to the only discursive explicit referent. This can be explained by the fact that Chinese is considered as a discourse-oriented language and, as discussed previously, coreference in anaphora conditions is determined, mainly, by discursive factors. The second aspect that deserves further and dedicated research is the difference between forward and backward anaphora conditions. Results from backward anaphora conditions display a different pattern from that found in forward anaphora conditions. In backward anaphora conditions with only one referent in the following sentence, null pronouns are interpreted as referring to the subject, while overt pronouns are interpreted as referring to some other entity, not referred in the discourse. These strategies are different from those adopted for pronoun resolution in forward anaphora conditions, as the native speakers prefer the non-subject antecedent for overt pronoun in backward anaphora, which corroborates again the analyses in Chapters 2 and 4. The results also reflect the strategies adopted by consistent null subject languages in backward anaphora

conditions (see Lobo & Silva, 2016, for instance), as in these languages, an overt pronoun is preferentially interpreted as having a non-subject antecedent. It seems that the factors that determine overt and null pronoun resolution in forward and backward anaphora conditions in Chinese are not the same; however, further research is needed to explore these differences in greater depth, namely with conditions which present more than one referent in the sentence, as has been done in similar studies.

6.3.2. EP test

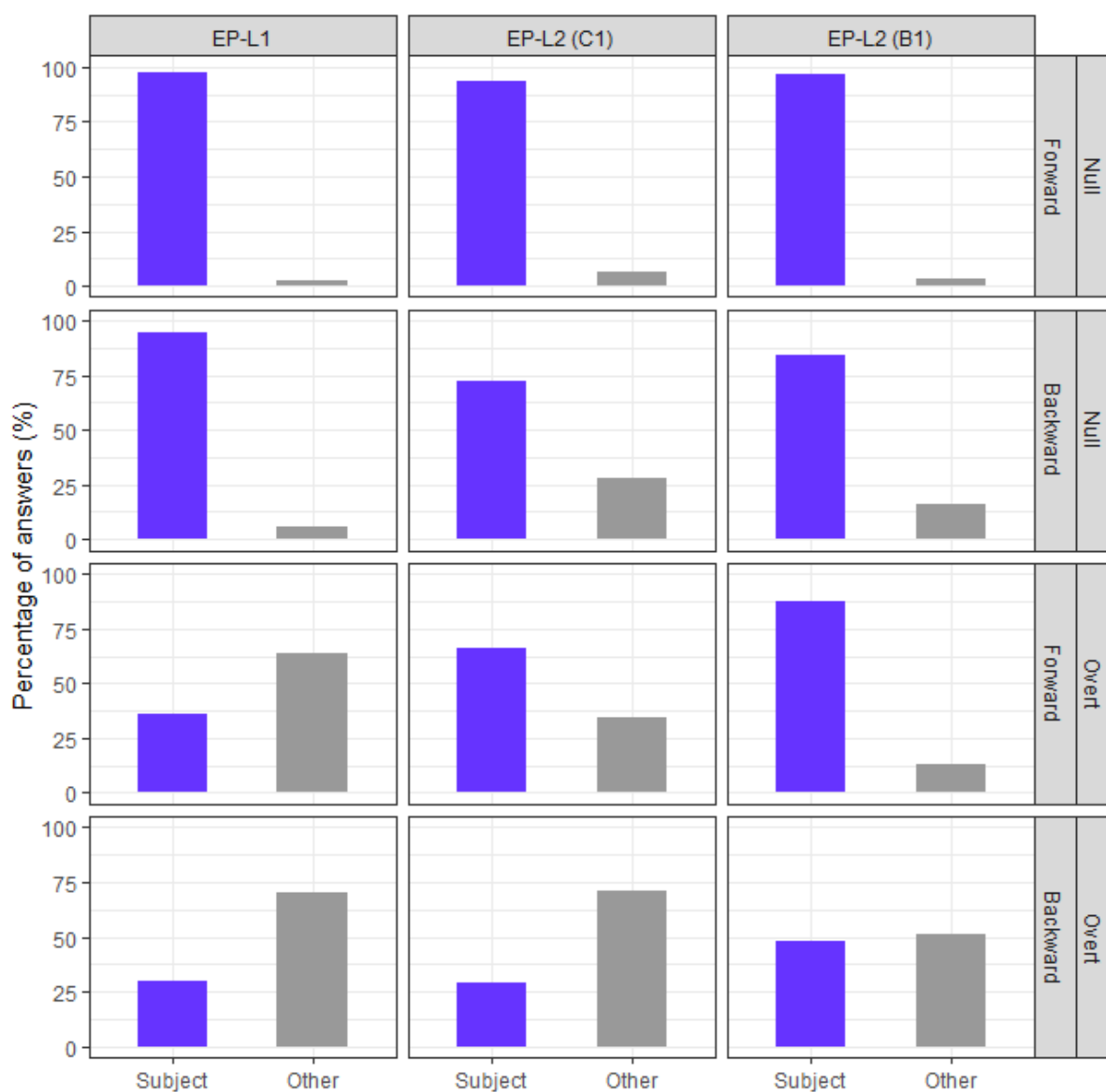
I would like to start the section by presenting the descriptive statistics of the individual performances for all the groups, which can be seen in Table 11. I opted to include the results from the group of the Chinese pre-test to allow for a comparison between learners and their native L1 group although the results are not directly comparable since the test applied to the Chinese native speaker group was a translation of the sentences used in the EP test, that is, it is not the same experiment. This table allows for the comparison of the variability among the different groups in what concerns subject choice (Minimum, Maximum, Mean, Standard Deviation and Subject choice over 50%).

			EP	C1	B1	Native Chinese
			(n=31)	(n=23)	(n=21)	(n=24)
Null	Forwa.	Min-Max	50%-100%	50%-100%	83%-100%	50%-100%
		Mean (SD)	98% (9)	93% (14)	97% (7)	90% (15)
		Choice over 50%	97% (30/31)	96% (22/23)	100% (21/21)	96% (23/24)
	Backw.	Min-Max	17%-100%	0%-100%	17%-100%	50%-100%
		Mean (SD)	94% (18)	72% (32)	84% (22)	88% (17)
		Choice over 50%	94% (29/31)	74% (17/23)	86% (18/21)	92% (22/24)
Overt	Forwa.	Min-Max	0%-100%	0%-100%	0%-100%	33%-100%
		Mean (SD)	36% (35)	66% (37)	87% (24)	74% (25)
		Choice over 50%	29% (9/31)	70% (16/23)	90% (19/21)	71% (17/24)
	Backw.	Min-Max	0%-100%	0%-100%	0%-100%	0%-83%
		Mean (SD)	30% (32)	29% (33)	48% (38)	26% (28)
		Choice over 50%	26% (8/31)	22% (5/23)	43% (9/21)	13% (3/24)

Table 11 Descriptive statistics of subject preferences in all conditions and all groups (Native Chinese group is in gray since results are from the test in Chinese, not in EP)

This table allows for the direct comparison among groups and is especially helpful for the interpretation of statistical results. The results show that the greater difference is in the anaphora condition with overt pronoun when comparing EP with the other three groups and in the B1 group, in the overt backward anaphora condition, when compared with the other groups, but especially with C1 and with EP native speakers.

Graphic 2 presents the general results of the EP test, including choice percentages of the control group and the two learner groups.



Graphic 2: General results of the EP test

In the next sections I present the results for the EP group and also for the contrast between each pair of groups.

6.3.2.1. Native EP

Table 12 presents the LMM effects for the control group of the EP test. As can be seen, there is only main effect of pronoun, which means that the participants show a higher choice for the subject antecedent in null pronoun conditions (75%) than in overt pronoun conditions (25%). However, there is no main effect of anaphoric expression, nor interaction between the pronoun and anaphoric expression. The results of the native EP group reveal that in this language, concerning the structure tested in the current experiment, the native speakers clearly prefer subject antecedents for the null pronoun, regardless of whether it is forward or backward anaphora, while they prefer the non-subject antecedent for the overt pronoun, in both forward and backward anaphora structures.

	Estimate	Std.Error	z-value	Pr(> z)
(Intercept)	-4.2363	1.4278	-2.967	0.003
Anaphoric Expression	1.2098	0.8350	1.449	0.147
Pronoun	11.9107	2.8809	4.134	< 0.001
Anaphoric Expression * Pronoun	-0.4399	1.3753	-0.320	0.749

Table 12: LMM effects of Portuguese L1 speakers

6.3.2.2. L2 EP

In this section, I report results contrasting B1 with C1 and, afterwards, each learner level group, B1 and C1, with EP native speakers.

We first compare the B1 and C1 groups. As can be seen in Table 13, there is a main effect of anaphoric expression, pronoun and group. There is also an interaction of anaphoric expression and pronoun. Overall there is a preference, in the two groups, to choose more subject answers in forward anaphora conditions (86%) when compared to backward anaphora ones (58%). Pronoun effect reflects a preference to interpret the subject as the antecedent of the null pronoun

(87%) when compared to the overt pronoun (58%). Moreover, there is a group effect that reflects a subject preference higher in the B1 group (79%) than in the C1 group (65%). Overall, B1 learners show a subject preference in all conditions except for the overt pronoun in backward anaphora conditions, in which choice is at chance level. The interaction effect between anaphoric expression and pronoun indicates that there are more subject choices in forward anaphora with null (95%) than in backward anaphora with overt (39%).

	Estimate	Std.Error	z-value	Pr(> z)
(Intercept)	-2.180	0.46	-4.79	0.001
Anaphoric Expression	3.349	0.69	4.84	< 0.001
Pronoun	2.662	0.72	3.70	< 0.001
Group	1.835	0.82	2.23	0.026
Anaphoric Expression * Pronoun	1.706	0.69	2.46	0.014
Anaphoric Expression * Group	-0.388	1.18	-0.33	0.743
Pronoun * Group	2.060	1.41	1.46	0.145
Anaphoric Expr. * Pronoun * Group	-0.321	1.35	-0.24	0.812

Table 13: LMM results of B1 versus C1

Now we compare the B1 group with the native speakers. As can be seen in Table 14, there is a main effect of anaphoric expression and pronoun, and an interaction of anaphoric expression * group and pronoun * group. Overall there is a preference, in the two groups, to choose more subject answers in forward anaphora conditions (80%) when compared to backward anaphora ones (64%). Pronoun effect reflects a preference to interpret the subject as the antecedent of the null pronoun (93%) when compared to the overt pronoun (50%). The interaction between anaphoric expression and group indicates that there are more subject choices in forward anaphora conditions in the B1 group (92%) than in the forward anaphora condition in the EP group (67%). The interaction between pronoun and group is a result of less subject choices with the overt pronoun in the EP native speakers' group (33%) when compared to the B1 group (68%). Overall, the B1 group prefers to interpret the subject as the antecedent of both the null and the overt pronouns in all conditions, except in the overt backward anaphora condition, in which the preference for subject or other is at chance level.

	Estimate	Std.Error	z-value	Pr(> z)
(Intercept)	-3.058	0.59	-5.19	<0.001
Anaphoric Expression	2.037	0.60	3.42	<0.001
Pronoun	6.226	1.00	6.20	<0.001
Group	0.202	0.99	0.20	0.839
Anaphoric Expression * Pronoun	0.384	0.90	0.43	0.670
Anaphoric Expression * Group	-2.433	1.11	-2.19	0.029
Pronoun * Group	7.912	1.98	4.00	<0.001
Anaphoric Expr. * Pronoun * Group	-2.241	1.67	-1.34	0.179

Table 14: LMM results of B1 versus EP

Finally, we compare the C1 group and the native speakers. As can be seen in Table 15, there is a main effect of anaphoric expression and pronoun and also two interactions: Anaphoric expression * Group and Pronoun * Group. The main effect of anaphoric expression reflects a preference for subject choice in forward anaphora conditions (73%) when compared to backward anaphora structures (56%). The main effect of pronoun reveals a preference of subject choice with the null pronoun (90%) when compared to the overt pronoun (40%). The interaction of anaphoric expression and group reflects fewer subject choices in forward anaphora condition for the EP group (52%) compared to the C1 (61%), and the interaction of pronoun and group reflects a higher subject choice with null pronoun in the EP group (75%) than in the C1 group (64%).

	Estimate	Std.Error	z-value	Pr(> z)
(Intercept)	-1.542	0.33	-4.74	<0.001
Anaphoric Expression	1.263	0.40	3.13	0.002
Pronoun	4.741	0.32	14.69	<0.001
Group	-0.541	0.53	-1.02	0.306
Anaphoric Expression * Pronoun	0.570	0.65	0.88	0.378
Anaphoric Expression * Group	-1.710	0.69	-2.47	0.013
Pronoun * Group	3.299	0.52	6.32	<0.001
Anaphoric Expr. * Pronoun * Group	0.123	1.04	0.12	0.906

Table 15: LMM results of C1 versus EP

6.4 Discussion

Firstly, I would like to start with the data of the EP control group. There is only a main effect

of pronoun, as the native speakers clearly prefer the subject antecedent for null pronouns, and the non-subject antecedent for overt pronouns, no matter what kind of anaphoric expression is involved. As suggested by the statistical analyses, there is neither main effect on expression, nor interaction between pronoun and expression. Thus, considering the data of this study, it appears that anaphoric resolution in EP (at least for the structures investigated in this study) is linked to the form of the anaphoric expression, but not to the position of the anaphoric expression.

It should be noticed that some other studies have shown that native speakers of EP may be more likely to accept the subject antecedent in the backward overt anaphora condition than in the forward overt anaphora condition (see Lobo & Silva, 2016 and Lobo et al., 2017), although showing the same interpretation bias for the two forms (null-subject and overt-object, although with lower percentages in the latter)⁸⁸. However, as discussed in section 6.2.3, Experiment 1 is different from those applied in Lobo & Silva 2016 and Lobo et al. 2017, in which participants were forced to choose the antecedent of the pronouns between the matrix subject and matrix object. As a result, differences between the results of Experiment 1 and these other studies are expected, since different experimental designs (stimuli and paradigms) were applied.

Considering the results from the group of B1 level learners, it has been found that they do not show many differences, when compared to the native speakers, in the null pronoun conditions, as they prefer to choose the subject antecedent in both the forward null anaphora and the backward overt anaphora conditions. These results confirm that the interpretation of the null pronoun is not problematic in L2 acquisition, even if the L1 and L2 do not present the same properties regarding null subjects (in our case is Chinese vs. EP).

However, for the condition of forward overt anaphora, the B1 learners differ significantly from the native speakers of EP, as they prefer to interpret the matrix subject as the antecedent of the embedded overt pronoun, which is in contrast to the interpretation of the EP native speakers. There are two ways to explain this phenomenon: on the one hand, there may be some L1 influence on the B1 learners, as in their L1 there is also preference to interpret the subject as

⁸⁸ Interestingly, in the on-line test of the current study, which will be described in the next chapter, the native speakers of EP preferred the subject antecedent in forward anaphora, but not in backward anaphora.

the antecedent of the overt pronoun in these structures; on the other hand, previous studies (e.g. Margaza & Bel, 2006) have shown that L2 learners may always prefer the subject antecedent for overt pronouns, no matter what kind of L1 they speak, especially for the learners from lower proficiency levels.

The most striking part of the results comes from the backward overt anaphora condition, as the B1 learners did not show any preference for subject or extra-discursive antecedents. This interpretation is not only different from that of the target language (EP), but also distinct from that of their native language. Actually, in this condition, EP and Chinese native speakers preferentially interpret the overt pronoun in their L1 as referring to an extra-discursive referent. Looking into the individual results, it has been found that there is a major variation inside this group, as 6 of the 21 B1 participants show a clear preference for the non-subject antecedent (over 83% of acceptance), while 8 of them show a clear preference for the subject antecedent (overt 83% of acceptance). The rest of the participants show a chance-level interpretation. We will return to this point after presenting the discussion of the C1 group.

The C1 learners, like the B1 learners, also prefer subject antecedents in both the forward and backward anaphora conditions, when interpreting the null pronouns. However, for the overt pronoun conditions, they behave differently with respect to forward and backward anaphora. In the forward anaphora condition, they also prefer the matrix subject as the antecedent of the overt pronoun, which is similar to the interpretation of the B1 learners. In the backward anaphora condition, however, they prefer the non-referred entity as the antecedent of the overt pronoun.

Overall, the findings of the test confirm the following facts. First, learners from the two levels do not experience difficulties in their interpretation of null pronouns in EP, either in forward or in backward anaphora. This shows that the interpretation of null pronouns is not problematic in L2 acquisition at least when the L1 is a null subject language.

The results also show that the learners have a non-native-like behavior in the interpretation of overt pronouns in forward anaphora. Such a non-native-like behavior has been found in various previous studies, so it is not a surprising result that the Chinese learners from both B1 and C1 levels prefer the matrix subject as the antecedent of the overt pronoun, which contrasts with the judgement of the native speakers. This finding is also consistent with the IH, since the anaphoric

resolution is in the interface domain of syntax and discourse-pragmatics and should be difficult to acquire. The interpretation of the Chinese learners may also be explained by the fact that their L1 also prefers the matrix subject in these structures.

The most interesting part are the results of the condition of overt pronoun in backward anaphora structures. The B1 learners do not perform as consistently as the C1 learners in the overt backward anaphora condition, as some of them display a target interpretation, while others do not. However, C1 learners already exhibit a clear preference for the non-subject antecedent, which is similar to that shown by the native EP speakers.

The research question of the study is whether the L2 learners' null and overt pronoun interpretation, as well as their pace of development, is influenced by their L1, which is answered in the following way. For the forward anaphora structure with overt pronouns, both the B1 and C1 learners showed non-target interpretation when compared to the native speakers, and no development was attested from B1 to C1 level. However, for the backward anaphora structure with overt pronouns, only some of the B1 learners maintained a non-target interpretation, while other B1 learners and most of the C1 learners already interpret the overt pronoun in the same way as the native speakers do. As a result, some of the learners show a different pace of development regarding different structures. Hence the developmental differences between forward and backward anaphora may be caused by L1 influence.

As discussed in section 2.5, Chinese and EP display different interpretations in the forward anaphora with overt pronoun condition. However, they are identical in the case of backward anaphora with overt pronoun condition, since both in Chinese and EP, an overt pronoun in backward anaphora is not preferentially interpreted as referring to a subject antecedent. As discussed in section 2.5 (for EP) and Chapter 4 (for Chinese), such an interpretative preference in both languages can be explained by discourse-pragmatic properties, though there is some difference between the two languages. For example, in EP, the anaphoric resolution is related to the structural/syntactic position, while in Chinese it is related to topic chains. But anyway, the two languages share some similarities on the properties concerning backward anaphoric resolution.

As a consequence, the differences regarding L2 development between forward and backward anaphora attested in this experiment may be a direct cause of the similarities and differences

between the two languages. In forward anaphora, when EP and Chinese have different interpretative preferences, the L2 learners have some difficulties in mastering the anaphoric resolution of EP, which can be explained by the IH. However, in backward anaphora, the L1 of the learners is similar with EP regarding the interpretative preference for overt pronoun. As a result, such a similarity may reduce the acquisition difficulty predicted by the IH and facilitate the L2 learners' acquisition of the interpretative preference of EP, which results in the improvement from the B1 to the C1 level.

In fact, a similar phenomenon is also tested in studies like Madeira et al. 2009 and Lobo et al. 2017. In these studies, two groups of L2 learners of EP were compared, which performed in different ways concerning their development. Both the studies compared L2 learners whose L1 is a Germanic language (lacking referential null subjects) and a Romance null subject language. In both of these studies, it was found that L2 learners of both groups from the initial levels show difficulties in interpreting the overt pronouns in EP. However, the L1 Romance language group showed quicker and earlier development from the lower to the higher levels than the L1 Germanic language group. Thus, the authors of these two studies concluded that there was L1 influence, which determined the differences regarding the pace of development between the Romance and Germanic groups.

As a result, it is possible that the L1 helped some of the B1 and most of the C1 Chinese learners to develop a similar interpretation to the native speakers' in the backward overt anaphora condition. However, for the forward overt anaphora condition, where the L1 Chinese differs from the target language, no such development has been found in the current experiment.

6.5 Conclusions of the off-line task

The current off-line questionnaire task aimed to test how two groups of Chinese learners (B1 and C1 level) of L2 EP interpret null and overt subject pronouns in EP and compare the conditions of forward and backward anaphora, considering the influence from their L1. The experimental test provided the following findings.

Firstly, the study demonstrates that, in EP, the native speakers show a clear interpretative bias for the null and overt pronouns, regardless of whether forward or backward anaphora is concerned. They prefer to interpret a null pronoun as referring to a subject antecedent, while

interpreting the overt pronoun as referring to a non-referred antecedent in both kinds of anaphora.

Secondly, like other previous studies, the current study also demonstrates that the L2 learners do not show much difficulty in acquiring the interpretative properties of the null pronouns, as the learners at both intermediate and advanced proficiency levels tend to accept a subject antecedent for the null pronouns of EP, and there is no difference between forward and backward anaphora.

However, results from the current experiment show that the L2 learners display some non-native-like behaviors regarding the interpretation of overt pronouns, in both forward and backward anaphora, which can be explained by the IH. The non-native-like behaviors appear to be more apparent in the forward anaphora condition than in the backward anaphora condition. In the former case, learners from both levels prefer the subject antecedent and there is no change from the B1 to the C1 level. In the latter case, though some of the B1 learners show a non-native-like interpretation, the others already have a native-like interpretation, and the C1 learners also exhibit a native-like interpretation in the backward overt anaphora condition. The different developmental route for forward and backward anaphora can be explained by L1 influence, as, in Chinese, native speakers also prefer a non-subject antecedent in the backward overt anaphora condition, while in the forward anaphora condition, they prefer the subject antecedent for overt pronouns.

As a result, these findings indicate that, in L2 acquisition, L1 influence may play an important role in the learners' development of anaphoric interpretation. The similarity between L1 and the target language may facilitate the development in their L2 acquisition. The following chapter presents the second experiment of the current thesis, an on-line self-paced reading task.

Chapter 7 Experiment 2 Self-paced reading task

7.1 Research questions

The off-line test of this thesis shows that the L2 learners of EP who speak Chinese as their L1 present different interpretative preferences in forward and backward anaphora, and such differences are supposed to be caused by their L1 influence. Now the current on-line test aims to test if the L2 learners present the same difference in real time processing and how their L1 influences their real time processing.

Concerning the forward anaphora conditions, the L2 learners prefer the subject antecedent for both the null and overt pronominal embedded subject, while the native speakers only have such a preference with the null embedded subject. It is crucial to test if this difference between the native and non-native speakers is reflected in on-line language processing.

For the case of backward anaphora, according to the active search mechanism of Kazanina et al. 2007, or the active parser of Sorace & Filiaci 2006, once the parser encounters a referential dependent constituent (the null or overt pronoun at the start of the sentence, inside the subordinate clause), such a mechanism will be activated, requiring the parser to encounter its antecedent as soon as possible, which favors a subject antecedent reading. Sorace & Filiaci 2006 even argue that such a mechanism should be adopted by the L2 learners, since they have limited resources to establish anaphoric relations with constituents outside of the utterance they read. As a result, they should prefer the matrix subject as the antecedent of the embedded subject. However, the performances of the L2 learners in Experiment 1 only favor this mechanism in the null subject condition of backward anaphora. In backward anaphora with overt pronoun, both the B1 and the C1 learners show a tendency to accept the non-subject antecedent. Hence it would be interesting to investigate whether the active search mechanism is completely absent for the case under study, or whether it is still activated in real time processing, so that their L1-like interpretation in the off-line test is only a result of a late decision influenced by the L1. In order to answer this question, the current on-line test will compare the processing of the native speakers and that of the L2 learners in both forward and backward anaphora structures.

In summary, these are the two research questions of this on-line test:

- (1) Will the L2 learners show the same interpretative preference of the off-line test also in the on-line test, that is to say, will they treat forward and backward anaphora differently in the on-line test?
- (2) Will the L2 be influenced by the active search mechanism? That is to say, is their interpretative preference in the off-line test only a late decision?

7.2 Self-paced reading study

A self-paced reading test was conducted to investigate the research questions presented in (1) and (2). Like the off-line test, in this on-line test I also manipulated the order of presentation of overt and null pronouns, in backward and forward anaphora, and tested one group of native speakers of EP and one group of L2 learners of EP who speak Chinese as their L1. I consider that this study will allow us to further understand the real time processing of overt and null pronoun in both forward and backward conditions by Chinese-speaking learners of L2 EP, and analyze whether there are some differences between real time processing and off-line interpretation.

7.2.1 Participants

The Experiment 2 was administered to two groups of participants: a group of native speakers of EP, the control group, and one group of Chinese learners of L2 EP, who acts as the experimental group⁸⁹. The L2 level of the experimental group is different from that in the off-line test. In this on-line test, I only tested L2 learners with upper intermediate (B2) and advanced (C1) level. This is due to the fact that the self-paced reading requires the participant to have a higher proficiency for the target language, and may therefore be difficult for learners at B1 level. As a result, Experiment 2 did not include learners at B1 level, which was different from the off-line test.

The control group consists of 25 native speakers of EP, who were attending a bachelor's degree at the School of Arts and Humanities of the University of Lisbon, or at the Polytechnic Institute of Leiria, with an age range from 18 to 56, (M=22; SD=7.51). The L2 group consists of 30 Chinese learners of L2 EP, who were attending the Portuguese Language and Culture course in the School of Arts and Humanities of the University of Lisbon, or at the Polytechnic Institute

⁸⁹ None of the participants of Experiment 1 (native speakers and L2 learners) participated in Experiment 2.

of Leiria, with an age range from 19 to 24, ($M=21.39$; $SD=1.23$). All participants gave their informed consent and participated voluntarily in the experiment. The test was approved by the ethics commission of the School of Arts and Humanities of the University of Lisbon.

Concerning their L2 proficiency, the L2 learners completed a placement test before the start of their course, which determined their L2 EP proficiency level. Two participants from the L2 group were excluded as they failed to reach an accuracy rate of 70% in the filler items when processing the on-line test (see the descriptions below) and one other was excluded because she started learning Portuguese at the age of 10.

In the end, I analyzed the data of the other 27 L2 participants, who were 10 B2 learners and 17 C1 learners. Since the number of participants for both levels does not reach 24, which is a plausible number to make statistical analysis, I analyzed the L2 participants as a unique group. In fact, in a complementary analysis, I compared the performances of the B2 and C1 learners and did not find differences between the two levels, except an interaction effect of match and level in the critical region of forward anaphora⁹⁰. Though such an effect deserves to be analyzed, since there is no difference in other regions, and due to the difficulty in finding additional learners at these levels, I decided to treat the learners from the two levels as a unique group.

Like in the off-line test, I also present some descriptive statistics of the characteristics of my samples considering the short questionnaire about language knowledge and use that the L2 participants completed at the beginning of the experimental session⁹¹. This questionnaire can be found in Appendix 2. As can be seen in Table 16, the average learning time of Portuguese of the L2 learners is three years (with a range from 1 year to 5 years⁹²). The average length of residence in Portugal is less than one year (with a range from 4 months to two years). All the L2 participants use EP regularly on a daily basis.

⁹⁰ The effects manipulated in Experiment 2 will be presented in section 7.2.2.

⁹¹ In this task, the introduction, which was presented in the informed consent sheet, has a Portuguese version and a Chinese translation, while the questions about the participants' profile, were presented only in Portuguese.

⁹² One participant did not give this information.

n=23	Time (in years) of learning (SD)	3.00 (1.71)
	Time (in years) since arrival (SD)	0.67 (0.40)
	Daily contact	100%
	Languages the learner was exposed to before 6 years old ⁹³	64% none; 36% Engl.
	Other known languages ⁹⁴	32% none; 64% Engl.; 7% other ⁹⁵

Table 16: L2 learners' characteristics in Experiment 2

7.2.2 Materials

Like the off-line test, the current on-line test was also designed to test four different conditions: forward anaphora with null pronoun; forward anaphora with overt pronoun; backward anaphora with null pronoun and backward anaphora with overt pronoun. In order to force or partially force an interpretation towards the subject or 'someone else', gender information was used. As can be seen in examples (3) to (10), the gender of the pronoun or the past participle either matches or mismatches the gender of the embedded subject (forward anaphora) or of the matrix subject (backward anaphora).

The forward and backward anaphora conditions were treated as separate experiments. As a result, inside the forward or backward anaphora conditions, two independent variables were manipulated, namely Pronoun Type (Null Pronoun vs. Overt Pronoun) and Match Type (Subject Match vs. Subject Mismatch).

The test sentences were presented segment-by-segment (segment breaks are illustrated by the slash sign (/) in the examples), which can be exemplified from (3) to (10). The full set of the test sentences in the forward and backward anaphora conditions can be found in Appendix 4.

Let us first see the forward anaphora conditions:

(3) Condition 1 Null-Match

Quando/a Rita/contou/a história,/ficou relaxada/com a reação/dos amigos.

When/the Rita/told/the story,/became relaxed.fem.sg/with the reaction/from+the friends

⁹³ In the EP group the results were: 68%, none (no other language), 28%, English, and 12% spoke other languages (French and Spanish, some of them spoken by the same participant).

⁹⁴ In the EP group the results were: 28%, none (no other language), 72%, English, and 48% spoke other languages (French, German, Spanish, Italian and Chinese, some of them spoken by the same participant).

⁹⁵ This includes Japanese and Spanish. One participant speaks both English and Spanish.

(4) Condition 2 Overt-Match

Quando/a Rita/contou/a história,/ela ficou relaxada/com a reação/dos amigos.

When/the Rita/told/the story,/she became relaxed.fem.sg/with the reaction/from+the friends

(5) Condition 3 Null-Mismatch

Quando/a Rita/contou/a história,/ficou relaxado/com a reação/dos amigos.

When/the Rita/told/the story,/became relaxed.masc.sg/with the reaction/from+the friends

(6) Condition 4 Overt-Mismatch

Quando/a Rita/contou/a história,/ele ficou relaxado/com a reação/dos amigos.

When/the Rita/told/the story,/he became relaxed.masc.sg/with the reaction/from+the friends

‘When Rita told the story, (he/she) became relaxed with the reaction from the friends.’

As can be seen from the examples, each item consists of a matrix clause plus a temporal adverbial clause, which precedes the matrix clause. In each condition, the subordinate clause contains a DP which may serve as a potential antecedent of the matrix subject, thus yielding a forward anaphora situation. It should be noticed that this design for the forward anaphoric sentences is different from that in the off-line test, in which the matrix sentences always precede the subordinate clause in forward anaphora. Just as Carminati 2002 pointed out, in an on-line test, a subordinate-matrix structure is better than a matrix-subordinate structure, since in the former structure, it is easier for the parser to maintain the information obtained in the first clause. Considering this issue, I also adopt such a structure in my test.

Condition 1 contains a null matrix subject that matches, in gender, the subordinate subject. Condition 2 contains an overt pronoun in the matrix clause, and there is gender match between it and the subordinate subject. Condition 3 contains a null matrix subject, and there is gender mismatch between it and the subordinate subject. Condition 4 contains an overt pronoun in the matrix clause, also with gender mismatch between it and the subordinate subject. In this way, there are two independent variables with two levels each: Pronoun Type (Null Pronoun vs. Overt Pronoun) and Match Type (Subject Match vs. Subject Mismatch).

To force the interpretation of the null pronoun, I adopted structures with passive or adjectival form, where the past participle or adjective has gender information at the final syllable. In this way, the passive sentences allow us to conduct the gender disambiguation, which will be explained in the following part.

In all of the forward anaphora conditions, the segment that contains the main subject serves as the critical region, while the following segment (a PP) serves as the post-critical region. The final segment then serves as the post-post-critical region. The length (number of characters and spaces) of the critical region and the post-critical region were properly controlled.

It should be noticed that, in the Null-Mismatch and Overt-Mismatch conditions, since there is gender mismatch between the main and the subordinate subjects, the participants are forced to interpret the matrix null subject or overt pronoun as referring to an entity other than the subordinate subject, while in the Null-Match and Overt-Match conditions, the subordinate subject may serve as the antecedent of the matrix null or overt pronoun, though a disjoint interpretation is also possible⁹⁶.

Thus, the disambiguation process can be explained in the following way. If the participant prefers the coreferential reading between the subordinate subject and the matrix subject (null or overt), then he or she will spend more time reading sentences with gender mismatch between the subordinate and the main subject than sentences with gender match. This is because, in the case of gender mismatch, the gender information prevents the participant from accepting the interpretation he or she prefers, thus increasing the reading time. Such reading time differences may be shown by the reading time of the critical region or the post-critical region or even the post-post-critical region.

Let us see the conditions of backward anaphora:

(7) Condition 1 Null-Match

Enquanto/estava/virada/para o restaurante,/a Soraia/apagou/o cigarro/tranquilamente.
While/was/turned.fem.sg/to the restaurant,/the Soraia/put out/the cigarette/quietly.

(8) Condition 2 Overt-Match

Enquanto/ela estava/virada/para o restaurante,/a Soraia/apagou/o cigarro/tranquilamente.
While/she was/turned.fem.sg/to the restaurant,/the Soraia/put out/the cigarette/quietly.

(9) Condition 3 Null-Mismatch

Enquanto/estava/virado/para o restaurante,/a Soraia/apagou/o cigarro/tranquilamente.
While/was/turned.masc.sg/to the restaurant,/the Soraia/put out/the cigarette/quietly.

⁹⁶ So, in these conditions, the interpretation of the null or overt pronoun is only a preference by the participants, and not forced by grammar.

(10) Condition 4 Overt-Mismatch

Enquanto/ele estava/virado/para o restaurante,/a Soraia/apagou/o cigarro/tranquilamente.
While/he was/turned.masc.sg/to the restaurant,/the Soraia/put out/the cigarette/quietly.

‘While (he/she) was turned to the restaurant, Soraia put out the cigarette, quietly.’

As can be seen from the examples, each item consists of a matrix clause plus a temporal adverbial clause, which precedes the matrix clause. In each condition, the matrix clause contains a DP which may serve as a potential antecedent of the subordinate subject, thus yielding a backward anaphora situation.

Condition 1 contains a null embedded subject, where there is gender match between it and the matrix subject. Condition 2 contains an overt pronoun in the subordinate clause, where there is gender match between it and the matrix subject. Condition 3 contains a null embedded subject, where there is gender mismatch between it and the matrix subject. Condition 4 contains an overt pronoun in the subordinate clause, where there is also gender mismatch between it and the matrix subject. In this way, the presence or not of an overt pronoun and the gender match or mismatch correspond to the Pronoun Type and Match Type variables, respectively. In this way, there are two independent variables with two levels each: Pronoun Type (Null Pronoun vs. Overt Pronoun) and Match Type (Subject Match vs. Subject Mismatch).

Like the conditions of forward anaphora, all of the subordinate clauses in the backward anaphora conditions are also in the passive or adjectival form, in order to enable the forced interpretation. In all of the backward anaphora sentences, the matrix subject serves as the critical region, while the matrix verb serves as the post-critical region. The matrix object serves as the post-post-critical region. The length (number of characters and spaces) of the critical region, the post-critical region and the post-post-critical region were properly controlled.

It should be noticed that in the Null-Mismatch and Overt-Mismatch conditions, the participants are forced to accept the disjoint reading between the main and subordinate subjects, while in the Null-Match and Overt-Match conditions, both the coreferential and the disjoint reading are possible. The gender disambiguation process works in the same way as in the forward anaphora conditions.

As can be seen from the examples, in both the forward and the backward anaphora, there is only one potential antecedent in the sentence for the null or overt pronoun. Such a design serves

to establish similar conditions in comparison to the off-line test.

Both the forward anaphora and the backward anaphora conditions contain twenty experimental items, with the structure described above. Inside each condition (forward and backward anaphora), sentences were distributed by four lists crossing Pronoun Type and Match Type according to a Latin Square design: all participants saw all sentences in different conditions. However, items from the forward anaphora condition are not related to items from the backward anaphora condition, and hence, for example, a given item in forward anaphora will not have a counterpart in backward anaphora, unlike what happened in the off-line test. Forward and backward anaphora conditions are, therefore, considered as two different experiments. The option of including backward and forward anaphora conditions as separate experiments in the same experiment was due to time and limitations in access to participants. Since it was not easy to find L2 PE participants with B2 or C1 level, I decided to test both conditions in the same experiment, but treating them as separate experiments. Moreover, it was not possible or, at least, not straightforwardly possible to directly compare both structures (backward and forward conditions), that is, to compare regions with null and overt pronouns (in forward condition) and regions with full DPs (in backward conditions) – this was another reason why I decided to treat conditions as different experiments. The best option was to build two different experiments, but since it was not easy to find participants, as I explained above, I opted for this solution.

In addition to the experimental sentences, forty filler items were constructed and presented in between the experimental items, in a pseudo-randomized order: every experimental item was always followed by one filler item. All of the filler items were also segmented in the same way as the experimental items. The filler items do not vary between different lists and some filler items are exactly the same ones that appear in the off-line test.

Every item (experimental and filler) was followed by one comprehension question with two possible answers. For the experimental items, one of the options corresponds to a coreferential interpretation between the null or overt pronoun and the DP, and the other option corresponds to a disjoint reading. For the filler items, one option gives a true information about the sentence and the other conveys incorrect information about the sentence. Thus, for the filler items, there is only one possible (correct) answer. The options are exemplified as follows:

(11) For experimental items

Quando / a Carla / encomendou / um bolo, / ele ficou curioso / com o recheio / escolhido.

When/the Carla/ordered/one cake,/he became curious.masc.sg/with the filling/chosen.

‘When Carla ordered one cake, he became curious about the chosen filling.’

A. a Carla ficou curiosa com o recheio escolhido ‘Carla became curious about the chosen filling.’

B. uma outra pessoa ficou curiosa com o recheio escolhido ‘another person became curious about the chosen filling.’

(12) For filler items

A Marta/pediu/à Susana/uma opinião/sobre a decoração/da casa.

the Marta/asked/to+the Susana/one opinion/about the decoration/of+the home

‘Marta asked Susana an opinion about the decoration of the home.’

A. A Susana pediu uma opinião. ‘Susana asked for an opinion’ B. A Marta pediu uma opinião.

‘Marta asked for an opinion’

7.2.3 Procedure

The on-line test was presented on a laptop and completed in a silent room. All of the participants took part in the test individually, according to a pre-booked schedule. Participants were instructed to carefully read the instruction of the test, presented on a separate page, and then to fill in their personal information such as gender, age, knowledge of foreign languages and also information about EP proficiency, use and knowledge⁹⁷. Then they were asked to operate on the laptop the rest of the test.

The self-paced reading task was presented in Psychopy (Peirce 2007), in the Moving Window mode. All of the items were presented in *Lucida Console* font with a letter height of 0.9⁹⁸ at the center of a white screen. Before each sentence, a cross appeared on the left side of the screen. After the participant pressed the space bar, the sentence appeared with every segment masked by an underscore (white spaces between segments were preserved). After pressing the space bar again, the first segment appeared, while the other segments were still being masked. Afterwards, every time the participant pressed the space bar, the next segment appeared and the previous segment was re-masked by an underscore. After reading the full sentence, two options about the interpretation of the sentence appeared on the screen. The participant was asked to select one option by pressing ‘s’ (answer on the left side) or ‘l’ (answer on the right side) on the

⁹⁷ The introduction of the test, as well as the personal information questions, were presented in Portuguese. However, L2 learners were provided with a Chinese translation of the introduction and the informed consent.

⁹⁸ The letter width is then determined by the font.

keyboard. The same procedure was repeated for every item until the end of the test. Before the experimental test, the participants were given a training session to get used to the procedure. Two breaks were included during the reading of the testing sentences.

During the process of reading, the software recorded the reaction time of every segment read by the participants, as well as their answers for the options. As a result, for the experimental items, the reaction time of the critical region, post-critical region and post-post-critical region (only for backward anaphora) and the answers are the dependent variables.

7.3 Analysis

The statistical analysis for the current on-line test is composed of three parts: accuracy rate for filler items, answers for experimental items and reaction time for experimental items.

In the first part of the analysis, two L2 learners were found to have failed to reach the accuracy level of 70% in the filler items. These participants were excluded and so, as mentioned before, only the data from 27 participants was analyzed.

The second part of the analysis concerns the selection of antecedent for a referentially dependent element, which is similar to that of the off-line test. As a result, I also implemented an LMM with *lme4* package (Bates, Maechler, Bolker, & Walker, 2015) in *R* (R Development Core Team, 2008). I also fitted a mixed logit regression model using *glmer* with “logit” link function (Jaeger, 2008), since the dependent variable in the current analysis is binomial. The data for forward anaphora and backward anaphora were treated separately. In each condition (forward or backward anaphora), I included in the model all main effects and interactions. The following main effects were included: Pronoun (Null Pronoun vs. Overt Pronoun), Match (Subject Match vs. Subject Mismatch), Language (EP native speakers vs. L2 learners) and all interactions between these effects. I also included items and participants as random intercepts as well as by-item and by-participants slopes for all main effects. All independent variables were centered to avoid collinearity effects (Jaeger, 2008; Field, Miles, & Field, 2012).

For the third part of the analysis, the reaction time is the dependent variable. As a result, various LMMs⁹⁹ tests were manipulated. Like the analyses for the answers, the forward and backward

⁹⁹ I used the same statistical analysis in the part two, only without the logit function.

anaphora conditions were also treated separately in the analysis of reaction time. In each condition (forward or backward anaphora), the following main effects were included: Pronoun (Null Pronoun vs. Overt Pronoun), Match (Subject Match vs. Subject Mismatch), Language (EP native speakers vs. L2 learners). I also included items and participants as random intercepts as well as by-item and by-participants slopes for all main effects.

For the reaction times, I eliminated all values below 250ms and above 3000ms. After this, I eliminated all the values which are higher than the mean + 2.5 SD. After deleting outliers, reaction times were log transformed.

7.3.1 Off-line results: comprehension question

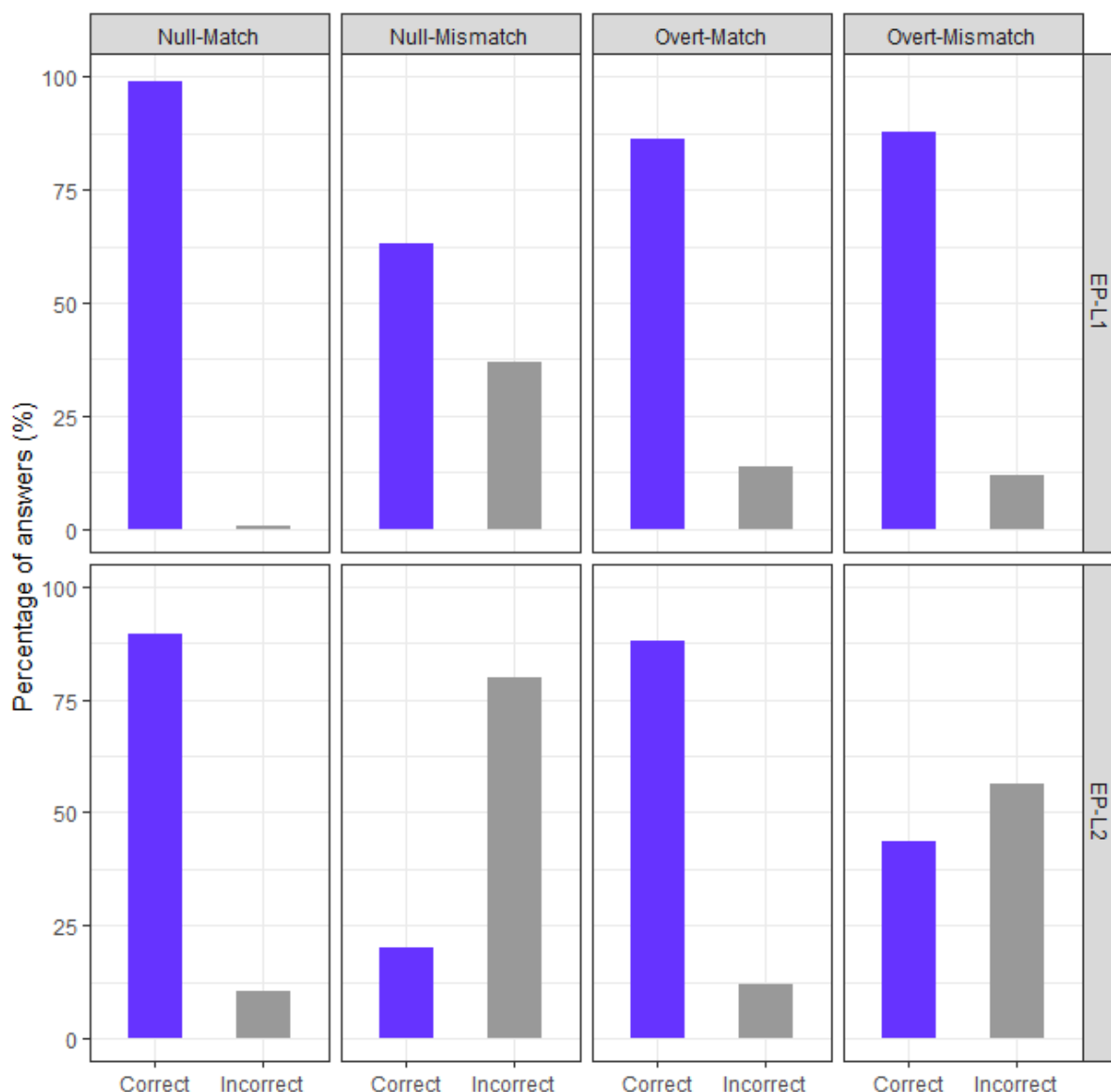
Before making the statistical analyses, I first codified the answers for the off-line questions. Each item was assigned a correct answer and an incorrect answer¹⁰⁰, which can be summarized in the following way. This codification is valid for both the forward and backward anaphora conditions.

- (13) Null-Mismatch: choosing subject antecedent = incorrect
choosing other antecedent = correct
- Null-Match: choosing subject antecedent = correct
choosing other antecedent = incorrect
- Overt-Mismatch: choosing subject antecedent = incorrect
choosing other antecedent = correct
- Overt-Match: choosing subject antecedent = correct
choosing other antecedent = incorrect

Firstly, let us consider the answers of the participants. The answers for the forward anaphora conditions are presented in Graphic 3, while statistical analyses are in Tables 17 to 19. It should be noticed that the native speakers clearly prefer the subject antecedent in Null-Match condition. However, in the Null-Mismatch condition, there are also some participants who choose the subject antecedent, which is obviously incompatible with the gender agreement (there is interaction between pronoun*match, cf. Table 17). For the conditions with overt pronouns, they chose the subject antecedent when the gender match allows such interpretation, and chose the

¹⁰⁰ It should be noticed that in some conditions, especially those conditions with gender match, the ‘incorrect answer’ does not indicate that it is incorrect in interpretation, but only serves as a codification for analyses. Since the off-line answer is only a secondary test of the self-paced reading task, I chose to keep this codification.

non-subject antecedent when there is gender mismatch.



Graphic 3 Results of answers in forward anaphora

As for the L2 participants in these conditions, they also prefer the subject antecedent in the Null-Match condition. However, in the Null-Mismatch condition, they still chose subject antecedent, which goes against the gender match (there is match effect and interaction between match*pronoun cf. Table 18). In the Overt-Match condition, like the native speakers, the L2 participants also prefer subject antecedents, but for the Overt-Mismatch condition, more than half of the answers correspond to the subject antecedent, which again goes against the gender

agreement.

	Estimate	Standard Error	z-value	p-value
(Intercept)	5.150	1.313	3.924	<0.001
Pronoun	-0.785	1.487	-0.528	0.598
Match	2.330	1.594	1.461	0.144
Pronoun * Match	-14.880	5.229	-2.846	0.004

Table 17 Statistical analysis for native speakers in forward anaphora - answers

	Estimate	Standard Error	z-value	p-value
(Intercept)	0.858	0.152	5.659	<0.001
Pronoun	0.919	0.450	2.044	0.041
Match	4.190	0.746	5.618	<0.001
Pronoun * Match	-1.724	0.611	-2.819	0.005

Table 18 Statistical analysis for L2 learners in forward anaphora - answers

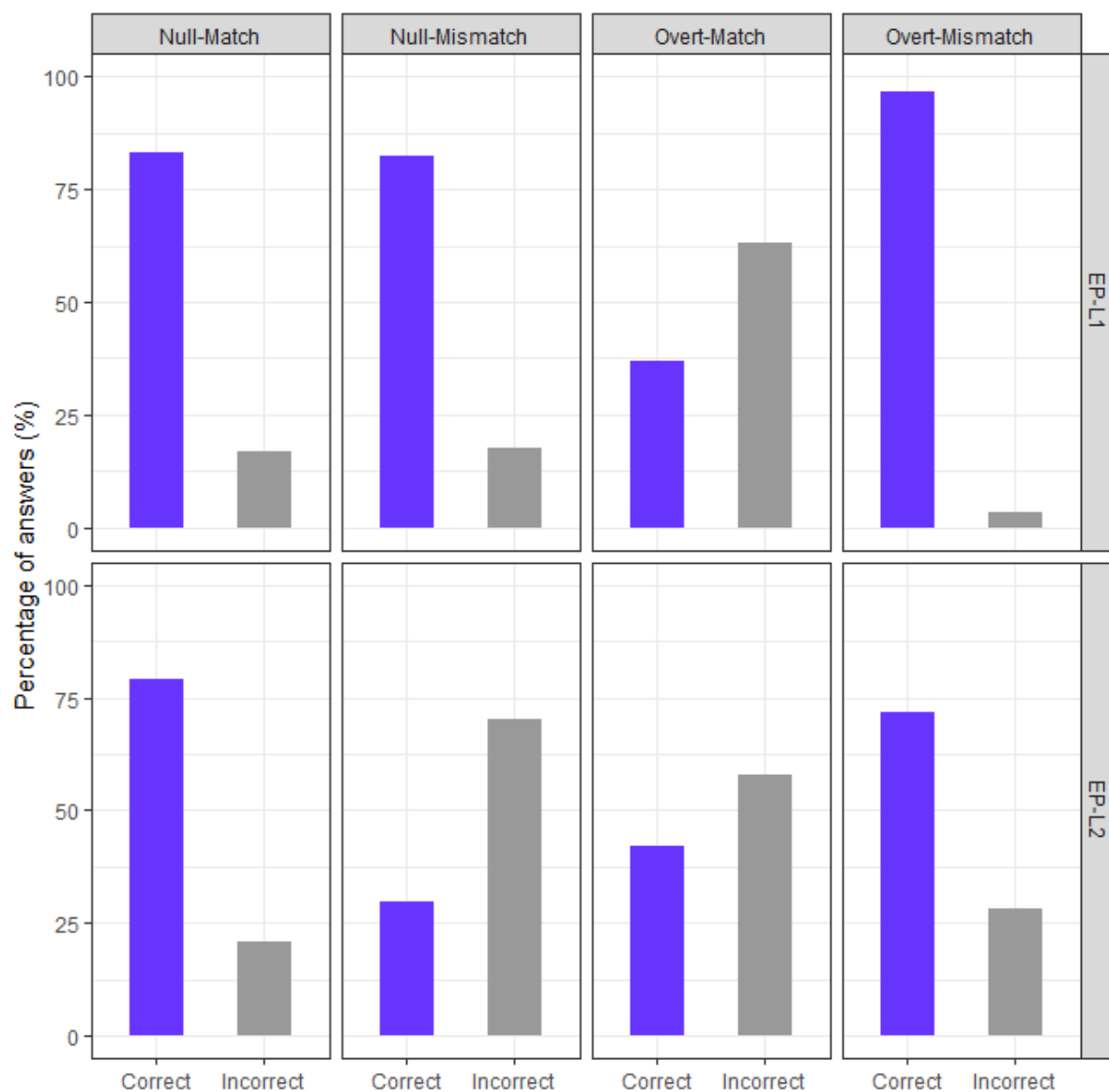
	Estimate	Standard Error	z-value	p-value
(Intercept)	1.862	0.251	7.431	<0.001
Pronoun	0.429	0.451	0.950	0.342
Match	2.850	0.634	4.498	<0.001
Language	1.218	0.225	5.424	<0.001
Pronoun * Match	-3.987	0.892	-4.472	<0.001
Pronoun * Language	-0.522	0.395	-1.320	0.187
Match * Language	-1.202	0.540	-2.228	0.026
Pronoun * Match * Language	-2.088	0.772	-2.703	0.007

Table 19 Comparison between native speakers and L2 learners in forward anaphora - answers

Comparing the two groups (see statistical analyses in Table 19), there is interaction between match*language, meaning that the L2 learners are more likely to choose the incorrect answer (with gender mismatch), and this phenomenon is more relevant in the Null-Mismatch condition, with interaction between pronoun*match*language.

Now let us consider the results for the backward anaphora conditions (Graphic 4); the statistical analyses are presented in Tables 20 to 22. As for the native speakers, they clearly respect gender agreement in the null subject conditions. In the overt pronoun conditions, they accept the non-subject antecedent when the gender match does not favor the subject antecedent. Even in the cases of Overt-Match, when the gender information permits the subject antecedent, they still

showed a higher preference for the non-subject antecedent. There is a match effect and interaction between match*pronoun (cf. Table 20).



Graphic 4 Results of answers in backward anaphora

As for the L2 participants, they show a preference for the subject antecedent in the null subject conditions, irrespective of whether there is gender match. For the overt pronoun conditions, they choose the non-subject antecedent when the gender information does not permit the subject antecedent, though some of the answers still correspond to the subject antecedent. As for the Overt-Match condition, when the subject antecedent is possible, they also showed a higher

preference for the non-subject antecedent, which is also revealed by the interaction between pronoun*match (cf. Table 21).

	Estimate	Standard Error	z-value	p-value
(Intercept)	2.063	0.393	5.246	<0.001
Pronoun	-0.463	0.812	-0.570	0.569
Match	-2.466	0.915	-2.694	0.007
Pronoun * Match	-5.437	1.604	-3.390	0.001

Table 20 Statistical analysis for native speakers in backward anaphora - answers

	Estimate	Standard Error	z-value	p-value
(Intercept)	0.419	0.153	2.736	0.006
Pronoun	0.130	0.351	0.369	0.712
Match	0.486	0.481	1.010	0.312
Pronoun * Match	-4.606	0.805	-5.721	<0.001

Table 21 Statistical analysis for L2 learners in backward anaphora - answers

	Estimate	Standard Error	z-value	p-value
(Intercept)	1.138	0.138	8.249	<0.001
Pronoun	-0.058	0.314	-0.183	0.854
Match	-0.905	0.404	-2.238	0.025
Language	0.769	0.119	6.743	<0.001
Pronoun * Match	-4.924	0.642	-7.664	<0.001
Pronoun * Language	-0.225	0.262	-0.858	0.391
Match * Language	-1.485	0.382	-3.890	<0.001
Pronoun * Match * Language	-0.328	0.589	-0.556	0.578

Table 22 Comparison between native speakers and L2 learners in backward anaphora - answers

The comparison between the two groups can be shown by the statistical analyses in Table 22. There is effect of Language, meaning that the L2 learners had more incorrect answers than the native speakers, and this phenomenon is more relevant in the mismatch conditions, with an interaction between language*match. There is also an interaction between pronoun*match, with more correct answers in the null condition in the EP native group.

7.3.2 Discussion

Let us first consider the conditions of forward anaphora. It is interesting that both the native speakers and the L2 learners may ignore the gender information in the Null-Mismatch condition, and the L2 learners even choose more incorrect answers than correct ones. Such an interpretation may suggest that participants from both groups show a strong preference for the subject antecedent in the null subject conditions, which is predicted by the PAH¹⁰¹.

However, for the Null-Match condition, both groups show a significant accuracy, since they hardly choose the non-subject antecedent. This fact also reinforces the idea that they strongly prefer the subject antecedent for the null subject.

Now we turn to the overt pronoun conditions. When there is gender mismatch, the accuracy of the native speakers is relatively higher than that of the L2 learners. It is possible that the L2 learners have a strong preference for the subject antecedent, which leads them to ignore the gender information.

For the Overt-Match condition, when there is gender match, both the native speakers and the L2 learners prefer the subject antecedent. It should be noticed that such an interpretation by the L2 learners corroborates the results of the off-line test and the interpretation in their L1 (which is Chinese Mandarin). However, for the native speakers, such an interpretation is different from that found in the off-line test, where the native speakers prefer the non-subject antecedent. It should be noticed that there is only one potential antecedent in the sentence for the overt pronoun. For example, Carminati 2002 also found the same interpretation when there is only one antecedent for overt pronoun. This may suggest that some native speakers were eager to choose an antecedent in the sentence, rather than searching for an antecedent outside the context, which increases the processing cost, especially in an on-line test. The same may also be true for the L2 learners, who were reported to have difficulty in retrieving the discursive information in their L2 processing (Sorace & Filiaci 2006; Zhao 2012). Another explanation could be L1 influence, since in Chinese, the native speakers clearly prefer the subject antecedent in the equivalent condition.

¹⁰¹ It is also possible that gender sensitivity accounts for such an interpretation, since, in the null subject conditions, the gender information is only revealed in the past participle, which may lead participants to disregard this information, especially in the case of the L2 learners.

Now we turn to the conditions of backward anaphora. As for the native speakers, unlike in the forward anaphora condition, they show a greater accuracy in the Null-Mismatch condition, where there is no gender match, and in the Null-Match condition, they clearly prefer the subject antecedent, which corresponds to the results of the off-line test. For conditions with overt pronoun, they prefer the non-subject antecedent when the gender information permits the subject antecedent (Overt-Match condition) and show high accuracy when the gender agreement does not permit the subject antecedent (Overt-Mismatch condition). These facts show that the native speakers prefer the non-subject antecedent in backward anaphora structures with overt embedded pronoun, which also corroborates the results of the off-line test.

For the L2 learners, they behave in the same way as the native speakers, except in the Null-Mismatch condition. In this situation, the L2 learners choose the subject antecedent, which is incompatible with the gender information. This phenomenon may also be explained by the fact that they may not be sensitive to the gender information contained in the past participle. Furthermore, since the past participle precedes the matrix clause in this condition, its processing cost also increases as a result¹⁰². However, the L2 learners prefer the subject antecedent whenever such an interpretation is allowed by the grammar (Null-Match condition), which is again consistent with the results of the off-line test.

For the conditions with overt pronoun, the L2 learners show greater accuracy in the Overt-Mismatch condition, where the subject antecedent is not allowed by the grammar. However, such accuracy by the L2 learners has not been found in the forward anaphora condition, which may suggest that their choice may also be influenced by the structure, that is to say, they indeed prefer the non-subject antecedent in backward anaphora, but not in forward anaphora. This hypothesis is also strengthened by the fact that, in the Overt-Match condition, the L2 learners still prefer the non-subject antecedent, which goes against the results in the corresponding condition in forward anaphora. Interestingly, in the off-line test, such a difference between forward and backward anaphora was also attested for the L2 learners, which suggests L1 influence, since, in Chinese, native speakers indeed show such an interpretative difference for the overt pronoun between forward and backward anaphora.

¹⁰² It is also possible that they indeed prefer the subject antecedent, which would lead them to ignore the gender information.

Generally speaking, the results of the answers in this on-line test corroborate those in the off-line test of Experiment 1, except in the case of the native speakers in the Overt-Match condition of forward anaphora. This phenomenon may be influenced by the difference between Experiment 1 and Experiment 2, since in the latter some conditions have a forced interpretation, while in the former there is no forced interpretation. For the null subject condition in forward anaphora, both the native speakers and the L2 learners prefer the subject antecedent, while they show the same preference in the overt pronoun condition. In backward anaphora, both groups prefer the subject antecedent in the null subject condition and the non-subject antecedent in the overt pronoun condition.

7.3.3 Reaction times: forward anaphora

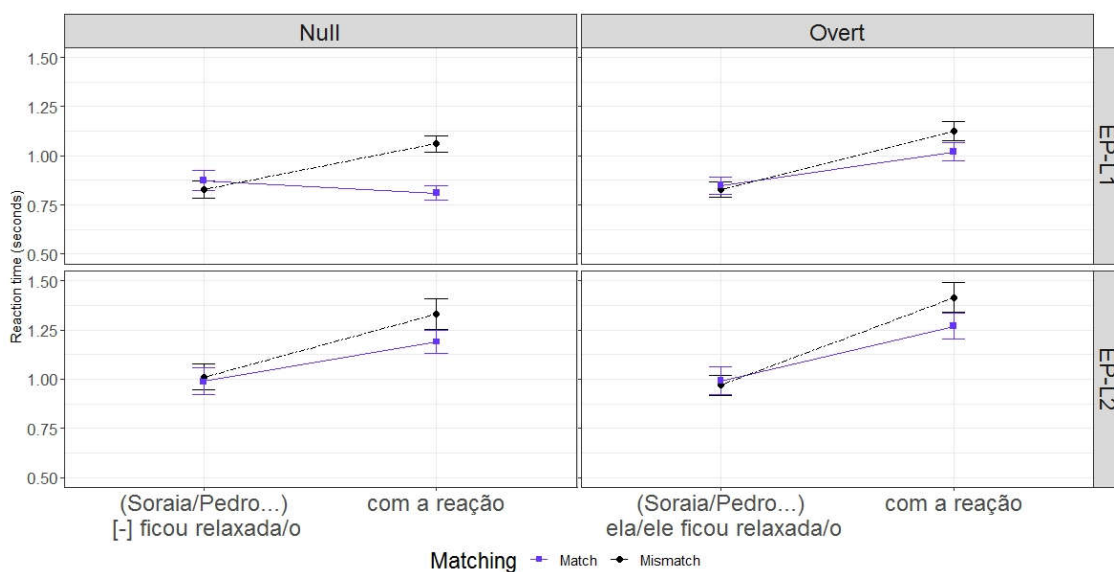
7.3.3.1 Critical region

Now we consider the reaction times for all the conditions. We first check the reaction times for the forward anaphora conditions. Graphic 5 shows the reaction time for all the regions in the forward anaphora conditions, while Table 23-25 show the statistical analysis for the critical region. It has been found that there is no effect inside the control group and the L2 group, and the only effect is the difference between the two groups, without any interaction effects (cf. Table 25, only a marginal language effect). These results show that there is no difference in reaction times between conditions, inside both groups, and the only difference is that the L2 learners used more time to read than the native speakers, in all of the four conditions¹⁰³.

	Estimate	Standard Error	t-value	p-value
(Intercept)	-0.239	0.042	-5.709	<0.001
Pronoun	0.002	0.031	0.053	0.958
Match	0.026	0.029	0.906	0.365
Pronoun * Match	-0.027	0.056	-0.485	0.627

Table 23 Statistical analysis for native speakers in forward anaphora – critical region

¹⁰³ I will not consider this effect as relevant, since it reflects level of proficiency and it is expected that non-native speakers take longer to read and decode information than native speakers.



Graphic 5 Reaction time in forward anaphora – null and overt pronoun

	Estimate	Standard Error	t-value	p-value
(Intercept)	-0.120	0.056	-2.142	0.032
Pronoun	-0.011	0.039	-0.291	0.771
Match	-0.012	0.042	-0.290	0.772
Pronoun * Match	-0.019	0.077	0.251	0.802

Table 24 Statistical analysis for L2 learners in forward anaphora – critical region

	Estimate	Standard Error	t-value	p-value
(Intercept)	-0.178	0.036	-4.975	<0.001
Pronoun	-0.007	0.024	-0.285	0.776
Match	0.007	0.022	0.307	0.759
Language	0.119	0.066	1.786	0.074
Pronoun * Match	-0.002	0.049	-0.034	0.973
Pronoun * Language	-0.015	0.045	-0.339	0.735
Match * Language	-0.036	0.044	-0.825	0.410
Pronoun * Match * Language	0.049	0.098	0.494	0.621

Table 25 Comparison between native speakers and L2 learners in forward anaphora – critical region

7.3.3.2 Post-Critical region

Now we turn to the post-critical region for the forward anaphora conditions. Graphic 5 show the reaction time in this respect, while Table 26-28 show the statistical analysis. It is clear that in both groups the reaction time is faster in the Match conditions than in Mismatch conditions,

which means that the participants from both groups read quicker when there is gender match between the embedded subject and the matrix subject. As can be seen in Table 26, there is an effect of pronoun, with more time in overt pronoun condition, and effect of match, with more time in Mismatch, for the control group, as well as a marginal interaction effect concerning pronoun * match, with less time in Null-Match. These results show that the native speakers read quicker in the null subject conditions and in the conditions with gender match of subject.

The L2 learners (cf. Table 27) also show a significant match effect and a marginal pronoun effect, meaning that they also read quicker in the null subject conditions and in the conditions with gender match. However, there is no interaction between the two factors.

	Estimate	Standard Error	t-value	p-value
(Intercept)	-0.069	0.033	-2.060	0.039
Pronoun	0.140	0.034	4.166	< 0.001
Match	-0.169	0.032	-5.317	< 0.001
Pronoun * Match	0.141	0.076	1.869	0.062

Table 26 Statistical analysis for native speakers in forward anaphora – post critical region

	Estimate	Standard Error	t-value	p-value
(Intercept)	0.178	0.044	4.061	<0.001
Pronoun	0.072	0.038	1.913	0.056
Match	-0.096	0.038	-2.506	0.012
Pronoun * Match	0.027	0.075	0.359	0.720

Table 27 Statistical analysis for L2 speakers in forward anaphora – post critical region

	Estimate	Standard Error	t-value	p-value
(Intercept)	0.055	0.028	1.993	0.046
Pronoun	0.106	0.025	4.265	< 0.001
Match	-0.132	0.028	-4.760	< 0.001
Language	0.248	0.052	4.774	< 0.001
Pronoun * Match	0.080	0.043	1.874	0.061
Pronoun * Language	-0.069	0.048	-1.423	0.155
Match * Language	0.075	0.045	1.680	0.093
Pronoun * Match * Language	-0.118	0.086	-1.381	0.167

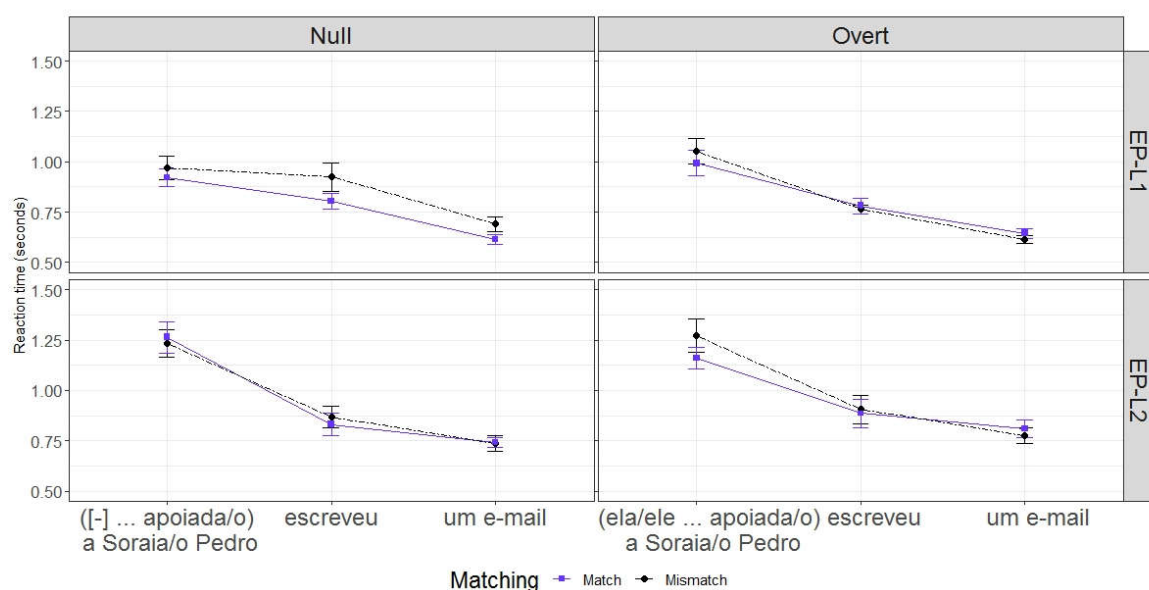
Table 28 Comparison between native speakers and L2 learners in forward anaphora – post critical region

Now consider the comparison between the groups (cf. Table 28): there is an effect of pronoun, match and language. The first two effects indicate that participants from both groups read faster in null subject conditions and in conditions with gender match. The last effect also indicates that the L2 learners read slower than the native speakers.

7.3.4 Reaction times: backward anaphora

7.3.4.1 Critical region

Now we consider the reaction time for the backward anaphora conditions. As in the forward anaphora conditions, there is no significant effect in the critical region. Reaction times are shown in Graphic 6, while the statistical analyses are in Tables 29-31. There is only marginal pronoun effect for the native speakers (cf. Table 29), meaning that they read slower in overt pronoun conditions, and language effect when comparing the two groups (cf. Table 31), which again shows that the native speakers spent less time reading than the L2 learners.



Graphic 6 Reaction time in backward anaphora – null and overt pronoun

	Estimate	Standard Error	t-value	p-value
(Intercept)	-0.101	0.056	-1.813	0.070
Pronoun	0.050	0.028	1.749	0.080
Match	-0.039	0.028	-1.390	0.165
Pronoun * Match	-0.014	0.055	-0.260	0.795

Table 29 Statistical analysis for native speakers in backward anaphora – critical region

	Estimate	Standard Error	t-value	p-value
(Intercept)	0.120	0.060	2.012	0.044
Pronoun	-0.011	0.038	-0.276	0.782
Match	-0.018	0.036	-0.498	0.618
Pronoun * Match	-0.109	0.079	-1.375	0.169

Table 30 Statistical analysis for L2 learners in backward anaphora – critical region

	Estimate	Standard Error	t-value	p-value
(Intercept)	0.012	0.046	0.251	0.802
Pronoun	0.019	0.023	0.819	0.413
Match	-0.029	0.021	-1.395	0.163
Language	0.220	0.068	3.232	0.001
Pronoun * Match	-0.060	0.048	-1.232	0.218
Pronoun * Language	-0.062	0.043	-1.449	0.147
Match * Language	0.020	0.041	0.497	0.619
Pronoun * Match * Language	-0.092	0.097	-0.944	0.345

Table 31 Comparison between native speakers and L2 learners in backward anaphora – critical region

7.3.4.2 Post-Critical region

Coming into the post-critical region (see Graphic 6), according to the statistical analysis (Table 32-34), there is significant effect of pronoun for the native speakers (Cf. Table 32), which indicates that they spent more time reading sentences with null embedded subject. This result seems to be mainly influenced by the time spent reading the Null-Mismatch condition, though there is no significant interaction between pronoun*match.

	Estimate	Standard Error	t-value	p-value
(Intercept)	-0.273	0.041	-6.699	<0.001
Pronoun	-0.077	0.038	-2.029	0.042
Match	-0.044	0.039	-1.134	0.257
Pronoun * Match	0.106	0.076	1.397	0.162

Table 32 Statistical analysis for native speakers in backward anaphora – post critical region

However, for the L2 learners, there is again absence of any statistical effects. When comparing the two groups, there is interaction between pronoun and language (cf. Table 34), which means

that the native speakers spent more time reading sentences with null subject conditions.

	Estimate	Standard Error	t-value	p-value
(Intercept)	-0.241	0.055	-4.339	<0.001
Pronoun	0.031	0.035	0.886	0.376
Match	-0.026	0.035	-0.751	0.453
Pronoun * Match	0.010	0.058	0.171	0.864

Table 33 Statistical analysis for L2 learners in backward anaphora – post critical region

	Estimate	Standard Error	t-value	p-value
(Intercept)	-0.257	0.036	-7.221	<0.001
Pronoun	-0.021	0.024	-0.889	0.374
Match	-0.034	0.027	-1.261	0.207
Language	0.032	0.068	0.471	0.638
Pronoun * Match	0.058	0.042	1.373	0.170
Pronoun * Language	0.107	0.047	2.289	0.022
Match * Language	0.019	0.043	0.435	0.664
Pronoun * Match * Language	-0.095	0.084	-1.138	0.255

Table 34 Comparison between native speakers and L2 learners in backward anaphora – post critical region

7.3.4.3 Post- Post-Critical region

Finally, let us consider the reaction time for the post-post-critical region. As in the post-critical region, the native speakers took longer time with sentences from the Null-Mismatch condition (see Graphic 6), which was supported by the statistical analysis (Table 35-37), showing that there is an interaction between pronoun and match for the control group, (cf. Table 35). This interaction effect seems to show a spillover effect, since, in the post-critical region, although there is an apparent difference between reaction time of Null-Mismatch and other conditions, no statistical effect was verified.

There is still no significant effect from the L2 group. When comparing the two groups, there is an effect of language (cf. Table 37), which indicates again that the L2 group spent more time reading than the native speakers. There is also a marginal interaction between pronoun, match and language (cf. Table 37), which suggests that it was the native speakers who spent more time reading sentences from the Null-Mismatch condition, when gender mismatch does not allow

the coreferential reading between the embedded null subject and the matrix subject.

	Estimate	Standard Error	t-value	p-value
(Intercept)	0.640	0.022	29.584	<0.001
Pronoun	-0.022	0.025	-0.890	0.373
Match	-0.020	0.024	-0.807	0.420
Pronoun * Match	0.105	0.045	2.333	0.020

Table 35 Statistical analysis for native speakers in backward anaphora – post post critical region

	Estimate	Standard Error	t-value	p-value
(Intercept)	-0.345	0.038	-9.092	<0.001
Pronoun	0.046	0.035	1.335	0.182
Match	0.024	0.029	0.816	0.414
Pronoun * Match	-0.014	0.068	-0.210	0.834

Table 36 Statistical analysis for L2 learners in backward anaphora – post post critical region

	Estimate	Standard Error	t-value	p-value
(Intercept)	-0.417	0.026	-16.200	<0.001
Pronoun	0.014	0.025	0.542	0.588
Match	0.003	0.021	0.126	0.900
Language	0.150	0.048	3.145	0.002
Pronoun * Match	0.057	0.045	1.279	0.201
Pronoun * Language	0.070	0.043	1.642	0.101
Match * Language	0.044	0.039	1.126	0.260
Pronoun * Match * Language	-0.145	0.083	-1.745	0.081

Table 37 Comparison between native speakers and L2 learners in backward anaphora – post post critical region

7.3.5 Discussion: Reaction times

7.3.5.1 Forward anaphora

Now we discuss the reaction time results previously presented. We first check the forward anaphora condition. As has been shown in the previous section, there is no statistically significant effect in the critical region, except for language between the two groups, which means that it took longer for the L2 learners to read the sentences than for the native speakers. Such an effect is not relevant for our analysis since it is natural that L2 speakers spend more

time reading than native speakers. A similar phenomenon has been attested in almost all the other regions and conditions. For the relevance of information, I will not discuss the language effect alone in the following conditions.

There are significant effects in the post-critical region, for both the control group and the L2 group. For the native speakers, the effect of match and pronoun and the interaction of pronoun and match show that they spent less time with the sentences with gender match and this is more relevant in the condition with null subject. This reaction time difference for the null subject condition indicates that the native speakers prefer the coreference between the sentence-initial subject and the matrix null subject, and thus spent less time reading when there was gender match between them. This preference is consistent with what PAH predicts for null subjects and the results from Experiment 1.

However, it should be noticed that, in the overt pronoun condition, the native speakers also displayed a similar reaction time difference, though this may not be as relevant as in the null subject condition. This indicates that the native speakers also accept the subject antecedent in this condition, which is also corroborated by their answers (see Graphic 3). It is obvious that such an interpretation is different from what the PAH predicts and the results of the off-line test. There are three possible explanations for this result. First, there may be some difference between a pure off-line test and an on-line test. As a result, when the procedure is an on-line test where there are time limitations, the participants are more likely to accept a coreferential interpretation, which does not involve any information from outside the context. Second, the sentences in this on-line test are not exactly the same as those in the off-line test, since the sentence-initial subject is contained in a subordinate clause. Moreover, as Carminati 2002 reported, when there is only one potential antecedent in the sentence, the participant may have a higher preference for a subject antecedent for the overt pronoun, compared to the sentences with two potential antecedents.

Now we consider the L2 learners, who show match effect in this condition, without interaction between pronoun and match, which indicates that they spent less time when reading sentences with gender match between the sentence-initial subject and the matrix subject, no matter whether it is the null subject or the overt pronoun condition. As can be seen in Graphic 5, they spent less time in the Null-Match condition than in the Null-Mismatch condition, and less time

in the Overt-Match condition than in the Overt-Mismatch condition. Such reaction time differences indicate that the L2 learners prefer the subject antecedent, irrespective of type of pronoun (overt or null). This interpretation is consistent with what was found in the off-line test, and also corresponds to the interpretation in their L1 Chinese.

Here, it should be noticed that there is in fact no difference between the null and the overt pronoun conditions. Though there is a marginal pronoun effect, meaning higher reaction time for the overt pronoun condition, the lack of interaction between pronoun and match indicates that the differences in reaction time between match and mismatch conditions inside the null and pronoun conditions are the same. In the off-line test, I proposed that such a phenomenon can be explained by the fact that their L1 Chinese does not show such a difference in the interpretation of null and overt pronouns in this condition, which may influence their L2 interpretation. The on-line test results seem to corroborate this hypothesis, since, according to the reaction time difference, the L2 learners show equal preference for the coreferential reading. However, it is worth noting that the native speakers also show a similar preference in the case of the overt pronoun, which may indicate that the coreferential reading between the sentence initial-subject and a matrix overt pronoun can be a universal strategy for the structure concerned.

Here, I would argue that there is a difference between the native speakers and the L2 learners. For the native speakers, there is an interaction effect which indicates that their preference for the subject antecedent is stronger in the null subject condition than in the overt pronoun condition. This fact may imply that the interpretative biases found in the null subject condition and in the overt pronoun condition may be caused by different reasons. However, for L2 learners, there is no such interaction, so it is not possible to consider that their preference for the subject antecedent is stronger in one condition than in the other. Due to this fact, I still propose that there is a difference between the L2 learners and the native speakers.

7.3.5.2 Backward anaphora

Now we analyze the reaction time for the backward anaphora conditions. Concerning the native speakers, it has been found that they again did not show any relevant effect in the critical region, except a marginal pronoun effect, meaning they spent less time reading sentences with null embedded pronoun, especially in the Null-Match condition. The L2 learners do not show any

significant effect in this region.

Now we consider the post-critical region. For the native speakers, there is only pronoun effect without interaction between pronoun and match, meaning they spent more time in the null subject conditions, irrespective of whether there is gender match or not. But, from Graphic 8, it is clear that they spent longer in the Null-Mismatch condition, which can be deduced as they accept the least the sentences which do not allow the coreferential reading between the backward anaphoric null subject and the matrix subject. This result in fact corroborates their responses (see Graphic 4) and the results of the off-line test, since the native speakers show a clear preference for the coreferential reading in the null subject conditions of backward anaphora.

However, there is still no time difference in the condition of overt pronoun, although according to their responses, the native speakers prefer the non-subject antecedent in the Overt-Match condition, when both the subject and the non-subject antecedent are available.

However, the reading times of the native speakers in this test do not seem to follow such logic. This puzzle may be explained in the following way: since the coreferential reading is not the preferred one, it is clear that the unmatched gender information will not increase the reading time. But, at the same time, the matched gender information does not necessarily increase the reading cost, since if the parser does not prefer the coreferential reading between the overt pronoun and the matrix subject, then it does not matter if the pronoun matches or not the matrix subject in gender. So, it is completely possible to result in a similar reading time between the gender matched and unmatched conditions.

A similar situation was tested also in Kazanina et al. 2007 and Kazanina & Philips 2010. For the case of Principle C of the Binding Theory, since it is not possible to establish an anaphoric relation between a sentence initial pronoun and a c-commanded DP, no reaction time differences were attested between sentences with gender match and those without gender match.

Now we quickly analyze the case of Post-post-critical region. There is an interaction between pronoun and match in the control group, which indicates that the native speakers spent more time reading sentences in Null-Mismatch condition, showing that they prefer the coreferential reading in the null subject condition. There is also no difference for the conditions of overt pronouns. As can be seen, the results of the native speakers in the post-post-critical region are

quite similar to those in the post-critical region. As for the L2 learners, there is still no significant effect, which is the same as in the post-critical region. A marginal interaction effect between pronoun, match and language indicates that only the native speakers, but not the L2 learners, show a significant preference for the coreferential interpretation in the null subject condition.

There are two points which seem to be relevant. First, it is unclear why the native speakers in the Null-Match and Null-Mismatch conditions of backward anaphora do not show an effect in the critical region, but rather in the post-critical region. Unlike the sentences of forward anaphora where the gender information is only revealed at the final part of the critical region, in backward anaphora, such information can be acquired right at the start of the region, where a definite article clearly indicates the gender of the DP. One explanation may be as follows: since the native speakers read the sentences so quickly, it is possible that some effect only occurs in a later region, resulting in the spillover effect, especially when the critical region is relatively short, only containing a DP.

The second question is, why do the L2 learners show an effect in the null subject conditions in forward anaphora, but not in backward anaphora? It is clear that, in both the off-line test and the responses of the current test, the L2 learners prefer the subject antecedent for null subject in both the forward and backward anaphora conditions. However, there is only statistical effect in the forward anaphora condition in this respect, while in backward anaphora, they almost spent equal time reading sentences with and without gender match in the null subject conditions of backward anaphora. One explanation should be as follows: in backward anaphora with null embedded subject, the gender information is only contained in the past participle which is located at the start of the sentence. Since the L2 learners might not be so sensitive to the gender information as the native speakers are, they may be unaware of such gender information when reading the first clause of the backward anaphora sentences. As a result, when they read the critical or post critical regions, such gender information may have been lost or become unclear in their memory, which results in the absence of effect. This explanation may be further strengthened by the low accuracy rate produced by the L2 learners in the Null-Mismatch condition of the off-line answer task. For the conditions of overt pronoun in backward anaphora, there is also a pronoun in the first clause of the sentence, which may reinforce the gender

information in the memory of the L2 learners. Thus, they did show some effect in this respect. Now we consider the case of forward anaphora, in which the past participle appears in the critical region where the participants were forced to establish an anaphoric relation, in order to complete the processing of the sentence. In this case, even if the L2 learners are not sensitive to the gender information contained in the past participle, they are still forced to pay attention to such gender information, because they are required to decide if there is an anaphoric relation between the null (or overt) pronoun and the previous subject DP. This explains why there is such difference between forward and backward anaphora for the L2 learners.

7.4 Summary of the on-line task

Let us now consider the overall results of the on-line test. For the native speakers, there are differences from the off-line test in Chapter 6, since they prefer the subject antecedent both for the null subject and the overt pronoun, in the case of forward anaphora, whereas in the offline test they prefer subject antecedent for null subject and non-subject antecedent for overt pronoun. Some differences between the design of the on-line and off-line tests may account for such results. However, even with the current results, the native speakers still show a stronger preference for the subject antecedent in the condition of null subject than in the condition of overt pronoun. For the case of backward anaphora, it can be confirmed that the native speakers behave alike in the on-line and off-line tests in null subject conditions, since they prefer the subject antecedent in this situation. However, there is no conclusive results to show their preference for overt pronoun condition in backward anaphora from their reaction times. Of course, their answers for the comprehension questions show that they prefer the non-subject antecedent in this condition.

For the L2 learners, they show a similar performance to the off-line test in the conditions of forward anaphora, since they always prefer the subject antecedent, irrespective of what kind of pronoun it is. Such an interpretative pattern is also consistent with their L1, which implies L1 influence in this respect, just as what the off-line test shows. However, the most unexpected results were found in the backward anaphora conditions. The L2 learners did not show a match effect in the condition of null and overt subject, which makes it impossible to determine their preference or make a comparison with the off-line test and their L1. Again, the L2 participants

may have some difficulties in processing backward anaphora, which results in the lack of effects. Since there are no significant results found in the backward anaphora conditions for L2 participants, the research question regarding this part ((2) of this Chapter, which is repeated as (14) below) cannot be resolved from the data of the current on-line task.

Finally, I repeat the two research questions of the current experiment as (13) and (14):

(13) Will the L2 learners show the same interpretative preference of the off-line test also in the on-line test, that is to say, will they treat forward and backward anaphora differently in the on-line test?

(14) Will the L2 be influenced by the active search mechanism? That is to say, is their interpretative preference in the off-line test only a late decision?

In response to these two questions, the answers should be as follows: the L2 learners use the same strategy of the off-line test to interpret forward anaphora structure in on-line test. However, the results found in the current test are not able to explain their processing in backward anaphora.

Chapter 8 Conclusions

The current thesis uses a theoretical approach and an experimental approach to investigate the anaphoric resolution of EP and Chinese and compares the interpretation of forward and backward anaphora, as well as the acquisition of the related properties of L2 EP by learners who speak Chinese as L1.

The first objective of the current thesis is to explore the syntactic and discourse-pragmatic properties of null subjects and anaphoric resolution under a theoretical approach. In both the GB and the MP frameworks the null subjects in consistent null subject languages are considered to be legitimated by rich verbal agreement. Under the recent proposal of Holmberg 2010, the key factor that allows the consistent null subject languages to have referential null subjects is the Definite feature in T. These facts imply that in null subject languages such as EP, since the phi-feature of a null pronoun can be recovered from the verbal agreement, the overt pronoun is in fact a redundant form, unless it conveys some additional information. This results in a division of labor between null and overt pronouns in EP and in other consistent null subject languages, which can be summarized as follows: null subjects are predominantly used in topic maintenance contexts, while overt pronouns are used in topic shift contexts. Such a division of labor can also be explained by the PAH, which predicts that the interpretation of null and overt pronouns is closely related to the structural position (which is closely related to the syntactic position) of the antecedent, though other factors such as order of reference (Luegi 2012) may also influence the pronominal resolution in the case of forward anaphora. As a result, it is assumed that the pronominal resolution in EP is constrained by both syntactic principles (since it cannot violate the Binding Principles) and discourse-pragmatic factors (PAH, among others). On the other hand, Chinese, a discourse null subject language, does not show a division of labor between null and overt pronouns, since both can refer to a subject antecedent in forward anaphora, which is different from EP, where only null subjects favor a subject antecedent. So, it is reasonable to consider why there is such a difference between EP and Chinese. It should be noticed that Chinese does not have verbal agreement at all, which means that an overt pronoun may not be redundant, since the phi-features of a null pronoun cannot be recovered from verbal agreement. As a result, Chinese overt and null pronouns can be used alternatively

in forward anaphora.

If this is on the right track, then the following questions must be answered:

- (1) If there is no verbal agreement in Chinese, how is the null subject legitimated by syntax?
- (2) Why do native speakers of Chinese use the null and overt pronouns alternatively in forward anaphora, instead of always using one form?
- (3) Furthermore, in backward anaphora in Chinese, there is a division of labor between null and overt pronouns. Why is there such an asymmetry between forward and backward anaphora?

For Question (1), the current thesis compared some previous studies on Chinese null subjects and A'-dependency. In the GB framework, Huang 1984 proposed that the Chinese null subjects can be a *pro* legitimated by GCR or a variable generated by A'-movement. In the MP framework, Liu 2014 proposed that the Chinese null subject is a *pro* which agrees with a topic. In the current thesis I propose that Chinese null subjects can be legitimated by different mechanisms according to the context where they appear. I assume that null subjects in subordinate clauses (including the case of island) are *pro*, legitimated and recovered by GCR, and null subjects in matrix clauses are also *pro*, which can be legitimated by an Agree operation, in the sense of Pan 2016, 2017¹⁰⁴. In the latter case, I adopt the idea of Li 2007 and assume that *pro* contains an uD feature, which is valued by the interpretable D feature of its antecedent. In this sense, it is plausible to consider that the key factor that permits null subjects in Chinese is the D feature contained in *pro*.

Interestingly, the Agree operation of Pan 2016, 2017 may be consistent with the A'-movement proposal of Huang 1984, since, in MP, movement is analyzed as Merge and Agree (feature valuing). In the current thesis, I propose that *pro* should not have a variable feature, which does not favor the proposal to analyze Chinese null subjects in matrix clauses as variable. However, future studies can analyze more structures and explore the relationship between the Agree operation of Pan and the A'-movement of the GB framework. Of course, there is also a possibility that the Agree operation analysis can be replaced by the GCR, since it is possible to consider that *pro* in matrix subject position is controlled by a topic. This question can also be addressed in future studies.

¹⁰⁴ I adopt the Agree operation of Pan 2016, 2017, but not that of Liu 2014. The former does not allow agree inside an island, while the latter allows it.

For Question (2), I propose in this thesis an analysis based on the topic chain theory developed by Pu & Pu 2014. I firstly adopt the proposal of Pu & Pu 2014 that several sentences/clauses in Chinese may form a topic chain. Then I propose that each clause in Chinese may have a syntactic topic position, which can be filled by an operator without phonetic realization, and an abstract topic, which is determined by the context. The abstract topic gives a referential value to the syntactic topic (the operator), which agrees with the null subjects inside the topic chain and shares with them the referential value (this corresponds to the Agree operation I discussed for Question (1)). An overt pronoun in Chinese may also serve as the head of a topic chain, marking the beginning of a new topic chain. Meanwhile, the referential value of the overt pronoun may be inherited from the previous topic chain. This explains why in Chinese an overt pronoun is also acceptable or even preferred in topic maintenance contexts and answers question (2), that is to say, although a null subject and an overt pronoun may be used in the same context, the overt pronoun implies the beginning of a new topic chain, while the null pronoun does not.

For Question (3), the current thesis reviewed the previous studies that analyzed this issue in purely syntactic terms (Huang 1982, Zhao 2014, Lust et al. 1996) and showed that these approaches have some shortcomings. As a result, I examined this issue using the topic chain analysis.

Under this proposal, the backward anaphora structures with embedded overt pronoun can be analyzed as one sentence containing two topic chains. Since the matrix subject is DP, it is not proper for it to inherit its referential value from a pronoun, which explains why Chinese does not favor the coreferential interpretation in such backward anaphora structures, see (1):

(1) [subordinate pronoun ...] [matrix DP ...]
 Topic chain 1 Topic chain 2

However, such an analysis is not applicable to the null subject in backward anaphora, since studies like Biller-Lapin 1983 have argued that a null subject can serve as a conventional way to introduce a new entity. Considering this fact, I propose that, for backward anaphora with an embedded null subject, there is only one topic chain, thus, the null subject can share the referential value of the matrix subject, which serves as the head of the topic chain. See (2):

(2) [subordinate [-] ...] [matrix DP ...]

Topic chain 1

In this way, the current thesis proposes that the interpretative asymmetry for overt pronoun between forward and backward anaphora of Chinese is also constrained by discourse-pragmatic factors, namely the topic chain theory.

It is also worth noting that the proposal of topic chain may be extended to other null subject languages, such as the Romance null subject languages. Since pronominal resolution in both the Romance null subject languages and Chinese is related to discourse pragmatic factors, it will be interesting to test in future studies whether the rules for topic maintenance and switch in the Romance null subject languages can be compared to the Maximum Discourse Coherence and Minor Thematic Discontinuity proposed by Pu & Pu 2014 for Chinese.

Such a comparison between Chinese and Romance null subject languages such as EP may lead us to reconsider the difference between discourse null subject languages and consistent null subject languages. There is a possibility (which should be tested by future studies) that these two groups of languages are not significantly different in the way that they resolve anaphoric relations in terms of discourse-pragmatics, since they may use a similar mechanism to recover the referential value of a null or overt pronoun. If this is true, then the real difference between these two groups of null subject languages only resides in the mechanism with which the null subject is licensed. In Romance null subject languages such as EP, the null subject licensing involves the presence of verbal agreement, while in null subject languages without verbal agreement, such as Chinese, the null subject licensing relies on the GCR, or the Agree operation with a topic in the left periphery of the sentence. It is worth noting that both the verbal agreement of EP and the Agree operation in Chinese involve a D feature, and even for the GCR analysis for Chinese, the D feature is also involved, in accordance with Li 2007. In this way, it can be concluded that in both languages the D feature is the key factor to permit referential null subjects, and the only difference between the two languages is how the D feature is presented in these languages.

The second objective of the thesis is to investigate how these properties of Chinese influence the L2 acquisition of EP by learners who speak Chinese as L1, in an experimental approach.

The first experiment (off-line questionnaire) compares the interpretation of null and overt pronouns in forward anaphora with their interpretation in backward anaphora. The results show that the native speakers prefer the subject antecedent for null pronouns and a non-subject (extra-discursive) antecedent for overt pronouns in both the forward and backward anaphora conditions. These results seem to prove that the interpretation of null and overt pronouns in EP is decided by the syntactic position of the antecedent.

However, the performance of the Chinese learners differs in forward and backward anaphora. In forward anaphora, where Chinese is different from EP, the learners' interpretation seems to be influenced by their L1, since they prefer the subject antecedent for overt pronoun and there is no change from B1 level to C1 level.

In backward anaphora, the L2 learners' interpretation differs from the forward anaphora interpretation, since the B1 learners show an undetermined interpretation in overt pronoun conditions and the C1 learners already have a native-like interpretation. Thus, an improvement from B1 to C1 level has been found in this condition.

Now we consider if these findings can be explained by the IH. The properties tested in Experiment 1 involve anaphoric resolution and are clearly inside the domain of interface between syntax and discourse-pragmatics. At first glance, the results in forward anaphora seem to confirm the IH, since the learners from both levels prefer the subject antecedent for the overt pronoun, which is not the one favored by the native speakers; However, in the case of backward anaphora, the C1 learners already have a native-like performance, and there is also an improvement from B1 to C1 level, which has not been found in forward anaphora conditions. In this case, it seems more plausible to consider that the L1 influences may facilitate the acquisition of the properties which are predicted to be difficult by the IH in certain contexts. When the L1 is consistent with the L2 (as is the case with backward anaphora in the current study), the L2 learners show improvement from B1 to C1 level; when the L1 is different from the L2 (forward anaphora in the current study), the L2 learners in both levels have difficulties and no change was found from B1 to C1 level.

If the previous analysis for Experiment 1 is on the right track, it may indicate that the degree of difficulty for the L2 learners in mastering anaphoric resolution depends on the similarity between the target language and the learners' L1, as far as the combination of L2 EP and L1

Chinese is considered.

The second experimental task (self-paced reading) aims to test the on-line processing of forward and backward anaphora by the native speakers and the L2 learners, as well as to test whether the L2 learners are influenced by the active search mechanism of Kazanina et al. 2007. The results for forward and backward anaphora structures are however analyzed separately considering their differences.

In forward anaphora, unlike in the off-line task, the native speakers seem to prefer the subject antecedent in both the null and overt pronoun conditions. This fact may be caused by the properties of the on-line task. The L2 learners also have a similar interpretation, as they also accept the subject antecedent in both conditions. Since there is no major difference between the L1 and L2 participants, it is not clear whether the L2 learners are exactly influenced by their L1, or in the on-line task the preference for subject antecedent is the universal processing strategy.

In backward anaphora, the native speakers seem to prefer the subject antecedent in null pronoun conditions, since there is a marginal effect and an effect in the post-critical and post-post critical regions, respectively. However, there is no effect in the overt pronoun conditions, which may be caused by the properties of the on-line task. In this case, the only finding for the backward anaphora conditions consists of their off-line answers, which show a preference for the non-subject antecedent in the Overt-Match condition, in both the native and L2 groups.

For the L2 learners, there is no significant effect found in the current test. As a consequence, it is not possible to conclude if the active search mechanism has an influence on the L2 processing from the results of the current thesis. This fact may be caused by the design of the current study, which will be discussed below.

In fact, as discussed in Chapter 5, there are not many studies which analyzed backward anaphora with an on-line paradigm, especially in case of null subject languages, which have two pronominal forms. Experiment 2 of the current thesis explores this domain, which does not result in a very clear result. In backward anaphora conditions, there is only an effect in the null subject condition for native speakers, which may corroborate the active search mechanism of Kazanina et al. 2007. However, there is no effect in the overt pronoun condition, either for the L1 or the L2 participants. This fact may suggest that the design of the current experiment should

be improved. Experiment 2 only tested sentences with one referent in the matrix clause, which require the participant to establish an anaphoric relation with an extra-linguistic antecedent, if they do not accept the only referent presented in the sentence. This fact may increase the processing cost, which eventually results in the lack of effect. In this sense, future studies should test backward anaphora with sentences with two referents in the matrix clause in order to investigate whether the choice between the matrix subject and object may influence the participants' behavior. Another aspect that should be considered in future studies is the type of online tasks. For example, it would be interesting to compare the results of a self-paced reading task with those of an eye-tracking task.

In fact, the question of the number of potential antecedents should also be addressed for the offline task of the current thesis, as it differs from previous studies. In studies on Italian, such as Sorace & Filiaci 2006, Serratrice 2007 and Belletti et al. 2007, the authors allowed the participants to choose the antecedent of the pronoun from three possibilities: a subject, an object and an extra-linguistic entity. In studies on EP, Lobo et al. 2017 allowed the participants to choose from two possibilities: a subject and an object. In the current study, the participants also had two options: a subject and an extra-linguistic entity. Another difference is the fact that the previous studies mentioned here used a picture verification task, while the current thesis used a comprehension task. As a result, these differences may be an artifact: the comparison is not straightforward since different experimental paradigms were used.

Considering these factors, future studies should test other types of structures, namely those with a subject and an object in matrix clause of backward anaphora, and compare those results with the ones presented in this study. It is also important to compare the results between different types of tasks, for example, comparing comprehension task with picture verification task.

Bibliographic References

- Adger, D. & G. Ramchand. 2001. Phases and interpretability. *Proceedings of WCCFL 20*: 1 – 14. Somerville, Mass.: Cascadilla Press.
- Adger, D. & G. Ramchand. 2005. Merge and Move: Wh-dependencies revisited. *Linguistic Inquiry* 36: 161 – 193.
- Alexiadou, A. & E. Anagnostopoulou. 1998. Parametrizing Agr: word order, verb-movement and EPP-checking. *Natural Language and Linguistic Theory*, 16: 491-539.
- Almor, A. 1996. *NP anaphora and focus - the informational load hypothesis*. Unpublished Ph.D., Brown University, Providence, RI.
- Almor, A. 1999. Noun-phrase anaphora and focus: The informational load hypothesis. *Psychological Review*, 106, pp. 748-765.
- Almor, A. 2000. Constraints and mechanisms in theories of anaphor processing. In M. W. Crocker, M. Pickering, & C. Clifton, *Architectures and Mechanisms for Language Processing*, pp. 341-354. Cambridge, UK: Cambridge University Press.
- Ambar, M. 1988. *Para uma Sintaxe da Inversão Sujeito Verbo em Português*. PhD dissertation, University of Lisbon.
- Ariel, M. 1990. *Accessing noun-phrase antecedents*. London: Routledge.
- Ariel, M. 2001. Accessibility theory: an overview. In Sanders, T., J. Schilperoord & W. Spooren (eds.). *Text representation: linguistic and psycholinguistic aspects*. Amsterdam: John Benjamins.
- Barbosa, P. 1995. *Null Subjects*. Ph.D. dissertation. MIT, Cambridge, Mass.
- Barbosa, P. 2000. Clitics: a Window into the Null Subject Property. In J. Costa (org.), *Essays in Portuguese Comparative Syntax*. New York: Oxford Press.
- Barbosa, P. 2009. Two kinds of subject pro. *Studia Linguistica* 63-1, 2-58.
- Barbosa, P. 2011. Partial *pro*-drop as null NP-anaphora. *Proceedings of NELS 41*, Amherst (Mass): GLSA Publications.
- Barbosa, P. 2013. *Pro* as a minimal NP: Towards a unified approach to *pro*-drop. <http://ling.auf.net/lingbuzz/001949>.
- Barss, A. 1986. *Chains and anaphoric dependence: on reconstructions and its implications*. PhD dissertation, Massachusetts Institute of Technology.
- Bates, D., Maechler, M., Bolker, B. & Walker, S. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1-48. doi: 10.18637/jss.v067.i01
- Beck, M.L. 1998. L2 acquisition and obligatory head movement: English-speaking learners of German and the local impairment hypothesis. *Studies in Second Language Acquisition* 20, 311-48.
- Belletti, A. 2004. Aspects of the low IP area. In Rizzi, L. (ed.), *The structure of CP and IP*, pp. 16-51, Oxford University Press.
- Belletti, A., Bennati, E. & Sorace, A. 2007. Theoretical and developmental issues in the syntax of subjects: evidence from near-native Italian. *Natural Language and Linguistic Theory* 25, 657-689.
- Biller-Lappin, Y. 1983. *Backwards Anaphora in Discourse*, MA thesis, Tel-Aviv University.
- Bini, M. 1993. La adquisición del italiano: mas allá de las propiedades sintácticas del pará

- metro pro-drop. In J.M. Liceras (ed.) *La lingüística y el análisis de los sistemas no nativos*. Doverhouse, 126-39.
- Bley-Vroman, R. 1990. The logical problem of foreign language learning. *Linguistic Analysis* 20, 3-49.
- Brito, A. M. 1991. Ligação, co-referência e o princípio evitar pronome. *Encontro de Homenagem a Óscar Lopes*. (pp. 101-121). Lisboa: Associação Portuguesa de Linguística.
- Brito, A. M., Duarte, I. & Matos, G. 2003. Estrutura da frase simples e tipos de frases. In Mateus et al. (Eds.), *Gramática da Língua Portuguesa* (pp. 433-505). Lisboa: Editorial Caminho.
- Büring, D. 2005. *Binding theory*. Cambridge Textbooks in Linguistics, Cambridge University Press.
- Calabrese, A. 1986. Pronomina: some properties of the Italian pronominal system. Cambridge, Massachusetts: *MIT working papers in Theoretical Linguistics*, vol 8, 1-46.
- Camacho, José. 2013. *Null subjects*. Cambridge University Press.
- Canceiro, N. 2016. *Coordenação, subordinação adverbial e relações referenciais entre sujeitos*. MA dissertation, Universidade de Lisboa.
- Cardinaletti, A. & Starke, M. 1999. The typology of structural deficiency: A case study of the three classes of pronouns. In H. Van Riemsdijk (ed), *Clitics in the Languages of Europe*. Berlin: de Gruyter, 145-235.
- Carminati, M. N. 2002. *The processing of Italian subject pronouns*. PhD dissertation, University of Massachusetts at Amherst, Amherst (Mass): GLSA Publications.
- Chao, Y. R. 1968. *A grammar of spoken Chinese*. Los Angeles: California University Press.
- Chomsky, N. 1980. On binding. *Linguistic Inquiry* 11, 1-46.
- Chomsky, N. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.
- Chomsky, N. 1982. *Some Concepts and Consequences of the Theory of Government and Binding*. Linguistic Inquiry Monograph Six. Cambridge: The MIT Press.
- Chomsky, N. 1986. *Knowledge of Language – its nature, Origins and use*. New York: Praeger.
- Chomsky, N. 1995. *The Minimalist Program*. Cambridge, Mass.: MIT Press.
- Chomsky, N. 2001. Derivation by phase. Kenstowicz M., *Ken Hale: A Life in Language*, Cambridge Mass.: MIT Press, pp. 1-52.
- Clahsen, H. & Hong, U. 1995. Agreement and null subjects in German L2 development: new evidence from reaction-time experiment. *Second Language Research* 11, 1, 57-87
- Clahsen, H. & Muysken, P. 1989. The UG paradox in L2 acquisition. *Second Language Research* 5: 1-29
- Clark, H.H. & Marshall, C.R. 1981. Definite reference and mutual knowledge, In A.K. Joshi et al. (eds), pp. 10-63.
- Cole, P. & Sung, M.-L. 1994. Head movement and long-distance reflexives. *Linguistic Inquiry* 25: 355-406.
- Costa, J. 2000. Word Order and Discourse Configurability in European Portuguese. In *Portuguese Syntax. New Comparative Studies.*, ed. J. Costa, 94 - 115. New York: Oxford University Press.
- Costa, A., Faria, I. & Matos, G. 1998. Ambiguidade referencial na identificação do sujeito em estruturas coordenadas. In Mota, M. A. and R. Marquilha (eds) *Actas do XIII Encontro Nacional da Associação Portuguesa de Linguística*. Lisboa: Associação Nacional de Linguística, 173-188.

- Costa, A., Faria, I. & Matos, G. 1999. Competitive information sources in referential ambiguity resolution. In Pinto, M^a da Graça, J. Veloso and B. Maia (eds) *Psycholinguistics on the Threshold of the Year 2000 — Proceedings of 5th International Congress of the International Society of Applied Pshycholinguistics (ISAPL 97)*. Porto: Faculdade de Letras da Universidade do Porto, 133-138.
- Costa, J. & Ambulate, J. 2010. The acquisition of embedded subject pronouns in European Portuguese, in M. Iverson et al. (orgs.) *Proceedings of the 2009 Mind/Context Divide Workshop*. Somerville, MA: Cascadilla Proceedings Project, pp. 1–12.
- Costa, A. & Mato, Gs. 2012. Processamento da correferência e sujeitos anafóricos – dados sobre o Português Europeu e Brasileiro. *Revista Linguística - Correferência anafórica: representação, aquisição e processamento*, Vol. 8 N.2., pp. 289-323.
- Crain, S. & Fodor, J. D. 1985. How can grammars help parsers. In D. R. Dowty, L. Karttunen, & A. M. Zicky (Eds.), *Natural language parsing: psychological, computational, and theoretical perspectives* (pp. 94 - 128). Cambridge: Cambridge University Press.
- Déchaine, R.-M. & Wiltschko, M.. 2003. On pro-nouns and other ‘pronouns’. In Coene, M. & Y. D'hulst (eds), *From NP to DP. Volume 1: The syntax and semantics of noun phrases*. Amsterdam: John Benjamins, (=Linguistik Aktuell / Linguistics Today 55), 71-89.
- Drummer, J.-D. & Felser, C. 2018. Cataphoric pronoun resolution in native and non-native sentence comprehension. *Journal of Memory and Language* 101, 97-113. doi: 10.1016/j.jml.2018.04.001.
- Duarte, I. 1987. *A construção de topicalização na gramática do português: Regência, ligação e condições sobre movimento*. PhD Dissertation, Universidade de Lisboa.
- Eliseu, A. 1984 *Verbos Ergativos do Português. Descrição e Análise*. Faculdade de Letras da Universidade de Lisboa, Lisboa.
- Epstein, S., Flynn, S. & Martohardjono, G. 1996. Second language acquisition: Theoretical and experimental issues in contemporary research. *Brain and Behavioral Sciences*, 19, 677-758.
- Eubank, L. 1993/94. On the transfer of parametric values in L2 development. *Language Acquisition*, 3, 183-208.
- Eubank, L. 1996. Negation in early German-English Interlanguage: more Valueless Features in the L2 initial state. *Second Language Research* 12, 73-106.
- Fedele, E. & Kaiser, E. 2014. Looking Back and Looking Forward: Anaphora and Cataphora in Italian. In *Proceedings of the 37th Annual Penn Linguistics Conference*, Volume 20, Issue 1, pp. 81-90.
- Field, A. Miles, J., Field, Z. 2012. *Discovering Statistics Using R*. SAGE Publications.
- Filik, R. & A. J. Sanford. 2008. When is cataphoric reference recognised? In *cognition*, 107, (2008), pp. 1112-1121.
- Flynn, S. 1987. L2 Acquisition of Pronoun Anaphora: Resetting the Parameter." In B. Lust (ed.), *Studies in the Acquisition of Anaphora: Defining the Constraints*, Vol. 2. Dordrecht: D. Reidel, pp. 227-243, 1987.
- Flynn, S. 1996. A parameter setting approach to second language acquisition. In W. Ritchie and T. Bhatia (eds.), *Handbook of language acquisition* (pp. 121-158). San Diego: Academic Press.
- Flynn, S. & Martohardjono, G. 1994. Mapping from the initial state to the final state: the

- separation of universal principles and language-specific principles. In B. Lust, M. Suñer and J. Whitman (eds.), *Syntactic theory and first language acquisition: crosslinguistic perspectives*. Vol. 1: *Heads, projections and learnability* (pp. 319-335). Hillsdale, NJ: Lawrence Erlbaum.
- Frascarelli, M. 2007. Subjects, topics and the interpretation of referential *pro*. *Natural Language and Linguistic Theory*, 25-4, 691-734.
- Frazier, L. & Clifton, C. 1989. Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes*, 4, 93 - 126.
- Frazier, L., C. Clifton & J. Randall. 1983. Filling gaps: decision principals and structure in sentence comprehension. *Cognition*, 13, pp. 187-222.
- Fukumura, K. & R.P.G van Gompel. 2015. Effects of order of mention and grammatical role on anaphora resolution. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 41, 2, pp. 501-525.
- Gernsbacher, M., & Hargreaves, D. 1988. Accessing sentence participants: The advantage of first mention. *Journal of Memory and Language*, 27, pp. 699-717.
- Gernsbacher, M. 1989. Mechanisms that improve referential access. *Cognition* 32, pp. 99-156.
- Givón, T. 1983. Topic continuity and word order pragmatics in Ute. In T. Givón (ed.), *Topic Continuity in Discourse: Quantitative Cross Language Studies*. Amsterdam: John Benjamins, 343-363.
- Grice, H. P. 1975. Logic and Conversation. In P. Cole & J. Morgan (Eds.), *Syntax and Semantics 3: Speech Acts* (pp. 41-58). London: Academic Press.
- Grosz, B. J. & C. L. Sidner. 1986. Attentions, Intentions and the structure of discourse. *Computational Linguistics*, 12: 175-204.
- Grosz, B. J., A. K. Joshi & S. Weinstein. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21, 203-225.
- Haegeman, L. 2012. Adverbial Clauses, Main Clause Phenomena, and Composition of the Left Periphery. In *The Cartography of Syntactic Structures*, Volume 8. Oxford: Oxford University Press.
- Hawkins, R. & Chan, C. Y. 1997. The partial availability of Universal Grammar in second language acquisition: The 'failed functional features hypothesis'. *Second Language Research* 13: 187-226.
- Heim, I. 1982, *The Semantics of Definite and Indefinite NP's*. PhD dissertation. University of Massachusetts, Amherst.
- Holmberg, A. 2005. Is There a Little *Pro*? Evidence from Finnish. *Linguistic Inquiry* 36, pp. 533-564.
- Holmberg, Anders. 2010. Null Subject Parameters. In T. Biberauer, A. Holmberg, I. Roberts & M. Sheehan (eds), *Parametric Variation: Null subjects in minimalist theory*, Cambridge: Cambridge University Press, 88-124.
- Holmberg, A. & M. Sheehan 2010. Control into finite clauses in partial null-subject languages. In T. Biberauer, A. Holmberg, I. Roberts, & M. Sheehan (eds) *Parametric Variation: Null Subjects in Minimalist Theory*, Cambridge: Cambridge University Press, pp. 125-152.
- Huang, C.-T. J. 1982. *Logical relations in Chinese and the theory of grammar*. PhD dissertation, MIT.
- Huang, C.-T. J. 1984. On the Distribution and Reference of Empty Pronouns. *Linguistic Inquiry*

- 15, 531-574.
- Huang, C.-T. J. 1989. Pro-drop in Chinese: A generalized control theory. In Jaeggli, O. & K. Safir (eds) *The Null Subject Parameter*, Reidel, Dordrecht, pp. 185-214.
- Huang, C.-T. J. & B. C.-Y. Yang. 2013. Topic Drop and MCP. *The 87th Annual Meeting of the Linguistic Society of America*, Jan. 3-6, 2013, Boston, MA.
- Huang, C.-T. J., Y.-H. A. Li & Y. Li. 2009. *The syntax of Chinese*. Cambridge: Cambridge University Press.
- Jaeger, T. F. 2008. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59(4), 434-446. doi: 10.1016/j.jml.2007.11.007.
- Kato, M. 1999. Strong Pronouns, Weak Pronominals and the Null Subject Parameter, *Probus* 11:1-37.
- Kazanina, N. 2005. *The acquisition and processing of backwards anaphora*. Unpublished PhD dissertation, University of Maryland, College Park.
- Kazanina, N., E. F. Lau, M. Lieberman, M. Yoshida & C. Phillips. 2007. The effect of syntactic constraints on the processing of backwards anaphora. *Journal of Memory and Language*, v56, 3, 384-409.
- Kazanina, N. & C. Phillips. 2010. Differential effects of constraints in the processing of Russian cataphora. *The Quarterly Journal of Experimental Psychology*, 2010, 63 (2), 371-400.
- Keating, G., B. VanPatten & J. Jegerski. 2011. Who was walking on the beach? Anaphora resolution in Spanish heritage speakers and adult second language learners. *Studies in Second Language Acquisition* 33. 193-221.
- Laka, I. 1990. Negation in Syntax, on the nature of functional categories and projections. PhD dissertation, MIT.
- Lambrecht, K. 1994. *Information Structure and Sentence Form: Topic, Focus and the Mental Representation of Discourse Referents*. Cambridge: Cambridge University Press.
- Larson, R. & M. Luján. 1989. *Emphatic pronouns*. Ms. Stony Brook.
- Larson, R. & M. Luján. 1991. Focused Pronouns. Paper presented at *Conference on Japanese Linguistics*, Rochester University.
- Lasnik, H. 1991. On the necessity of binding conditions. In R. Freidin (ed.) *Principles and Parameters in Comparative Grammar*. MIT Press, pp. 7-28.
- Lasnik, H. & Saito, M. 1991. On the subject of infinitives. In L. Dobrin, L. Nichols, and R. Rodriguez (Eds.) *Papers from the 27th Regional Meeting of the Chicago Linguistic Society* (pp. 324-343). Chicago: Chicago Linguistic Society.
- Li, Y.-H. A. 2007. Beyond empty categories. *Bulletin of the Chinese Linguistic Society of Japan*, 254, 74-106.
- Li, C. N. and S. A. Thompson 1976. Subject and topic: A new typology of language. In C. N. Li (ed) *Subject and topic*, New York: Academic Press, pp. 457-489.
- Liu, C.-M. 2014. *A Modular Theory of Radical Pro Drop*. PhD dissertation, Harvard University Cambridge, Massachusetts.
- Lobo, M. 1994. *Para uma Redefinição do Parâmetro do Sujeito Nulo*. MA Thesis. Faculdade de Letras, Universidade de Lisboa.
- Lobo, M. 1995. Fenómenos Relacionados com o Parâmetro do Sujeito Nulo em português. *Atti del XXI Congresso Internazionale di Linguistica e Filologia Romanza*, Palermo, 18-24

- setembre 1995, Tübingen: Niemeyer.
- Lobo, M. 2003. *Aspectos da Sintaxe das Orações Subordinadas Adverbiais do Português*, Dissertação de Doutorado, Faculdade de Ciências Sociais e Humanas, Universidade Nova de Lisboa
- Lobo, M. & C. Silva. 2016. Ambiguidade pronominal em orações adverbiais do português europeu: crianças vs. adultos. *Revista da Associação Portuguesa de Linguística* 2, 319-338.
- Lobo, M., A. Madeira & C. Silva. 2017. Interpretação de pronomes sujeito anafóricos e catafóricos por falantes de português L2: efeitos da língua materna. *Revista da Associação Portuguesa de Linguística* 3, 135-153.
- Lozano, C. 2002. The interpretation of overt and null pronouns in non-native Spanish. *Durham Working Papers in Linguistics* 8: 53–66.
- Lozano, C. 2009. Selective deficits at the syntax-discourse interface: Evidence from the CEDEL2. In N. Snape, Y.I. Leung & M. Sharwood-Smith (eds). *Representational Deficits in SLA corpus* (pp. 127-166). Amsterdam: John Benjamins.
- Lǔ, S. 2000. *Xiandai Hanyu Babaici* [800 words in modern Chinese]. Beijing: Commercial Press.
- Luegi, P. 2012. *Processamento de sujeitos pronominais em Português: efeito da posição estrutural dos antecedentes*. Universidade de Lisboa, PhD dissertation.
- Luján, M. 1986. Stress and Binding of Pronouns, in A. Farley, P. Farley and K-E. McCullough (eds.) *Papers from the Parasession on Pragmatics and Grammatical Theory*. Chicago Linguistics Society, University of Chicago, Chicago, Illinois, 248-262.
- Lust, B., Y.-C. Chien, C.-P. Chiang and J. Eisele. 1996. Chinese Pronominals in Universal Grammar: A Study of Linear Precedence and Command in Chinese and English Children's First Language Acquisition. *Journal of East Asian Linguistics*, Vol. 5, No. 1 (Jan., 1996), 1-47.
- Madeira, A., M.F. Xavier & M.L. Crispim. 2009. A aquisição de sujeitos nulos em português L2. *Estudos da Língua(gem) (Pesquisas em Aquisição da Linguagem)* 7.2, Universidade Estadual do Sudoeste da Bahia, pp. 163-198.
- Madeira, A., Xavier, M., & Crispim, M. 2010. Interpretação semântica e/ou pragmático-discursiva de sujeitos na aquisição de português L2. *Textos seleccionados do XXV Encontro Nacional da Associação Portuguesa de Linguística* (pp. 513-529). Lisboa: Colibri.
- Madeira, A., M. F. Xavier & M. L. Crispim. 2012. Uso e interpretação de sujeitos pronominais em português L2. A. Costa, C. Flores & N. Alexandre (orgs.) *Textos Seleccionados do XXVII Encontro da Associação Portuguesa de Linguística*. Lisboa: Associação Portuguesa de Linguística, pp. 376-397.
- Margaza, P. & A. Bel. 2006. Null subjects at the syntax-pragmatics interface: evidence from Spanish interlanguage of Greek speakers, in: M. Grantham O'Brien, C. Shea, & J. Archibald (Eds.) *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference (GASLA 2006)*. Somerville, MA: Cascadilla Proceedings Project, pp. 88-97.
- Martins, A. M. 1993. Enclisis, VP-deletion and the Nature of Sigma (Σ), ms., Communication presented in *Going Romance 1993*, Utrecht.

- Matos, G. & I. Duarte 1984. Clíticos e Sujeito Nulo no Português: contribuições para uma teoria de *pro*. In *Boletim de Filologia*, tomo XXIX, Lisboa: Centro de Linguística da Universidade de Lisboa, pp. 479-538.
- Menuzzi, S. & M., Lobo 2016. Binding and Pronominal Forms in Portuguese. In Wetzels, Menuzzi & Costa (eds.), *Handbook of Portuguese Linguistics*. Wiley-Blackwell, 338-355.
- Mohanan, K.P. 1983. 'Functional and anaphoric control', *Linguistic Inquiry* 14:4, pp.641-74.
- Móia, T. 2001. Aspectos Sintactico-Semânticos das Orações Relativas com quando e como. *Actas do XVI Encontro Nacional da Associação Portuguesa de Linguística (Coimbra, 28-30 de Setembro de 2000)* (pp. 349-361). Lisboa: Associação Portuguesa de Linguística.
- Montalbetti, M. 1984. *After Binding- On the interpretation of pronouns*. PhD Dissertation. MIT. Cambridge, Massachusetts.
- Montrul, S. & C. Rodríguez-Louro. 2006. Beyond the syntax of the null subject parameter: a look at the discourse-pragmatic distribution of null and overt subjects by L2 learners of Spanish. V. Torrens & L. Escobar (orgs.) *The acquisition of syntax in Romance languages*. Amsterdam: John Benjamins, pp. 401-418.
- Morgado, S. 2012. *Processamento da co-referência pronominal: informação sintáctica e semântica*. Tese de Mestrado apresentada à Faculdade de Letras da Universidade de Lisboa.
- Neeleman, A. & K. Szendrői 2007. Radical pro-drop and the morphology of pronouns, *Linguistic Inquiry* 38, pp. 671-714.
- Neeleman, A. & Weerman, F. 1997. L1 and L2 word order acquisition. *Language Acquisition* 6:125-170.
- Pan, V. J. 2014. Wh-ex-situ in Chinese: Mapping Between Information Structure and Split CP. *Linguistic Analysis, Volume 39* (3-4), p. 371-413.
- Pan, V. J. 2016. *Resumptivity in Mandarin Chinese: A Minimalist Account*. [Trends in Linguistics. Studies and Monographs 298], Berlin: De Gruyter Mouton.
- Pan, V. J. 2017. Phasal Agree as Precondition on Multiple Transfers: Evidence From A-bar Dependency in Chinese. In 2017 *The Syntax-Morphology Interface in Generative Grammar: Proceedings of The 19th Seoul International Conference on Generative Grammar*, Changguk Yim (eds.), The Korean Generative Grammar Circle: 261-276.
- Pan, V. J. & W. Paul. 2018. The Syntax of Complex Sentences in Mandarin Chinese: A Comprehensive Overview with Analyses. In *Linguistic Analysis*, 42 1-2 2018.
- Partee, B. 1987. Noun Phrase Interpretation and Type-Shifting Principles. In Groenendijk et al. (eds) *Studies in Discourse Representation Theory and the Theory of Generalized Quantifiers*, Dordrecht: Foris, pp. 115-143.
- Peirce, J. W. 2007. PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods*, 162 (1-2):8-13 doi:10.1016/j.jneumeth.2006.11.017.
- Pérez-Leroux, A. & W. Glass. 1999. Null anaphora in Spanish second language acquisition: Probabilistic versus generative approaches. *Second Language Research* 15: 220-49.
- Perlmutter, D. 1971. *Deep and Surface Constraints in Syntax*. New York: Holt, Rinehart & Winston.
- Pesetsky, D. & E. Torrego 2004. "Tense, case, and the nature of syntactic categories." In J. Guéron and J. Lecarme (eds), *The Syntax of Time*, Cambridge, Mass.: MIT Press.
- Pollock, J.-Y. 1997. *Langage et Cognition: Introduction au Programme Minimaliste de la Grammaire Générative*. Paris: Presses Universitaires de France.

- Prince, E. F. 1981. Toward a taxonomy of Given-New information, in P. Cole (ed.), pp.223-55.
- Pu, M.-M. & Q. Pu. 2014. Zero Anaphora and Topic Chain: A Cross-Linguistic Study. *International Journal of Linguistics and Communication*, March 2014, Vol. 2, No. 1, 27-44.
- Ramchand, G. & Reiss, C. (Eds.) 2007. *The Oxford handbook of linguistic interfaces*. Oxford: OUP.
- Reinhart, T. 1983. Coreference and Bound Anaphora: A Restatement of the Anaphora Questions, *Linguistics and Philosophy* 6, pp. 47 – 88.
- Reinhart, T. & E. Reuland. 1993. Reflexivity. *Linguistic Inquiry* 24, pp. 657-720.
- Rigau, G. 1988. Strong Pronouns, *Linguistic Inquiry* 19:3, 503-511.
- Rizzi, L. 1982. *Issues in Italian Syntax*. Dordrecht: Foris.
- Rizzi, L. 1986. Null Objects in Italian and the Theory of *pro*, *Linguistic Inquiry* 17, 501-557.
- Rizzi, L. 1997. The fine structure of the left periphery. In L. Haegeman (Ed.), *Elements of grammar: Handbook in Generative Syntax* (pp. 281-337). Dordrecht: Kluwer.
- Rizzi, L. 2016. EPP and ECP revisited: The role of labeling, in Ernestina Carrillho, Alexandre Fiéis, Maria Lobo and Sandra Pereira (eds.), *Romance Languages and Linguistic Theory 10, Selected papers from Going Romance 28*, Lisbon: 211-232, John Benjamins Publish.
- Roberts, I. 2007. Taraldsen's Generalization and Language Change: Two Ways to Lose Null Subjects. Ms, University of Cambridge.
- Roberts, I. 2010. A deletion analysis of null subjects. Biberauer, T., A. Holmberg, I. Roberts & M. Sheehan 2010 (eds.). *Parametric Variation: Null subjects in minimalist theory*. Cambridge, UK: Cambridge University Press, pp. 58-87.
- Roberts, I. & A. Holmberg. 2010. Introduction: parameters in minimalist theory. In Biberauer, T., Holmberg, A., Roberts, I., Sheehan, M. (eds.), *Parametric Variation: Null Subjects in Minimalist Theory*. Cambridge: Cambridge University Press, 1–57.
- Rothman, J. 2007. Pragmatic solutions for syntactic problems: understanding some L2 Syntactic errors in terms of pragmatic deficits. Baauw, S., F. Drijkoningen & M. Pinto (orgs.) *Romance languages and linguistic theory 2005*. Amsterdam: John Benjamins, pp. 297-318.
- Rothman, J. 2008. How Pragmatically Odd!: Interface Delays and Pronominal Subject Distribution in the L2 Spanish of English Natives. *Studies in Hispanic and Lusophone Linguistics*, 1, pp. 317-339.
- Rothman, J. 2009. Pragmatic deficits with syntactic consequences?: L2 pronominal subjects and the syntax-pragmatics interface. *Journal of Pragmatics* 41(5). 951-973.
- Rouveret, A. 2002. How are Resumptive Pronouns Linked to the Periphery? *Linguistic Variation Yearbook* 2: 123 – 184.
- Rouveret, A. 2008. Phasal Agreement and Reconstruction. In *Foundational Issues in Linguistic Theory* (ed. by Robert Freidin, Carlos P. Otero and Maria Luisa Zubizarreta), 167 – 195. Cambridge, Mass: The MIT Press
- Rouveret, A. 2019 Computational and semantic aspects of resumption, in *Interfaces of Grammar*, Jianhua Hu & Haihua Pan eds., [Language Faculty and Beyond (LFAB)], John Benjamins.
- Safir, K. 2013. Syntax, binding and patters of anaphora. In Den Dikken (ed.), *The Cambridge Handbook of Generative Grammar*, CUP:515-576.

- Saito, M. 2007. Notes on East Asian argument ellipsis. *Language Research* 43, pp. 203-227.
- Serratrice, Ludovica. 2007. Cross-linguistic influence in the interpretation of anaphoric and cataphoric pronouns in English–Italian bilingual children. *Bilingualism: Language and Cognition* 10 (3), 2007, Cambridge University Press, 225-238.
- Simpson, A., Wu, Z., & Li, Y. 2016. Grammatical roles, Coherence Relations, and the interpretation of pronouns in Chinese. *Lingua Sinica*. 2016, 2:2.
- Schachter, J. 1988. Second language acquisition and its relationship to Universal Grammar. *Applied Linguistics* 9, 219–35.
- Schwartz, B. D. 2003. Why child L2 acquisition? In J. van Kampen & S. Baauw (eds.) *Proceedings of GALA 2003*, Vol. 1. Utrecht: Netherlands Graduate School of Linguistics, 47-66.
- Schwartz, B. D. & Sprouse, R. 1994. Word order and nominative case in nonnative language acquisition: a longitudinal study of (L1 Turkish) German interlanguage. In T. Hoekstra and B. D. Schwartz (eds.), *Language acquisition studies in generative grammar* (pp. 317-368). Amsterdam: John Benjamins.
- Schwartz, B. D. & Sprouse, R. 1996. L2 cognitive states and the full transfer/full access model. *Second Language Research*, 12: 40-72.
- Sorace A. 2011. Pinning down the concept of ‘interface’ in bilingualism. *Linguistic Approaches to Bilingualism* 1: 1–33.
- Sorace, A. & F. Filiaci. 2006. Anaphora resolution in near-native speakers of Italian. *Second Language Research* 22:3, 339–368.
- Stowe, L. 1986. Evidence for online gap creation. *Language and Cognitive Processes*, 1, 227 – 245.
- Taraldsen, T. 1980. “On the NIC, vacuous application and the *that*-trace filter.” Indiana University Linguistics Club.
- Teng, T.-H. 1985. *On pronominal reference in Chinese*. Master’s thesis, National Taiwan Normal University, Taipei.
- Tomioka, S. 2003. The semantics of Japanese null pronouns and its crosslinguistic implications. In K. Schwabe & S. Winkler. *The interfaces: Deriving and interpreting omitted structures*, Amsterdam: John Benjamins, 321-340.
- Tsao, F. 1977. *A functional study of topic in Chinese: the first step toward discourse analysis*. Doctoral dissertation, USC, Los Angeles, California.
- Tsimpli, I., A. Sorace, C. Heycock & F. Filiaci 2004. First language attrition and syntactic subjects: a study of Greek and Italian near-native speakers of English. *International Journal of Bilingualism* 8, pp. 257–77.
- Tsimpli, I. & A. Sorace, 2006. Differentiating interfaces: L2 performance in syntax-semantics and syntax-discourse phenomena. *Proceedings of the Annual Boston University Conference on Language Development*, 30 (2), pp. 653-664.
- Vainikka, A. & Young-Scholten, M. 1994. Direct access to X-bar theory: Evidence from Korean and Turkish adults learning German. In T. Hoekstra & B.D. Schwartz (Eds.), *Language Acquisition Studies in Generative Grammar*, 265-316. Amsterdam: John Benjamins.
- Vainikka, A. & Young-Scholten, M. 1996a. Gradual development of L2 phrase structure. *Second Language Research* 12: 7-39.

- Vainikka, A. & Young-Scholten, M. 1996b. The early stages in adult L2 syntax: Additional evidence from Romance speakers. *Second Language Research* 12: 140-76.
- Valmala, V. 2009. On the position of central adverbial clauses, *Anuario del Seminario de Filología Vasca "Julio de Urquijo"*, XLIII, 951-970.
- Van Gompel, R. P. G., & Liversedge, S. P. 2003. The influence of morphological information on cataphoric pronoun assignment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 128 - 139.
- White, L. 1985. The 'pro-drop' parameter in adult second language acquisition. *Language Learning* 35, pp. 47-62.
- White, L. 1989. *Universal grammar and second language acquisition*. Amsterdam: John Benjamins.
- White, L. 2011. The Interface Hypothesis: How far does it extend? *Linguistic Approaches to Bilingualism* 1:1, Amsterdam: John Benjamins, pp. 108-110.
- Xing, F. 2001. *Hanyu fujū yanjiu* [A study on complex sentences]. Beijing: Shangwu.
- Xu, L. 1986. Free empty category. *Linguistic Inquiry* 17, pp. 75-93.
- Xu, J. 2003. *Anaphora in Chinese texts*. Chinese Academy of Social Science Press.
- Yang, C. L., P. C. Gordon, R. Hendrick & J. T. Wu. 1999. Comprehension of Referring Expressions in Chinese. *Language and Cognitive Processes*, 14:5-6, 715-743, DOI: 10.1080/016909699386248.
- Zhang, N. N. 2002. Island Effects and Episodic Eventualities in Chinese Topicalization. In *Linguistics by Heart: in honor of Horst-Dieter Gasde* (ed. by D. Hole, P. Law, and N. Zhang). ZAS-Berlin.
- Zhao, L. X. 2012. Interpretation of Chinese overt and null embedded arguments by English-speaking learners. *Second Language Research* 2012 28, 169-190.
- Zhao, L. X. 2014. Ultimate attainment of anaphora resolution in L2 Chinese. *Second Language Research*, 30:3, 381-407.
- Zheng, Y., P. Luegi, A. Madeira & G. Matos. 2018. How Chinese learners of L2 European Portuguese interpret null and overt pronouns in forward and backward anaphora. In *Experimental Studies on Portuguese, Revista Linguística*, 14-2, pp. 291-321.
- Zubizarreta, M. 1998. *Prosody, Focus and Word Order*. Cambridge, MA: MIT Press.

Appendix 1 Introduction for the off-line questionnaire task

Experiência de Questionário

Dados do/a participante

Nome: _____ Idade: _____

Sexo: _____

Língua materna: _____

Teve contacto com alguma outra língua para além da língua materna antes dos 6 anos?

Se for sim, qual/quais? _____

Outras línguas que fala:

1. _____

2. _____

3. _____

4. _____

Experiência de Questionário

Nível do português:

A1

B1

C1

A2

B2

C2

Quando chegou a Portugal? _____ (mês) _____ (ano)

Com que idade começou a aprender português? _____

Com que frequência fala português?

Diariamente

Pelo menos 2 vezes por semana

Pelo menos 1 vez por mês

Raramente

Experiência de Questionário

Instruções

Neste questionário, vai ler algumas frases. Depois de cada frase existe uma pergunta de compreensão sobre a frase lida. Por favor, escolha entre a opção A e B a resposta mais adequada para a pergunta.

Algumas perguntas não têm respostas certas. Isto quer dizer que, diferentes participantes podem ter respostas diferentes.

Só pode escolher uma opção para cada pergunta. Se pensar que ambas as opções são adequadas, por favor escolha a opção que ache melhor.

Muito obrigado pela sua colaboração!

(English translation)

Questionnaire Experiment

Participant's profile

Name: _____

Age: _____ Gender: _____

First Language: _____

Did you have contact with some languages other than your mother tongue before 6 years old?

If it is yes, what is/are the language(s)? _____

Other languages that you speak:

1. _____

2. _____

3. _____

4. _____

Questionnaire Experiments

Level of Portuguese:

A1

B1

C1

A2

B2

C2

When did you arrived Portugal? _____ (month) _____ (year)

With which age did you start to learn Portuguese? _____

How frequently do you speak Portuguese?

Daily

At least 2 times per week

At least 1 time per month

Rarely

Questionnaire Experiment

Introductions

In this questionnaire, you will read some sentences. After each sentence, there is a question about the comprehension of the sentence you read. Please choose between options A and B the answer which is more adequate for the question.

Some questions do not have a correct answer. That is to say, different participants may have different ideas about the question.

You can only choose one option for each question. If you think that both of the options are adequate, please choose the one that you think is better.

Many thanks for your collaboration!

Appendix 2 Personal profile questionnaire for the self-paced reading task

Dados do/a participante

Nome: _____

e-mail: _____

Idade: _____ Sexo: _____

Língua materna: _____

Teve contacto com alguma outra língua para além da língua materna antes dos 6 anos?

Se for sim, qual/quais? _____

Outras línguas que fala:

1. _____

2. _____

3. _____

4. _____

Nível do português:

A1

B1

C1

A2

B2

C2

Quando chegou a Portugal? _____ (mês) _____ (ano)

Com que idade começou a aprender português? _____

Com que frequência fala português?

Diariamente

Pelo menos 2 vezes por semana

Pelo menos 1 vez por mês

Raramente

(the translation for Appendix 2 is as the same as the first two pages of the translation for Appendix 1)

Appendix 3 Testing items for the Questionnaire task

Legend:

BO = Backward Overt anaphora

BN = Backward Null anaphora

FO = Forward Overt anaphora

FN = Forward Null anaphora

- (1) BO: Enquanto ele preparava o jantar, o João partiu uma chávena.
BN: Enquanto preparava o jantar, o João partiu uma chávena.
FO: O João partiu uma chávena enquanto ele preparava o jantar.
FN: O João partiu uma chávena enquanto preparava o jantar.
Quem é que preparava o jantar? A. o João B. uma outra pessoa
- (2) BO: Enquanto ela comia o bolo, a Maria apagou a luz.
BN: Enquanto comia o bolo, a Maria apagou a luz.
FO: A Maria apagou a luz enquanto ela comia o bolo.
FN: A Maria apagou a luz enquanto comia o bolo.
Quem é que comia o bolo? A. a Maria B. uma outra pessoa
- (3) BO: Enquanto ele fazia o trabalho, o Paulo desligou a televisão.
BN: Enquanto fazia o trabalho, o Paulo desligou a televisão.
FO: O Paulo desligou a televisão enquanto ele fazia o trabalho.
FN: O Paulo desligou a televisão enquanto fazia o trabalho.
Quem é que fazia o trabalho? A. o Paulo B. uma outra pessoa
- (4) BO: Enquanto ela procurava a chave, a Rita atendeu o telefone.
BN: Enquanto procurava a chave, a Rita atendeu o telefone.
FO: A Rita atendeu o telefone enquanto ela procurava a chave.
FN: A Rita atendeu o telefone enquanto procurava a chave.
Quem é que procurava a chave? A. a Rita B. uma outra pessoa
- (5) BO: Enquanto ele bebia a cerveja, o Carlos leu a ementa.
BN: Enquanto bebia a cerveja, o Carlos leu a ementa.
FO: O Carlos leu a ementa enquanto ele bebia a cerveja.
FN: O Carlos leu a ementa enquanto bebia a cerveja.
Quem é que bebia a cerveja? A. o Carlos B. uma outra pessoa
- (6) BO: Enquanto ela ouvia música, a Sara escreveu uma reportagem.
BN: Enquanto ouvia música, a Sara escreveu uma reportagem.
FO: A Sara escreveu uma reportagem enquanto ela ouvia música.
FN: A Sara escreveu uma reportagem enquanto ouvia música.
Quem é que ouvia música ? A. a Sara B. uma outra pessoa

- (7) BO: Enquanto ele esperava o autocarro, o Hugo limpou os óculos.
BN: Enquanto esperava o autocarro, o Hugo limpou os óculos.
FO: O Hugo limpou os óculos enquanto ele esperava o autocarro.
FN: O Hugo limpou os óculos enquanto esperava o autocarro.
Quem é que esperava o autocarro? A. o Hugo B. uma outra pessoa
- (8) BO: Enquanto ela via as notícias, a Laura bebeu um café.
BN: Enquanto via as notícias, a Laura bebeu um café.
FO: A Laura bebeu um café enquanto ela via as notícias.
FN: A Laura bebeu um café enquanto via as notícias.
Quem é que via as notícias? A. a Laura B. uma outra pessoa
- (9) BO: Enquanto ele escrevia o diário, o Pedro ouviu as notícias.
BN: Enquanto escrevia o diário, o Pedro ouviu as notícias.
FO: O Pedro ouviu as notícias enquanto ele escrevia o diário.
FN: O Pedro ouviu as notícias enquanto escrevia o diário.
Quem é que escrevia o diário? A. o Pedro B. uma outra pessoa
- (10) BO: Enquanto ela atravessava a avenida, a Helena fumou um cigarro.
BN: Enquanto atravessava a avenida, a Helena fumou um cigarro.
FO: A Helena fumou um cigarro enquanto ela atravessava a avenida.
FN: A Helena fumou um cigarro enquanto atravessava a avenida.
Quem é que atravessava a avenida? A. a Helena B. uma outra pessoa
- (11) BO: Enquanto ele andava de bicicleta, o Marcos viu um avião.
BN: Enquanto andava de bicicleta, o Marcos viu um avião.
FO: O Marcos viu um avião enquanto ele andava de bicicleta.
FN: O Marcos viu um avião enquanto andava de bicicleta.
Quem é que andava de bicicleta? A. o Marcos B. uma outra pessoa
- (12) BO: Enquanto ela conduzia o carro, a Joana contou uma história.
BN: Enquanto conduzia o carro, a Joana contou uma história.
FO: A Joana contou uma história enquanto ela conduzia o carro.
FN: A Joana contou uma história enquanto conduzia o carro.
Quem é que conduzia o carro? A. a Joana B. uma outra pessoa
- (13) BO: Enquanto ele limpava o vidro, o Diogo cantou uma canção.
BN: Enquanto limpava o vidro, o Diogo cantou uma canção.
FO: O Diogo cantou uma canção enquanto ele limpava o vidro.
FN: O Diogo cantou uma canção enquanto limpava o vidro.
Quem é que limpava o vidro? A. o Diogo B. uma outra pessoa
- (14) BO: Enquanto ela lavava a loiça, a Susana planeou uma viagem.

- BN: Enquanto lavava a loiça, a Susana planeou uma viagem.
FO: A Susana planeou uma viagem enquanto ela lavava a loiça.
FN: A Susana planeou uma viagem enquanto lavava a loiça.
Quem é que lavava a loiça? A. a Susana B. uma outra pessoa
- (15) BO: Enquanto ele copiava os dados, o Nuno abriu um livro.
BN: Enquanto copiava os dados, o Nuno abriu um livro.
FO: O Nuno abriu um livro enquanto ele copiava os dados.
FN: O Nuno abriu um livro enquanto copiava os dados.
Quem é que copiava os dados? A. o Nuno B. uma outra pessoa
- (16) BO: Enquanto ela comprava os legumes, a Paula teve uma ideia.
BN: Enquanto comprava os legumes, a Paula teve uma ideia.
FO: A Paula teve uma ideia enquanto ela comprava os legumes.
FN: A Paula teve uma ideia enquanto comprava os legumes.
Quem é que comprava os legumes ? A. a Paula B. uma outra pessoa
- (17) BO: Enquanto ele jogava computador, o Bruno comeu as batatas fritas.
BN: Enquanto jogava computador, o Bruno comeu as batatas fritas.
FO: O Bruno comeu as batatas fritas enquanto ele jogava computador.
FN: O Bruno comeu as batatas fritas enquanto jogava computador.
Quem é que jogava computador? A. o Bruno B. uma outra pessoa
- (18) BO: Enquanto ela lia o jornal, a Luísa pediu um croissant.
BN: Enquanto lia o jornal, a Luísa pediu um croissant.
FO: A Luísa pediu um croissant enquanto ela lia o jornal.
FN: A Luísa pediu um croissant enquanto lia o jornal.
Quem é que lia o jornal? A. a Luísa B. uma outra pessoa
- (19) BO: Enquanto ele abria a porta, o Ricardo recebeu uma mensagen.
BN: Enquanto abria a porta, o Ricardo recebeu uma mensagen.
FO: O Ricardo recebeu uma mensagen enquanto ele abria a porta.
FN: O Ricardo recebeu uma mensagen enquanto abria a porta.
Quem é que abria a porta? A. o Ricardo B. uma outra pessoa
- (20) BO Enquanto ela pintava um quadro, a Teresa fechou a janela.
BN Enquanto pintava um quadro, a Teresa fechou a janela.
FO A Teresa fechou a janela enquanto ela pintava um quadro.
FN A Teresa fechou a janela enquanto pintava um quadro.
Quem é que pintava um quadro? A. a Teresa B. uma outra pessoa
- (21) BO: Enquanto ele cortava o melão, o Jorge contou uma piada.
BN: Enquanto cortava o melão, o Jorge contou uma piada.
FO: O Jorge contou uma piada enquanto ele cortava o melão.

FN: O Jorge contou uma piada enquanto cortava o melão.
Quem é que cortava o melão? A. o Jorge B. uma outra pessoa

(22) BO: Enquanto ela secava a roupa, a Marta tomou uma decisão.
BN: Enquanto secava a roupa, a Marta tomou uma decisão.
FO: A Marta tomou uma decisão enquanto ela secava a roupa.
FN: A Marta tomou uma decisão enquanto secava a roupa.
Quem é que secava a roupa? A. a Marta B. uma outra pessoa

(23) BO: Enquanto ele imprimia o documento, o Rui tirou uma bolacha.
BN: Enquanto imprimia o documento, o Rui tirou uma bolacha.
FO: O Rui tirou uma bolacha enquanto ele imprimia o documento.
FN: O Rui tirou uma bolacha enquanto imprimia o documento.
Quem é que imprimia o documento? A. o Rui B. uma outra pessoa

(24) BO: Enquanto ela descia as escadas, a Diana chamou um táxi.
BN: Enquanto descia as escadas, a Diana chamou um táxi.
FO: A Diana chamou um táxi enquanto ela descia as escadas.
FN: A Diana chamou um táxi enquanto descia as escadas.
Quem é que descia as escadas? A. a Diana B. uma outra pessoa

Appendix 4 Testing items for the self-paced reading task

Legend:

- OM = Overt-Match
- NM = Null-Match
- OMM = Overt-Mismatch
- NMM = Null-Mismatch

Part 1 Forward anaphora conditions

- (1) OM: Quando/o Mário/ligou/a televisão,/ele ficou perturbado/com a notícia/de última hora.
 NM: Quando/o Mário/ligou/a televisão,/ficou perturbado/com a notícia/de última hora.
 OMM: Quando/o Mário/ligou/a televisão,/ela ficou perturbada/com a notícia/de última hora.
 NMM: Quando/o Mário/ligou/a televisão,/ficou perturbada/com a notícia/de última hora.
 A. O Mário ficou perturbado com a notícia de última hora.
 B. Uma outra pessoa ficou perturbado com a notícia de última hora.
- (2) OM: Quando/o João/recebeu /o prato,/ele ficou aliviado/com a comida/servida.
 NM: Quando/o João/recebeu /o prato,/ficou aliviado/com a comida/servida.
 OMM: Quando/o João/recebeu /o prato,/ela ficou aliviada/com a comida/servida.
 NMM: Quando/o João/recebeu /o prato,/ficou aliviada/com a comida/servida.
 A. O João ficou aliviado com a comida servida.
 B. Uma outra pessoa ficou aliviada com a comida servida.
- (3) OM: Quando/a Diana/atendeu /o telefone,/ela ficou admirada/com a decisão/do diretor.
 NM: Quando/a Diana/atendeu /o telefone,/ficou admirada/com a decisão/do diretor.
 OMM: Quando/a Diana/atendeu /o telefone,/ele ficou admirado/com a decisão/do diretor.
 NMM: Quando/a Diana/atendeu /o telefone,/ficou admirado/com a decisão/do diretor.
 A. A Diana ficou admirada com a decisão do diretor.
 B. Uma outra pessoa ficou admirada com a decisão do diretor.
- (4) OM: Quando/a Maria/fez/o discurso,/ela ficou animada/com as ideias/inovadoras.
 NM: Quando/a Maria/fez/o discurso,/ficou animada/com as ideias/inovadoras.
 OMM: Quando/a Maria/fez/o discurso,/ele ficou animado/com as ideias/inovadoras.
 NMM: Quando/a Maria/fez/o discurso,/ficou animado/com as ideias/inovadoras.
 A. A Maria ficou animada com as ideias inovadoras.
 B. Uma outra pessoa ficou animada com as ideias inovadoras.
- (5) OM: Quando/a Rita/contou/a história,/ela ficou relaxada/com a reação/dos amigos.
 NM: Quando/a Rita/contou/a história,/ficou relaxada/com a reação/dos amigos.
 OMM: Quando/a Rita/contou/a história,/ele ficou relaxado/com a reação/dos amigos.
 NMM: Quando/a Rita/contou/a história,/ficou relaxado/com a reação/dos amigos.
 A. A Rita ficou relaxada com a reação dos amigos.

- B. Uma outra pessoa ficou relaxada com a reação dos amigos.
- (6) OM: Quando/o Rui/arranjou/a janela,/ele ficou ocupado/com o trabalho/da decoração.
NM: Quando/o Rui/arranjou/a janela,/ficou ocupado/com o trabalho/da decoração.
OMM: Quando/o Rui/arranjou/a janela,/ela ficou ocupada/com o trabalho/da decoração.
NMM: Quando/o Rui/arranjou/a janela,/ficou ocupada/com o trabalho/da decoração.
A. O Rui ficou ocupado com o trabalho da decoração.
B. Uma outra pessoa ficou ocupada com o trabalho da decoração.
- (7) OM: Quando/o Hugo/parou/o carro,/ele ficou irritado/com o barulho/do motor.
NM: Quando/o Hugo/parou/o carro,/ficou irritado/com o barulho/do motor.
OMM: Quando/o Hugo/parou/o carro,/ela ficou irritada/com o barulho/do motor.
NMM: Quando/o Hugo/parou/o carro,/ficou irritada/com o barulho/do motor.
A. O Hugo ficou irritado com o barulho do motor.
B. Uma outra pessoa ficou irritada com o barulho do motor.
- (8) OM: Quando/a Carla/encomendou/um bolo,/ela ficou curiosa/com o recheio/escolhido.
NM: Quando/a Carla/encomendou/um bolo,/ficou curiosa/com o recheio/escolhido.
OMM: Quando/a Carla/encomendou/um bolo,/ele ficou curioso/com o recheio/escolhido.
NMM: Quando/a Carla/encomendou/um bolo,/ficou curioso/com o recheio/escolhido.
A. A Carla ficou curiosa com o recheio escolhido.
B. Uma outra pessoa ficou curiosa com o recheio escolhido.
- (9) OM: Quando/a Sara/alimentou/os peixes,/ela ficou chateada/com o cheiro/da comida.
NM: Quando/a Sara/alimentou/os peixes,/ficou chateada/com o cheiro/da comida.
OMM: Quando/a Sara/alimentou/os peixes,/ele ficou chateado/com o cheiro/da comida.
NMM: Quando/a Sara/alimentou/os peixes,/ficou chateado/com o cheiro/da comida.
A. A Sara ficou chateada com o cheiro da comida.
B. Uma outra pessoa ficou chateada com o cheiro da comida.
- (10) OM: Quando/o Tiago/submeteu/o artigo,/ele ficou ansioso/pela resposta/dos editores.
NM: Quando/o Tiago/submeteu/o artigo,/ficou ansioso/pela resposta/dos editores.
OMM: Quando/o Tiago/submeteu/o artigo,/ela ficou ansiosa/pela resposta/dos editores.
NMM: Quando/o Tiago/submeteu/o artigo,/ficou ansiosa/pela resposta/dos editores.
A. O Tiago ficou ansioso pela resposta dos editores.
B. Uma outra pessoa ficou ansiosa pela resposta dos editores.
- (11) OM: Quando/a Paula/corrigiu /o trabalho,/ela ficou nervosa/com os erros/evitáveis.
NM: Quando/a Paula/corrigiu /o trabalho,/ficou nervosa/com os erros/evitáveis.
OMM: Quando/a Paula/corrigiu /o trabalho,/ele ficou nervoso/com os erros/evitáveis.
NMM: Quando/a Paula/corrigiu /o trabalho,/ficou nervoso/com os erros/evitáveis.
A. Uma outra pessoa ficou nervosa com os erros evitáveis.
B. A Paula ficou nervosa com os erros evitáveis.

- (12)OM: Quando/o André/consultou/o horário,/ele ficou inquieto/com as horas/de partida.
NM: Quando/o André/consultou/o horário,/ficou inquieto/com as horas/de partida.
OMM: Quando/o André/consultou/o horário,/ela ficou inquieta/com as horas/de partida.
NMM: Quando/o André/consultou/o horário,/ficou inquieta/com as horas/de partida.
A. Uma outra pessoa ficou inquieta com as horas de partida.
B. O André ficou inquieto com as horas de partida.
- (13)OM: Quando/a Sofia/encontrou/os livros,/ela ficou zangada/com os riscos/em várias páginas.
NM: Quando/a Sofia/encontrou/os livros,/ficou zangada/com os riscos/em várias páginas.
OMM: Quando/a Sofia/encontrou/os livros,/ele ficou zangado/com os riscos/em várias páginas.
NMM: Quando/a Sofia/encontrou/os livros,/ficou zangado/com os riscos/em várias páginas.
A. Uma outra pessoa ficou zangada com os riscos em várias páginas.
B. A Sofia ficou zangada com os riscos em várias páginas.
- (14)OM: Quando/o Jaime/tocou/a guitarra,/ele ficou furioso/com a sugestão/do vizinho.
NM: Quando/o Jaime/tocou/a guitarra,/ficou furioso/com a sugestão/do vizinho.
OMM: Quando/o Jaime/tocou/a guitarra,/ela ficou furiosa/com a sugestão/do vizinho.
NMM: Quando/o Jaime/tocou/a guitarra,/ficou furiosa/com a sugestão/do vizinho.
A. Uma outra pessoa ficou furiosa com a sugestão do vizinho.
B. O Jaime ficou furioso com a sugestão do vizinho.
- (15)OM: Quando/o Paulo/filmou/as paisagens,/ele ficou surpreso/com as flores/coloridas.
NM: Quando/o Paulo/filmou/as paisagens,/ficou surpreso/com as flores/coloridas.
OMM: Quando/o Paulo/filmou/as paisagens,/ela ficou surpresa/com as flores/coloridas.
NMM: Quando/o Paulo/filmou/as paisagens,/ficou surpresa/com as flores/coloridas.
A. Uma outra pessoa ficou surpresa com as flores coloridas.
B. O Paulo ficou surpreso com as flores coloridas.
- (16)OM: Quando/a Marta/começou/as férias,/ela ficou cansada/com a viagem/de avião.
NM: Quando/a Marta/começou/as férias,/ficou cansada/com a viagem/de avião.
OMM: Quando/a Marta/começou/as férias,/ele ficou cansado/com a viagem/de avião.
NMM: Quando/a Marta/começou/as férias,/ficou cansado/com a viagem/de avião.
A. Uma outra pessoa ficou cansada com a viagem de avião.
B. A Marta ficou cansada com a viagem de avião.
- (17)OM: Quando/a Luísa/atirou/a bola,/ela ficou escondida/atrás do banco/do jardim.
NM: Quando/a Luísa/atirou/a bola,/ficou escondida/atrás do banco/do jardim.
OMM: Quando/a Luísa/atirou/a bola,/ele ficou escondido/atrás do banco/do jardim.
NMM: Quando/a Luísa/atirou/a bola,/ficou escondido/atrás do banco/do jardim.
A. Uma outra pessoa ficou escondida atrás do banco do jardim.
B. A Luísa ficou escondida atrás do banco do jardim.

- (18) OM: Quando/o Diogo/fechou/a porta,/ele ficou trancado/fora de casa/durante a tarde.
NM: Quando/o Diogo/fechou/a porta,/ficou trancado/fora de casa/durante a tarde.
OMM: Quando/o Diogo/fechou/a porta,/ela ficou trancada/fora de casa/durante a tarde.
NMM: Quando/o Diogo/fechou/a porta,/ficou trancada/fora de casa/durante a tarde.
A. Uma outra pessoa ficou trancada fora de casa durante a tarde.
B. O Diogo ficou trancado fora de casa durante a tarde.
- (19) OM: Quando/o Marco/conseguiu/o emprego, /ele ficou afastado/dos familiares/temporariamente.
NM: Quando/o Marco/conseguiu/o emprego, /ficou afastado/dos familiares/temporariamente.
OMM: Quando/o Marco/conseguiu/o emprego, /ela ficou afastada/dos familiares/temporariamente.
NMM: Quando/o Marco/conseguiu/o emprego, /ficou afastada/dos familiares/temporariamente.
A. Uma outra pessoa ficou afastada dos familiares temporariamente.
B. O Marco ficou afastado dos familiares temporariamente.
- (20) OM: Quando/a Joana/abriu/a prenda,/ela ficou encantada/com a beleza/do relógio.
NM: Quando/a Joana/abriu/a prenda,/ficou encantada/com a beleza/do relógio.
OMM: Quando/a Joana/abriu/a prenda,/ele ficou encantado/com a beleza/do relógio.
NMM: Quando/a Joana/abriu/a prenda,/ficou encantado/com a beleza/do relógio.
A. Uma outra pessoa ficou encantada com a beleza do relógio.
B. A Joana ficou encantada com a beleza do relógio.

Part 2 Backward anaphora conditions

- (1) OM: Enquanto/ele estava/deitado/em cima da cama,/o Manuel/cantou/uma canção/relaxante.
NM: Enquanto/estava/deitado/em cima da cama,/o Manuel/cantou/uma canção/relaxante.
OMM: Enquanto/ela estava/deitada/em cima da cama,/o Manuel/cantou/uma canção/relaxante.
NMM: Enquanto/estava/deitada/em cima da cama,/o Manuel/cantou/uma canção/relaxante.
A. O Manuel estava deitado em cima da cama.
B. Uma outra pessoa estava deitada em cima da cama.
- (2) OM: Enquanto/ela estava/sentada/ao balcão do bar,/a Sandra/chamou/um amigo/com urgência.
NM: Enquanto/estava/sentada/ao balcão do bar,/a Sandra/chamou/um amigo/com urgência.
OMM: Enquanto/ele estava/sentado/ao balcão do bar,/a Sandra/chamou/um amigo/com urgência.
NMM: Enquanto/estava/sentado/ao balcão do bar,/a Sandra/chamou/um amigo/com urgência.
A. A Sandra estava sentada ao balcão do bar.

- B. Uma outra pessoa estava sentada ao balcão do bar.
- (3) OM: Enquanto/ele esteve/internado/no hospital,/o Carlos/enviou/o projeto/científico.
NM: Enquanto/esteve/internado/no hospital,/o Carlos/enviou/o projeto/científico.
OMM: Enquanto/ela esteve/internada/no hospital,/o Carlos/enviou/o projeto/científico.
NMM: Enquanto/esteve/internada/no hospital,/o Carlos/enviou/o projeto/científico.
A. O Carlos esteve internado no hospital.
B. Uma outra pessoa esteve internada no hospital.
- (4) OM: Enquanto/ela estava/trancada/fora de casa,/a Marisa/apanhou/uma flor/perfumada.
NM: Enquanto/estava/trancada/fora de casa,/a Marisa/apanhou/uma flor/perfumada.
OMM: Enquanto/ele estava/trancado/fora de casa,/a Marisa/apanhou/uma flor/perfumada.
NMM: Enquanto/estava/trancado/fora de casa,/a Marisa/apanhou/uma flor/perfumada.
A. A Marisa estava trancada fora de casa.
B. Uma outra pessoa estava trancada fora de casa.
- (5) OM: Enquanto/ele estava/escondido/debaixo da mesa,/o Filipe/mandou/um vídeo/divertido.
NM: Enquanto/estava/escondido/debaixo da mesa,/o Filipe/mandou/um vídeo/divertido.
OMM: Enquanto/ela estava/escondida/debaixo da mesa,/o Filipe/mandou/um vídeo/divertido.
NMM: Enquanto/estava/escondida/debaixo da mesa,/o Filipe/mandou/um vídeo/divertido.
A. O Filipe estava escondido debaixo da mesa.
B. Uma outra pessoa estava escondida debaixo da mesa.
- (6) OM: Enquanto/ela esteve/adoentada/durante as férias,/a Isabel/acabou/o trabalho/de matemática.
NM: Enquanto/esteve/adoentada/durante as férias,/a Isabel/acabou/o trabalho/de matemática.
OMM: Enquanto/ele esteve/adoentado/durante as férias,/a Isabel/acabou/o trabalho/de matemática.
NMM: Enquanto/esteve/adoentado/durante as férias,/a Isabel/acabou/o trabalho/de matemática.
A. A Isabel esteve adoentada durante as férias.
B. Uma outra pessoa esteve adoentada durante as férias.
- (7) OM: Enquanto/ele estava/encostado/ao muro do castelo,/o Marcos/colocou/os óculos/cuidadosamente.
NM: Enquanto/estava/encostado/ao muro do castelo,/o Marcos/colocou/os óculos/cuidadosamente.
OMM: Enquanto/ela estava/encostada/ao muro do castelo,/o Marcos/colocou/os óculos/cuidadosamente.
NMM: Enquanto/estava/encostada/ao muro do castelo,/o Marcos/colocou/os óculos/cuidadosamente.
A. O Marcos estava encostado ao muro do castelo.

B. Uma outra pessoa estava encostada ao muro do castelo.

(8) OM: Enquanto/ela estava/virada/para o restaurante,/a Soraia/apagou/o cigarro/tranquilamente.

NM: Enquanto/estava/virada/para o restaurante,/a Soraia/apagou/o cigarro/tranquilamente.

OMM: Enquanto/ele estava/virado/para o restaurante,/a Soraia/apagou/o cigarro/tranquilamente.

NMM: Enquanto/estava/virado/para o restaurante,/a Soraia/apagou/o cigarro/tranquilamente.

A. A Soraia estava virada para o restaurante.

B. Uma outra pessoa estava virada para o restaurante.

(9) OM: Enquanto/ele estava/apoiado/na parede,/o Miguel/escreveu/um email/no telemóvel.

NM: Enquanto/estava/apoiado/na parede,/o Miguel/escreveu/um email/no telemóvel.

OMM: Enquanto/ela estava/apoiada/na parede,/o Miguel/escreveu/um email/no telemóvel.

NMM: Enquanto/estava/apoiado/na parede,/o Miguel/escreveu/um email/no telemóvel.

A. O Miguel estava apoiado na parede.

B. Uma outra pessoa estava apoiada na parede.

(10) OM: Enquanto/ela esteve/desocupada/no refeitório,/a Eunice/visitou/a cozinha/relaxadamente.

NM: Enquanto/esteve/desocupada/no refeitório,/a Eunice/visitou/a cozinha/relaxadamente.

OMM: Enquanto/ele esteve/desocupado/no refeitório,/a Eunice/visitou/a cozinha/relaxadamente.

NMM: Enquanto/esteve/desocupado/no refeitório,/a Eunice/visitou/a cozinha/relaxadamente.

A. A Eunice esteve desocupada no refeitório.

B. Uma outra pessoa esteve desocupada no refeitório.

(11) OM: Enquanto/ele esteve/dedicado/à análise dos dados,/o Rodrigo/entregou/o artigo/ao diretor.

NM: Enquanto/esteve/dedicado/à análise dos dados,/o Rodrigo/entregou/o artigo/ao diretor.

OMM: Enquanto/ela esteve/dedicada/à análise dos dados,/o Rodrigo/entregou/o artigo/ao diretor.

NMM: Enquanto/esteve/dedicada/à análise dos dados,/o Rodrigo/entregou/o artigo/ao diretor.

A. Uma outra pessoa esteve dedicada à análise dos dados.

B. O Rodrigo esteve dedicado à análise dos dados.

(12) OM: Enquanto/ela esteve/fechada/no quarto,/a Cláudia/limpou/a estante/com um pano.

NM: Enquanto/esteve/fechada/no quarto,/a Cláudia/limpou/a estante/com um pano.

OMM: Enquanto/ele esteve/fechado/no quarto,/a Cláudia/limpou/a estante/com um pano.

NMM: Enquanto/esteve/fechado/no quarto,/a Cláudia/limpou/a estante/com um pano.

A. Uma outra pessoa esteve fechada no quarto.

B. A Cláudia esteve fechada no quarto.

(13) OM: Enquanto/ele esteve/constipado/durante a viagem,/o Gustavo/terminou/um romance/aborrecido.

NM: Enquanto/esteve/constipado/durante a viagem,/o Gustavo/terminou/um romance/aborrecido.

OMM: Enquanto/ela esteve/constipada/durante a viagem,/o Gustavo/terminou/um romance/aborrecido.

NMM: Enquanto/esteve/constipada/durante a viagem,/o Gustavo/terminou/um romance/aborrecido.

A. Uma outra pessoa esteve constipada durante a viagem.

B. O Gustavo esteve constipado durante a viagem.

(14) OM: Enquanto/ela esteve/magoada/no joelho,/a Manuela/planeou/uma viagem/ao Japão.

NM: Enquanto/esteve/magoada/no joelho,/a Manuela/planeou/uma viagem/ao Japão.

OMM: Enquanto/ele esteve/magoado/no joelho,/a Manuela/planeou/uma viagem/ao Japão.

NMM: Enquanto/esteve/magoado/no joelho,/a Manuela/planeou/uma viagem/ao Japão.

A. Uma outra pessoa esteve magoada no joelho.

B. A Manuela esteve magoada no joelho.

(15) OM: Enquanto/ele estava/concentrado/no jogo de futebol,/o Eduardo/despiu/o casaco/lentamente.

NM: Enquanto/estava/concentrado/no jogo de futebol,/o Eduardo/despiu/o casaco/lentamente.

OMM: Enquanto/ela estava/concentrada/no jogo de futebol,/o Eduardo/despiu/o casaco/lentamente.

NMM: Enquanto/estava/concentrada/no jogo de futebol,/o Eduardo/despiu/o casaco/lentamente.

A. Uma outra pessoa estava concentrada no jogo de futebol.

B. O Eduardo estava concentrado no jogo de futebol.

(16) OM: Enquanto/ela estava/chateada/com a família,/a Mariana/preparou/uma festa/surpresa.

NM: Enquanto/estava/chateada/com a família,/a Mariana/preparou/uma festa/surpresa.

OMM: Enquanto/ele estava/chateado/com a família,/a Mariana/preparou/uma festa/surpresa.

NMM: Enquanto/estava/chateado/com a família,/a Mariana/preparou/uma festa/surpresa.

A. Uma outra pessoa estava chateada com a família.

B. A Mariana estava chateada com a família.

(17) OM: Enquanto/ele estava/distraído/com os filhos,/o Ricardo/fechou/as janelas/calmamente.

NM: Enquanto/estava/distraído/com os filhos,/o Ricardo/fechou/as janelas/calmamente.

OMM: Enquanto/ela estava/distraída/com os filhos,/o Ricardo/fechou/as

janelas/calmamente.

NMM: Enquanto/estava/distraída/com os filhos,/o Ricardo/fechou/as janelas/calmamente.

A. Uma outra pessoa estava distraída com os filhos.

B. O Ricardo estava distraído com os filhos.

(18) OM: Enquanto/ela esteve/suspensa/da associação,/a Adriana/começou/as férias/no estrangeiro.

NM: Enquanto/esteve/suspensa/da associação,/a Adriana/começou/as férias/no estrangeiro.

OMM: Enquanto/ele esteve/suspensa/da associação,/a Adriana/começou/as férias/no estrangeiro.

NMM: Enquanto/esteve/suspensa/da associação,/a Adriana/começou/as férias/no estrangeiro.

A. Uma outra pessoa esteve suspensa da associação.

B. A Adriana esteve suspensa da associação.

(19) OM: Enquanto/ele estava/aborrecido/com o trabalho,/o Gonçalo/arrumou/as compras/no armário.

NM: Enquanto/estava/aborrecido/com o trabalho,/o Gonçalo/arrumou/as compras/no armário.

OMM: Enquanto/ela estava/aborrecida/com o trabalho,/o Gonçalo/arrumou/as compras/no armário.

NMM: Enquanto/estava/aborrecida/com o trabalho,/o Gonçalo/arrumou/as compras/no armário.

A. Uma outra pessoa estava aborrecida com o trabalho.

B. O Gonçalo estava aborrecido com o trabalho.

(20) OM: Enquanto/ela estava/envolvida/no novo trabalho,/a Daniela/desligou/o telefone/inesperadamente.

NM: Enquanto/estava/envolvida/no novo trabalho,/a Daniela/desligou/o telefone/inesperadamente.

OMM: Enquanto/ele estava/envolvido/no novo trabalho,/a Daniela/desligou/o telefone/inesperadamente.

NMM: Enquanto/estava/envolvido/no novo trabalho,/a Daniela/desligou/o telefone/inesperadamente.

A. Uma outra pessoa estava envolvida no novo trabalho.

B. A Daniela estava envolvida no novo trabalho.