UNIVERSIDADE DE LISBOA

FACULDADE DE LETRAS

DEPARTAMENTO DE LINGUÍSTICA GERAL E ROMÂNICA

Acquisition of European Portuguese cleft structures by L1 Mandarin learners

Xinyi Li

Mestrado em Linguística

2021

UNIVERSIDADE DE LISBOA

FACULDADE DE LETRAS

DEPARTAMENTO DE LINGUÍSTICA GERAL E ROMÂNICA

Acquisition of European Portuguese cleft structures by L1 Mandarin learners

Xinyi Li

Thesis supervised by
Prof. Dr. Ana Lúcia Santos (FLUL)
Prof. Dr. Maria Lobo (FCSH NOVA)

Mestrado em Linguística

2021

"Porque os outros se mascaram mas tu não

Porque os outros usam a virtude

para comprar o que não tem perdão

Porque os outros têm medo mas tu não"

Sophia de Mello Breyner Andresen

ACKNOWLEDGEMENTS

Table of Contents

List of tables

List of graphs

Resumo

A presente dissertação debruça-se sobre a aquisição de estruturas clivadas do português europeu (PE) por aprendentes que têm chinês mandarim (CM) como L1 e inglês como L2. Foi realizado um estudo experimental, com base numa tarefa de juízo de aceitabilidade, sobre a aquisição das seguintes estruturas em PE: clivadas de *é que*, clivadas canónicas, semipseudoclivadas e pseudoclivadas; foi ainda testada uma estrutura superficialmente idêntica à clivada canónica mas sem C preenchido, agramatical em PE.

As questões de investigação que guiaram o presente trabalho são: (i) que diferenças estruturais apresentam as clivadas do PE e as do CM? (ii) as diferenças sintáticas levantam obstáculos à aquisição de clivadas do PE por sujeitos que têm o CM como L1? (iii) em caso afirmativo, os alunos podem superar os obstáculos potenciais? (iv) que tipo de clivadas do PE colocam mais dificuldade na aquisição por falantes L1-CM? (v) outras línguas previamente adquiridas pelos alunos influenciarão a aquisição destas estruturas do PE?

Têm sido avançadas propostas diversificadas da sintaxe das estruturas clivadas em PE. Discutindo se todos tipos de clivadas têm por base uma estrutura sintática idêntica, as análises sintáticas prévias dividem-se em dois tipos: análises unificadas (Ambar, 1999; Costa e Duarte, 2001, 2005) e análises não-unificadas (Lobo, 2006; Soares, 2006). No tocante à aquisição da linguagem, atestou-se uma escala de emergência e diferente frequência de diferentes estratégias de clivagem (Lobo, Santos e Soares-Jesel, 2016), o que forneceu argumentos para a análise não-unificada. Em linha com Lobo (2006), Soares (2006) e Lobo, Santos e Soares-Jesel (2016), assume-se que: i) as clivadas canónicas envolvem movimento do constituinte clivado para uma posição periférica associada a um traço [+foco] na oração encaixada; ii) uma estrutura mono-oracional é proposta para as clivadas de *é que*, em que esta expressão lexicaliza uma categoria funcional na periferia esquerda da frase, a que se associa um traço [+foco]; iii) as pseudoclivadas são estruturas identificacionais em que o verbo "ser" seleciona uma Oração Pequena composta por uma oração clivada que é semelhante a uma relativa livre e um constituinte focalizado externo à oração tipo CP; iv) nas semipseudoclivadas, o verbo "ser", reanalisado como um marcador de foco contrastivo, marca a fronteira da periferia esquerda de vP, e o

constituinte clivado está *in situ,* sob escopo do marcador de foco. No que respeita à aquisição de língua não materna, prediz-se que o desenvolvimento da gramática das diferentes estruturas clivadas de aprendentes não ocorre de forma homogénea. Para as estruturas clivadas do CM, comparando as diferentes análises (Huang, 1988, 1998; Simpson e Wu, 2002; Cheng, 2008; Paul e Whitman, 2008; Hole, 2011, etc.), propomos que: i) as pseudoclivadas do CM e do PE partilham na base uma estrutura sintática semelhante, i.e. uma estrutura identificacional; ii) nas clivadas de *shì* simples, está envolvido foco prosódico, não sendo a posição de focalização associada à periferia esquerda ou a uma projeção de foco interna; iii) nas clivadas *shì...de*, à luz de Belletti (2004), o traço [+foco] é associado a uma projeção abaixo de vP, FocP, para a qual sobe o sujeito ou adjunto clivado, e o objeto, quando recebe o foco, fica na posição final de frase, de acordo com a Regra do Acento Nuclear (*Nuclear Stress Rule*), em linha com Cinque (1993). É de notar que, no CM, o objeto não pode ser clivado nas clivadas de *shì* simples ou nas clivadas *shì* V O *de*, ao contrário do que acontece com sujeito. Em suma, o facto de o português dispor de categorias que marcam foco contrastivo e a associação do traço [+foco] a diferentes categorias funcionais nas duas línguas determinam as diferenças entre as clivadas do CM e a sua contrapartida portuguesa.

No que diz respeito à aquisição de língua não materna (ALNM), dado que estes aprendentes geralmente adquiriram inglês como L2 antes do PE, levou-se também em consideração a influência potencial da L2 inglês. À luz da Hipótese de Reconfiguração de Traços (Lardiere, 2008, 2009), assume-se que, para adquirir as estruturas clivadas do PE, um aprendente L1-CM teria de desassociar o traço [+ foco] do Foc^0 baixo no CM e associá-lo a categorias diferentes, mais especificamente, uma categoria funcional na periferia esquerda nas clivadas de *é-que* e nas clivadas canónicas; nas semipseudoclivadas, deve ser associado ao verbo "ser" na fronteira da periferia esquerda de vP; nas pseudoclivadas, assume-se que a configuração do PE é idêntica à do CM, e nesse caso, o simples mapeamento dos traços na L1 e na L2 poderá assegurar a aquisição da gramática-alvo, pelo que não se espera que haja problemas para os aprendentes. No entanto, considera-se também que a aquisição de clivadas canónicas pode ser facilitada pela L2 inglês, conforme predizem os modelos da aquisição de L3 como o Modelo da Melhoria Cumulativa (MMC, e.g., Flynn et al. 2004) e o Modelo de Primazia Tipológica (MPT, e.g., Rothman, 2011).

Para testar a aceitação, pelos falantes de L3, de diferentes estruturas clivadas do PE, foram

aplicadas duas tarefas de juízo de aceitabilidade, um teste centrado em diferentes clivadas de sujeito e objeto (de *é que*, canónicas, estruturas superficialmente idênticas às clivadas canónicas sem C preenchido, semipseudoclivadas e pseudoclivadas), e outro centrado na concordância sujeito-verbo em clivadas com sujeito plural. Foi usada uma escala de Likert com quatro pontos. Os testes foram aplicados a três grupos de falantes do CM (N = 60), subdivididos por nível de proficiência (B1-23; B2-20; C1-17), e a um grupo de controlo de 21 falantes nativos do PE. Adicionalmente, foram aplicados dois testes centrados em clivadas do inglês para examinar a transferência eventual da L2.

A observação global dos resultados mostra que os aprendentes não rejeitam clivadas de objeto. As clivadas canónicas de sujeito e objeto são aceites pelos participantes independentemente da sua proficiência, seguidas pelas pseudoclivadas. Ao contrário, as clivadas de *é que* de sujeito parecem problemáticas para os grupos dos níveis intermédios por contraste com o grupo de controlo; no caso das clivadas de *é que* de objeto, ainda que os aprendentes em geral rejeitassem a estrutura, a surpreendente baixa aceitação por parte do grupo de controlo torna a comparação mais complexa. As estruturas sem C preenchido e as semipseudoclivadas foram uniformemente rejeitadas por todos os aprendentes. Quanto às questões de concordância, analisadas no teste 2, a aceitação de clivadas canónicas sem concordância é sobretudo observada no grupo do nível mais baixo testado, B1. Contudo, a aceitação de pseudoclivadas com concordância não esperada é persistente em todos os níveis, sendo também atestada no grupo de controlo, embora de forma reduzida. Assim, embora os falantes aceitem no teste 1 as clivadas canónicas e pseudoclivadas, não adquiriram completamente a gramática-alvo.

Os resultados apresentados não questionam um possível efeito facilitador da L2, esperado de acordo com modelos como o MMC ou o MPT, uma vez que não se atesta rejeição generalizada de clivadas de objeto nem aceitação de C vazio em clivadas, o que implica que a associação do traço [+foco] à periferia esquerda está adquirida. No caso das clivadas de *é que* e das semipseudoclivadas, explorámos a hipótese de que a dificuldade dos falantes possa ser associar um traço [+foco] à expressão *é que* e a "ser" nas semipseudoclivadas. Quanto à concordância em pseudoclivadas e em clivadas canónicas, a persistência dos problemas pode sugerir uma influência da concordância semântica.

O presente estudo realizou uma análise comparativa da sintaxe das estruturas clivadas do PE e do CM, e evidenciou uma escala de dificuldade na aquisição de L3 de estruturas clivadas do PE por

falantes de L1-CM: clivadas canónicas > pseudoclivadas > clivadas de *é-que* > semipseudoclivadas. Não obstante, as estruturas sintáticas dos vários tipos de estruturas clivadas do PE, CM e inglês e a natureza de tais estruturas na interface sintaxe-discurso requerem uma análise mais aprofundada. A sensibilidade mostrada pelos falantes nativos aos diferentes padrões de clivadas e às diferentes funções sintáticas dos constituintes clivados também despertou a nossa atenção e pode levar a uma investigação mais aprofundada. Além de testar clivadas de sujeito e de objeto, mais testes poderão ser conduzidos sobre a clivagem de outros constituintes com diversas funções sintáticas. Os contrastes nas respostas dos falantes nativos e dos aprendentes no trabalho experimental deixam questões em aberto. Por exemplo, na área do processamento linguístico, seria interessante averiguar se os falantes nativos e os falantes L2 avançados processam certas estruturas distintamente. No que diz respeito a questões mais gerais no âmbito dos estudos em ALNM, embora não seja o objetivo deste estudo corroborar nenhum modelo de SLA, estudos futuros poderiam concentrar-se em discernir os potenciais efeitos de L1 ou/e de L2 na aquisição de L3 da sintaxe do PE.

Palavras-chave: Aquisição de língua não materna, Sintaxe, Estruturas clivadas, Português Europeu, Chinês Mandarim.

Abstract

This dissertation addresses the L3 acquisition of European Portuguese (EP) cleft structures by L1- Mandarin Chinese (MC) learners who acquired English as L2. An experimental study was conducted, adopting the method of acceptability judgement task, on the acquisition of the following structures in EP: *é-que* clefts, standard clefts, semipseudoclefts and pseudoclefts. Additionally, standard-cleft-like structures with empty C, ungrammatical in EP, were also tested.

The research questions that guided this study are: (i) what structural differences do EP clefts and MC clefts display? (ii) are the syntactic differences a source of difficulties for the acquisition of EP clefts by MC speakers? (iii) if so, can these learners overcome the potential obstacles? (iv) what kinds of EP clefts present more difficulty for the learners? (v) will other languages previously acquired by learners influence the acquisition of EP?

Various proposals have been put forward on the syntax of EP cleft structures. Debating whether different EP clefts have the same underlying syntactic configuration, previous syntactic analyses are divided into two types: unified analyses (Ambar, 1999; Costa & Duarte, 2001, 2005) and non-unified analyses (Lobo, 2006; Soares, 2006). L1 acquisition data on the frequency and order of emergence of the different clefting strategies (Lobo, Santos & Soares-Jesel, 2016) provided arguments for a non--unified analysis. In line with Lobo (2006) and Soares (2006), we assume that: i) the standard clefts involve the movement of the clefted constituent to the left periphery of an embedded clause, to which a [+focus] feature is associated; ii) a monoclausal structure is proposed for the *é-que* clefts, in which such an expression lexicalizes a functional category associated with a [+focus] feature in the left periphery of the clause; iii) pseudoclefts involve a small clause that consists of a base-generated focused constituent and an unselected (relative clause) CP; iv) in semipseudoclefts, the verb *ser* "be", reanalyzed as a contrastive focus marker, marks the boundary of the left periphery of vP and the clefted constituent stays in situ. In the case of second language acquisition, it is predicted that the development of the L2 grammar of the different cleft structures does not occur in a homogeneous way.

For the MC cleft structures, after comparing different analyses (Huang, 1988, 1998; Simpson & Wu, 2002; Cheng, 2008; Paul & Whitman, 2008; Hole, 2011), we propose that: i) the pseudoclefts of

MC share a similar underlying structure with EP, i.e. an identificational structure; ii) in bare-*shì* subject clefts, a prosodic focus is at play, and neither the left periphery nor an internal focus projection is projected; iii) in *shì...de* clefts, in light of Belletti (2004), the [+focus] feature is associated to a projection below vP, FocP, whose specifier position is the landing site for the clefted subject or adjunct, and the object bearing the focus stays in the original sentence-final position in light of the Nuclear Stress Rule, in line with Cinque (1993). It is worth noting that, in MC, the object cannot be clefted in bare-*shì* clefts or in *shì* V O *de* clefts, as opposed to the subject. In sum, it is the association of a [+focus] feature to different functional categories that determines the differences between cleft structures in EP and MC.

In terms of second language acquisition (SLA), given that Chinese learners generally have acquired English as L2 before learning EP, the potential influence of L2 English was taken into consideration. In the light of the Feature Reassembly Hypothesis (Lardiere, 2008, 2009), we assume that, to ultimately acquire the EP cleft structures, apart from the mastery of the lexical items, an L1-MC learner would have to untangle the [+focus] feature from the low Foc^0 in MC and reassembly it onto different categories, i.e., a functional category in the left periphery in *é-que* clefts and standard clefts; in semipseudoclefts, it should be attached to the verb *ser* "be"; in pseudoclefts, a learner would be expected to exhibit apparent knowledge of the target grammar at the mapping stage, resulting from mapping of MC pseudoclefts. The acquisition of standard clefts may be facilitated due to L2 English, as predicted by L3 acquisition models such as the Cumulative-Enhancement Model (CEM, e.g., Flynn et al. 2004) and the Typological Primacy Model (TPM, e.g., Rothman, 2011).

To test the acceptance by L3 learners of different clefts in EP, two acceptability judgment tasks were applied, one centered on different types of subject and object clefts (*é-que* clefts, standard clefts, standard-cleft-like structures with empty C, semipseudoclefts and pseudoclefts), and the other centered on the subject-verb agreement patterns in clefts when the clefted subject is plural. A four-point Likert scale was used. The tests were applied to three groups of MC speakers (N = 60), subdivided by proficiency level (B1-23; B2-20; C1-17), and a control group of 21 native EP speakers. The tested learners later participated in English tests to examine the possibility of L2 transfer.

General observation of the results shows that the learners do not reject object clefts. Subject and object standard clefts are accepted by the participants regardless of their proficiency, followed by pseudoclefts. On the contrary, subject *é-que* clefts seem problematic for the groups at intermediate proficiency levels when compared with the control group; in the case of object *é-que* clefts, even though learners in general rejected the structure, the surprisingly low acceptance by the control group makes the comparison more complicated. Standard-cleft-like structures with an empty C and semipseudoclefts were uniformly rejected by all learners. As for the questions concerning agreement patterns, which were analyzed in test 2, the acceptance of standard clefts without agreement is mainly observed in the group at the lowest proficiency level tested, B1. However, the acceptance of pseudoclefts with unexpected agreement is persistent at all levels, and is also attested in the control group, although at a reduced scale. Thus, although in test 1 the learners accepted standard clefts and pseudoclefts, they did not completely acquire the target grammar.

The results presented do not question a possible facilitating effect of L2, since there is no evidence of generalized rejection of object clefts or acceptance of empty C in clefts, which implies that the association of the [+focus] feature to the left periphery is acquired. In the case of *é-que* clefts and semipseudoclefts, we explored an explanation for the difficulty faced by the learners and we suggest that it lies in the association of a [+focus] feature to the expression *é que*, and to *ser* "be" in the semipseudoclefts. As for the agreement patterns in standard clefts and pseudoclefts, the persistence of the problems may suggest an influence of semantic agreement.

The study conducted a comparative analysis of the syntax of EP and MC cleft structures, suggesting a scale of difficulty concerning the L3 acquisition of EP cleft structures by L1-MC learners: standard cleft > pseudocleft > *é-que* cleft > semipseudocleft. Notwithstanding, the syntactic structures of various clefting strategies in EP, MC and English and the nature of such structures at the syntax-discourse interface require further analysis. The native speakers' sensitivity to the different clefting patterns considering the different syntactic functions of the clefted constituents also captured our attention and can lead to more profound investigation of these structures. Apart from testing the subject and the object, more testing could be conducted on other clefted constituents with diverse syntactic functions. The contrasts shown in the responses of the native speakers and the learners in the

experiment allow to formulate new research questions. For instance, in the field of linguistic processing, it would be relevant to examine whether the native speakers and the advanced L2 speakers process certain structures distinctly. As regards SLA, although it was not the goal of this work to corroborate any SLA model, future studies could focus on discerning the potential effects of the L1 or/and the L2 on the L3 acquisition of EP syntax.

Keywords: Second language acquisition, Syntax, Cleft structures, European Portuguese, Mandarin Chinese.

Chapter 1 Introduction

In this chapter, we introduce the main aims and objectives of the present study, providing clarification and delimitation of the scope of the current research, critical background information for a comprehension of the particularity of this study, and finally, we present the structure of the dissertation.

1.1 Aims and objectives

Clefting strategies vary crosslinguistically. As clefting strategies express focus through syntactic reordering, the study of clefts can also provide an insight into the syntax-semantics/pragmatics interface and the syntax-discourse interface, among others. In this work, we will only focus on the properties of different types of cleft structures in the syntactic domain. However, it is important to explain that, following authors that we cite in the literature review, we do not distinguish the notion of exhaustivity from that of exclusiveness when it comes to the semantic information that clefting structures encode.

Mandarin Chinese (henceforth, MC) differs typologically from European Portuguese (henceforth, EP), and displays divergent clefting strategies which face many syntactic and semantic constraints that are absent in clefts in some other languages, such as English and Romance languages. Meanwhile, EP exhibits a wide variety of clefting strategies, which involve different structures and movements, such as an operator-variable chain, the activation of the left periphery and high functional projections. The crosslinguistic differences certainly set challenges for second language acquisition. Although, since the 1970s, the syntax of MC clefts has been studied by many scholars, and the acquisition of MC *shì...de* clefts by English native speakers was explored (Mai & Yuan, 2016), to the best of my knowledge, there is a research gap on the interlanguage development and acquisition of such structures in Portuguese by L1-MC learners.

Thus, we dedicate the present work to the acquisition of EP clefts by L1-MC speakers, trying to answer the following questions, which, along the present study, will justify specific working hypotheses in Chapter 3:

(i) what structural differences do EP clefts and MC clefts display?

(ii) are the syntactic differences a source of difficulties for the acquisition of EP clefts by MC speakers?

(iii) if so, can these learners overcome the potential obstacles?

(iv) what kinds of EP clefts present more difficulty for the learners?

(v) Will other languages previously acquired by learners influence the acquisition of EP?

1.2 A General Typological Description of Mandarin Chinese

Worldwide grammarians have categorized the Chinese language family as an independent branch of the Sino-Tibetan language family. On account of the large population and vast geographical area of China, numerous varieties of Chinese exist and are classified mainly by their phonetic divergence into seven major dialect groups, whereby Mandarin, the standard variety, occupies the predominant position by covering approximately 70% of the population (Li & Thompson, 1981)[1]. In this work, to ensure that our description of MC was consensual, we conducted an acceptability judgment task to a group of MC native speakers. The official romanization system *pīnyīn* is used in this work to transcribe Chinese characters into the Roman alphabet using diacritic marks to denote tones. Hereupon, a general typological description based on the work of Li and Thompson (1981) is presented.

MC as well as other varieties of Chinese languages is an analytic language without inflectional morphemes, but with derivational morphemes, compound words, little morphological complexity and a low ratio of morphemes per word. Case, number, and agreement markers are generally absent, and their functions are assumed by word order, particles, prepositions or some affixes, as exemplified in (1).

(1) Péngyǒu-men   zuótiān   zài   wǒ   jiā   hē-le   jiǔ.

      friend-PL   yesterday LOC 1SG home drink -PFV wine

---

[1] Mandarin, referred to as *pǔtōnghuà* in mainland China, Hong Kong and Macau and *guóyǔ* in Taiwan, was proclaimed as the national language by the government of the People's Republic of China in 1955. Based on the speech of Beijing, the national language embodies "the pronunciation of the Beijing dialect, the grammar of northern Mandarin and the vocabulary of modern vernacular literature" (Li & Thompson, 1981). However, variations of "Mandarin" can exist among individuals as a result of the unavoidable influence imposed by different dialects, although the divergence has been reduced over the years due to general education, mass media, along with other factors.

'Some friends drank wine in my home yesterday.'

As opposed to English, MC is termed as a topic-prominent language. Depending on discourse factors, subjects and objects in a topic position can be omitted, resulting in "topic-drop", as demonstrated in (2a,b). Chinese is therefore classified as a discourse-oriented zero-topic language (Huang, 1991).

(2) a. Jiǔ,　　tā-men　　hē-le.

　　wine　　3-PL drink-PFV

　　'As for wine, they drank it.'

　　b. (Tā-men)　méi　　hē　　(jiǔ).

　　3-PL　　NEG　　drink　(wine)

　　'They didn't drink the wine.'

It is conventionally considered that the canonical word order of MC is SVO, which is inherited from archaic Chinese. However, as noted by Li and Thompson (1981), the language also exhibits certain features of SOV languages, such as the occurrence of (superficial) SOV sentences, PPs preceding the verb, relative clauses preceding the head noun, and aspectual markers following the verb[2]. Paul (2009) insisted on an SVO order for MC and arguments from archaic Chinese are presented in Paul (2015) supporting that SVO has always been the canonical order. Meanwhile, the author claimed that the language is head-final in the domain of CP. Hole and Zimmermann (2013) also maintained that MC is an SVO language with certain SOV traits. In this work, we follow Paul (2009, 2015) among others, and assume that the canonical order of MC is SVO, without excluding the possibility of head-final order in certain domains, as illustrated in (3).

(3) [CP [TP Wǒ zuótiān　　dào Zhāng jiā　　chī fàn] C0 le]

　　1SG yesterday go　Zhang home eat food　SFP[3]

　　'I went to Zhang's for dinner yesterday.'

　　(Paul, 2009, p.7, example adapted from Chao, 1968)

---

[2] Li and Thompson (1981) considered it difficult to establish a basic word order and assumed that Mandarin may be undergoing a transition from an SVO language to an SOV one. However, Huang (1994) argued that Chinese selects a head-initial structure only at the complement level and functions as a typical head-final language in other cases. See Huang (1994) for a detailed presentation.

[3] SFP: Sentence-final Particles, differentiated from perfective marker *-le*, which immediately follows the verb (Paul, 2009).

Following Huang (1994), scholars generally maintain that there is no V-to-T movement (or T-to-C movement) in Chinese, given the fact that there is no verb raising across negation (4a) or adverbials (4b). Meanwhile, as analyzed by Sybesma (2007), TP is assumed to be projected, since the absence of overt morphological evidence for a T node cannot justify the absence of tense (Huang, Li and Li, 2009; cf. Lin, 2006, 2010). The head of Infl/T can remain "covert", in the terms of Paul (2015), in that it needs not be lexically filled.

(4) a. Wǒ   méi   yǒu   mǎi /* mǎi   méi yǒu   shū.

   1SG   NEG have   buy/ buy NEG   have   book

   'I did not buy books.

   b. Wǒ   jīngcháng   mǎi /* mǎi   jīngcháng   shū.

   1SG   often   buy / buy   often   book

   'I often buy books.'

1.3 Background on the acquisition of EP as a foreign language by L1-MC learners

Studying the acquisition of EP by L1-MC learners faces the unneglectable problem that the majority of these learners acquired EP as their major of higher education in China. This implies that most of these learners acquired English before encountering EP.

English language education was first incorporated in the compulsory education in China, namely at the primary and secondary levels, in 1978. In 2001, the Ministry of Education of the People's Republic of China issued a directive requiring English classes to be offered at Primary 3 from the autumn of the same year (MOE, 2001a, as cited in Hu, 2004). Not only do most candidates choose English for their compulsory foreign language subject in the National College Entrance Examination, but also non-English-major students are required to pass College English Test-Band 4 (CET4)[4] in most Chinese universities (Shao, 2006), which offer tertiary English curricula in line with the College English Curriculum Requirements (CECR, 2004), according to Gao (2013). Therefore, the current

---

[4] A standardized exam administered by the Ministry of Education of China that assesses students' proficiency according to College English Syllabus and Teaching Requirements (2007). A higher level, CET 6, is also available. According to the syllabus, the testing guideline for CET contains a vocabulary list of 5,418 words (Gu, 2018). CECR 2004 recommends a vocabulary of at least 4,500 words and 700 phrases (Gao, 2013).

Chinese college students have been provided with at least nine years of English language learning inside the classroom since primary school and they continue to have access to formal English education until graduation.

Taking the long-term English learning by Chinese students into account, it is thus natural to assume that, in accordance with several models of L3/Ln acquisition presented in Chapter 3, the linguistic system of English as a second language may have a certain impact on the development of L3 EP grammar by L1 MC learners. More specifically, having acquired English cleft structures, namely *it*-clefts and pseudoclefts, may, in turn, exert an influence on the acquisition of cleft structures in EP by such learners.

1.4 Structure of the dissertation

The dissertation is composed of seven chapters, summarised as followed: in Chapter 2, we present the basic descriptions of cleft structures in EP and in MC, and review some of the previous literature on syntactic analyses of clefts in the two languages that guided the whole dissertation, especially the design of the experiment; Chapter 3 provides a brief exposition on the theoretical framework of generative L2 acquisition, hypotheses adopted in this study and the working hypotheses formulated taking into account the syntactic analyses of the structures and the acquisition theories; in Chapter 4, we describe the methodology adopted, the type of treatment and statistical analysis of the data, and present a complementary corpus search; Chapter 5 presents the results of the experimental task, and in Chapter 6 we discuss the data, taking into consideration the working hypotheses presented in Chapter 3; in Chapter 7, we present the main conclusion of this study and some suggestions for future research.

Chapter 2 Clefts in European Portuguese and Mandarin Chinese

In this chapter, we will explore the syntactic properties and some current syntactic analyses of cleft structures in European Portuguese and in Mandarin Chinese, with a brief note on cleft structures in English.

2.1. Clefts in European Portuguese

2.1.1 A descriptive presentation of European Portuguese clefts

In this section, six cleft structures in European Portuguese (EP) are presented following Ambar (1999), Costa and Duarte (2001, 2005), and Lobo (2006): a. standard cleft; b. *wh*-cleft; c. basic pseudocleft; d. inverted pseudocleft; e. *é-que* cleft; and f. semipseudocleft.

a) Standard cleft

In a standard cleft, the copular verb *ser* "be" occupies the sentence-initial position and is immediately followed by the clefted constituent, which precedes the embedded clause introduced by the complementizer *que* (5a,b).

(5) a. Foi   o   João   que   comprou   o   livro.
    be.PST.3SG   the   João   QUE   buy.PST.3SG   the   book
    'It was João that bought the book.'

   b.   Foi   o   livro   que   o João   comprou.
    be.PST.3SG   the   book   QUE   the João   buy.PST.3SG
    'It was the book that João bought.'

As shown in previous studies (Ambar, 1999; Lobo, 2006; Soares, 2006), in a standard cleft, if a subject is clefted, the embedded verb obligatorily agrees with the clefted subject in person and number (6a,b).

(6) a. Foram   os   teus   pais   que   telefonaram/* telefonou.
     be.PST.3PL the   your   parents   QUE   call.PST. 3PL/* call.PST.3SG
    'It was your parents that called.'

   b. Fui   eu   que   fiz/*fez   isso.

be.PST.1SG 1SG   QUE do.PST.1SG/*do.PP3SG   that

'It was I that did that.'

(Lobo, 2006, p. 3)

b) *Wh*-cleft

A *wh*-cleft consists of the verb *ser* "be" in the sentence-initial position, followed by the clefted constituent, and an embedded clause introduced by a *wh*-word, such as *o que* "what", *quem* "who", *quando* "when", *onde* "where", and *como* "how", all of which can function as the head of a relative clause. (7a,b) exemplifies subject and object *wh*-clefts, in which the embedded clauses are respectively introduced by *quem* "who" (7a) and *o que* "what", depending on the [+/- human] feature of the clefted DP (7b); (7c) exemplifies the possibility of clefting a verbal constituent.

(7) a. Foi   o   João   quem   comprou   o   livro.

be.PST.3SG   the   João   who   buy.PST.3SG   the   book

'It was João who bought the book.'

b. Foi   o   livro   o que   o João   comprou.

be.PST.3SG   the   book   what   the João   buy.PST.3SG

'It was the book that João bought.'

c. Foi   ler   livros o que o   João   fez   o   dia todo.

be.PR.3SG read.INF books   what the João   do.PST.3SG the day all

'What João did all day was to read books.'

As indicated by Lobo (2006) and Soares (2006), at least in the standard variety of EP, in a subject *wh*-cleft, the embedded verb always remains in 3rd person singular (8a,b).

(8) a. Foram   os   teus   pais   quem   telefonou.

be.PST.3PL   the   your   parents   who   call.PST.3SG

'It was your parents who called.'

b. Fui   eu   quem   fez   isso.

be.PST.1SG   1SG   who   do.PST.3SG   that

'It was I who did that.'

(Lobo, 2006, p. 3)

c) Basic pseudocleft and d) inverted pseudocleft

Specificational pseudoclefts feature a *wh*-clause, an XP, and the copular verb *ser* "be" linking the

two elements. In line with the literature on English pseudoclefts (Heycock & Kroch, 1996; Den Dikken, Meinunger & Wilder, 2000), EP pseudoclefts are divided into 2 subtypes: basic pseudoclefts and inverted pseudoclefts[5]. In basic pseudoclefts (9a), the *wh*-clause precedes the focused constituent, the DP *os alunos* "the students" in (9a). In inverted pseudoclefts (9b), it is the DP that precedes the *wh*-clause. The embedded verb always remains in 3rd person singular, when the subject is clefted.

(9) a. Quem   comprou   o livro   foram   os   alunos.

 who   buy.PST.3SG the book   be.PST.3PL   the   student.PL

'Who bought the book were the students.'

b. Os   alunos   foram   quem   comprou   o   livro.

 the.PL student.PL be.PST.3PL who buy.PST.3SG the book

'The students were who bought the book.'

Pseudoclefts and *wh*-clefts display similar behaviors in what concerns the agreement pattern and restrictions on the types of clefted constituents, e.g., there is no restriction on clefting a verbal constituent (10a,b), as stated by Lobo (2006).

(10) a. O que   o João   faz   o dia todo   é   ler   livros.

 what the João do.PRS.3SG the day all   be.PRS.3SG   read.INF books

'What João does all day is to read books.'

b. Ler   livros   é   o que   o João   faz   o dia todo.

 read.INF books   be.PRS.3SG t what the João   do.PRS.3SG the day all

'To read books is what João does all day.'

e). *É-que* cleft

In an *é-que* cleft, the focused constituent is placed in the sentence-initial position, followed by an expression *é-que* (11a,b).

(11) a. O João   é   que   comprou   o   livro.

 the João be.PRS.3SG QUE   buy.PST.3SG the book

'João (is that) bought the book.'

b. O livro   é   que   o   João   comprou.

---

5 Or Type A pseudoclefts and Type B pseudoclefts, in the terminology of Den Dikken, Meinunger & Wilder (2000) taken by Barbosa (2013).

the book be.PRS.3SG   QUE the João   buy.PST.3SG

'The book (is that) João bought.'

In the expression *é-que*, the verb *ser* "be" is invariable in person/number and in tense/mood (12a), irrespective of the verbal form in the embedded domain, and the *é* is not separable from the *que* (12b).

(12) a. *O João   foi   que   comprou   o   livro.

the João be.PST.3SG QUE   buy.PST.3SG the book

* *'João (was that) bought the book.'*

b. *A Ana   é   certamente que   sabe   isso.

the Ana   be.PRS.3SG   certainly QUE   know.PRS.3SG that

* *'Ana (is) certainly (that) knows this.'*   (Lobo, 2006, p. 8)

f). Semipseudocleft

In a semipseudocleft in EP, the verb *ser* "be" appears in the pre-focus position and after the lexical verb, generally without altering the canonical word order (13a), including the cases of clefting the subject of an unaccusative verb, which can inherently precede its subject (13b).

(13) a. O João   leu   foi   o   livro.

the João read.PST.3SG   be.PST.3SG the book

'João read (was) the book.'

b. Chegou   foi   a   Maria.

arrive.PST.3SG   be.PST.3SG the Maria

'There arrived (was) Mary.'

Not only direct objects and post-verbal subjects, but also indirect objects (14a) and adverbs (14b) can also be clefted in such structure.

(14) a. O João   deu   foi   o livro   à   Ana.

the João give.PST.3SG   be.PST.3SG the book to.the Ana

'João gave (was) the book to Ana.'

b. Ele   discursou   foi   muito bem.

3SG speak.PST.S3G be.PST.3SG very well

'He spoke (was) very well.'

(Costa & Duarte, 2001, p. 633)

21

2.1.2 Previous analyses of European Portuguese clefts

In this section, we will present different analyses of EP clefts divided in two kinds. The first kind of analyses, unified analyses, which include Ambar (1999) and Costa and Duarte (2001, 2005), assume that all different EP clefts have the same underlying syntactic configuration. The second kind of analyses, Lobo (2006) and Soares (2006), propose that these different cleft structures do not share an identical underlying syntactic structure and hence they assume a non-unified analysis of EP clefts.

2.1.2.1 Unified analyses: Ambar (1999) and Costa and Duarte (2001, 2005)

Ambar (1999)

Ambar (1999) proposed a unified analysis with the following characteristics: (i) the assumption of a projection in the left periphery of the clause, Topic_FocusP, where focus in EP is uniformly licensed; (ii) the relation between the verb and the focus, and the *Focus Tense Identification*; (iii) properties of the copula, BE (*ser*), which ends up in Topic_Focus^0 in both the BE_Focus structure and (pseudo-)cleft structures: when BE takes IP as its complement, the BE_Focus structure, also designated as semipseudocleft, is formed; when BE selects a CP, it yields (pseudo-)cleft structures, including other structures presented in the previous section. In what follows, I summarize the different elements of Ambar's analysis.

i) The assumption of Topic_FocusP.

According to Ambar (1999), both contrastive focus and presentational focus are always introduced with a given relation to old information. In a contrastive focus construction, the new information introduced in the Universe of Discourse is chosen from a set of entities within the interlocutors' shared knowledge. Hence, according to Ambar, a contrastive focus bears *topic-like* properties. In a presentational focus construction, the focus merely introduces new information without contrast, and a presupposed element precedes the focus and functions as its subject. This element, null or phonetically realized, refers to old information and is thus *topic-like*.

Therefore, the hypothesis is formed that focus obligatorily involves a *topic-like* element to be

licensed[6]. Ambar (1999) thus proposed a Topic_Focus Phrase in the left periphery where the focus is generally licensed. Topic_Focus^0 contains the topic feature of the *topic-like* element, [+topicf], and the focus feature [+focust][7].

ii) The verb and the focus & *Focus Tense Identification*.

Ambar suggested that, because focus is the identification of an event, the verb, more precisely, the Event component transmitted by Tense of the verb, has the capacity of licensing focus. With an exclusive interpretation on the focus, the verb bears a focus feature to be checked.

Contrastive focused constituents, as previously explained, bear both [topicf] and [focust] features, hence raise to Spec, Topic_FocusP to check both features. A contrastive focus can have either exclusive or non-exclusive reading, entailing different verb positions. When there is an exclusive reading, the verb raises to Topic_Focus^0 to check its focus feature. A Spec-Head relation is established, with the focus under the scope of the Event component. If the verb does not raise to such head, a non-exclusive reading is available.

A presentational focus always has exclusive interpretation. It is marked [+focus] instead of [+focust], which inhibits it from raising and keeps it as the complement of Topic_Focus^0 in a position within IP. Such a position counts as rightmost embedded position, satisfying Cinque (1993)'s algorithm for focal stress assignment. An element marked [+topicf][8] occupies Spec, Topic_FocusP. The [+focust] feature is checked by the verb, which obligatorily raises to the head and consequently c-commands the focused element in IP.

The author thus proposed the *Focus Tense Identification*: At the latest at LF, a full exclusive/restrictive focused element has to be under the scope of the lexical verb, more precisely under the scope of its Event component, whereby scope is defined either in terms of c-command or of Spec-Head agreement and Event is carried by Tense on the verb (Ambar, 1999, pp.37-38).

iii) The properties of BE & BE_Focus structure and (pseudo-)cleft structures.

---

[6] Ambar (1999, p. 23). As clarified in pp. 23-24, *topic-like* elements are distinct from true topics licensed in TopicP, which is situated above Topic_FocusP.

[7] Ambar (1999) explained that the [+topicf] feature differs from the true topic feature [+topic] checked in TopicP, and that a contrastive focus bears the [+focust] feature while a presentational focus bears [+focus] feature. The author also claimed that both [+topicf] and [+focust] features are interpretable but trigger movements.

[8] The *topic-like* element, phonetically realized or not, is assumed in the spirit of Rizzi (1995) and Ambar (1988) to be a non-quantificational anaphoric operator that binds its variable in IP.

Ambar (1999) maintained that BE is an expletive inserted in Topic_Focus^0, which, in a BE_Focus structure, selects IP, and in (pseudo-)cleft structures takes a CP as its complement. In these structures, the interpretation of the focus is always exclusive. According to Casteleiro (1976, 1977, *apud* Ambar, 1999), the expletive BE has the function of emphasizing the exclusive interpretation of the focus that it c-commands.

As BE is semantically empty, it does not carry Event specifications, and is intrinsically defective in its L-features. Therefore, it cannot check the V-features of functional heads – Tense and Agr – which have to be licensed (Chomsky, 1995), unless: (a) BE is in the scope of a lexical verb and the Tense indexation occurs between such verb and BE, permitted by its semantic emptiness; (b) BE bears *zero* Tense/Agr marks, i.e. the simple present and a simple agreement, which can escape from the checking domain; (c) the L-features of BE are no longer defective, namely, enriched by a lexical element in its checking domain, while non-*zero* Tense and Agr must occur.

According to the author, in a BE_Focus structure, the focused element stays in IP, and the lexical verb must raise, and in this case to Topic^0, as shown in (15a-b). The trace left in Topic_Focus^0 is then lexicalized by the inserted expletive BE. BE thus occurs in a chain headed by the lexical verb and, through the Tense indexation, inherits its Event specifications. Hence the *Focus Tense Identification* applies, and the exclusive interpretation is obtained.

(15) a. [$_{TopicP}$ a tarte$_i$ [$_{Topic''}$ comeu$_v$ -a$_i$ [$_{Topic\_FocusP}$ [$_{Topic\_Focus'}$ foi$_v$ [$_{IP}$ a Maria t$_v$ pro$_i$]]]

      the pie   eat.PST.3SG -it.ACCU   be.PST.3SG the Maria

   'The pie, ate it (was) Maria.'

b. [$_{TopicP}$ a Maria$_i$ [$_{Topic''}$ comeu$_v$ [$_{Topic\_FocusP}$ [$_{Topic\_Focus'}$ foi$_v$ [$_{IP}$ pro$_i$ t$_v$ a tarte$_i$]]]

      the Maria   eat.PST.3SG   be.PST.3SG   the pie

   'Maria ate (was) the pie.'

(Ambar, 1999, p. 37)

The author took the (pseudo-)cleft constructions to be sentential complements of BE, more precisely, of Tense, which bears a focus feature and hence lands in Topic_Focus^0. The author proposed an underlying structure of all the derived (pseudo-)clefts, as shown in (16):

(16) [$_{TopicP}$ [$_{Topic'}$ [$_{Topic\_FocusP}$ [$_{TopicFocus'}$ [$_{IP}$ **[e]** é [$_{CP}$ [$_{C'}$ [$_{C^0}$ que [$_{IP}$ os meninos ouviram as

                               be   that   the boys hear.PST.3PL the

sonatas]]]]]]]]]

sonata

When a constituent inside IP is marked as [+topicf, +focust], that is, it is a contrastive focus, it moves to Spec, Topic_FocusP. BE bears *zero* Tense and Agr marks, hence will not block the Tense/Agr raising. At LF, the Tense of the lexical verb raises to the Tense position of BE to transmit the Event specifications. BE/Tense raises to Topic_Focus^0 to check its focus feature and ends up in a Spec-Head relation with the focus. *Focus Tense Identification* applies, and thus the exclusive interpretation is obtained, yielding an *é-que* cleft (17):

(17) [TopicP[Topic' [Topic_FocusP os meninos$_i$ [Topic_Focus' é$_v$ [IP t$_i$ t$_v$ [CP t$_i$ [C' que [IP t$_i$ ouviram as sonatas]]]]]]]].

the boys   be   that   hear.PST.3PL the sonata

'The boys (is that) heard the sonatas.'

When there is a presentational focus, an Event Operator recuperating *topic-like* elements fills Spec, Topic_FocusP and checks both features of Topic_Focus^0, yielding standard clefts (18).

(18) [TopicP [Topic' [Topic_FocusP OP$_{Ev}$ [Topic_Focus'   foram$_v$ [IP os meninos$_i$ t$_v$ [CP t$_i$ [C'   que [IP t$_i$ ouviram

be.PST.3PL   the boys   that   hear.PST.3PL

as sonatas]]]]]]]].

the sonata

'It was the boys that heard the sonatas.'

Since the lexical element *os meninos* is in the checking domain of BE, non-*zero* Tense and Agr obligatorily occur. Tense raising is blocked by the visible Agr. However, BE lacks the Event specifications necessary for focus licensing. The author hence suggested a Head-Head relation to obtain Tense coindexation and thus convey the Event specifications.

As for pseudoclefts and *wh*-clefts, the author suggested that the *wh*-element results from "the Spec-Head relation between the trace of the moved focused element and the complementizer." (Ambar, 1999, p. 44). A contrastive focus yields (19a), as the focused element raises to Topic_FocusP; (19b) bears a presentational focus, in that BE precedes the focus. Spec, Topic_FocusP is thus filled by an Event Operator.

(19) a. [TopicP [Topic' [Topic_FocusP os meninos$_i$ [Topic_Focus' foram$_i$ $_v$ [IP t$_i$ t$_i$ $_v$ [CP t quem$_i$ [IP t$_i$ ouviu   as

the boys be.PST.3PL who hear.PST.3PL the

sonatas]]]]]]].

sonata

'The boys were who heard the sonatas.'

b. [$_{TopicP}$ [$_{Topic'}$ [$_{Topic\_FocusP}$ OP$_{Ev}$ [$_{Topic\_Focus'}$ foram$_v$ [$_{IP}$ os meninos$_i$ t$_v$ [$_{CP}$ t quem$_i$ [$_{IP}$ t$_i$ ouviu

be.PST.3PL the boys who hear.PST.3PL

as sonatas]]]]]]].

the sonata

'It was the boys who heard the sonatas.'

(Ambar, 1999, p. 44)

Since *quem* "who" is a lexical element in the checking domain of BE, true Tense obligatorily appears on the copula, matching the Tense of the lexical verb through coindexation. The Event specifications of the lexical verb is transmitted to BE/Tense via the lexical head of CP, *quem*.

When the entire CP refers to old information, functioning as the subject of focus in a presentational focus structure, the whole CP raises to Spec, Topic_FocusP, yielding (20).

(20) [$_{TopicP}$ [$_{Topic'}$ [$_{Topic\_FocusP}$ [$_{CP}$ quem$_i$ [$_{IP}$ t$_i$ ouviu   as sonatas]]$_k$ [$_{Topic\_Focus'}$ foram$_v$ [$_{IP}$ os meninos$_i$ t$_v$ t$_k$]]]]].

who hear.PST.3PL the sonata be.PST.3PL the boys

'Who heard the sonatas were the boys.'

(Ambar, 1999, p. 44)

Costa and Duarte (2001, 2005)

Costa and Duarte (2001, 2005) proposed a unified analysis for all six clefting strategies in EP, maintaining that clefting constructions originate in an identificational structure, in which the identification relation is established between two terms of a small clause. According to the authors, the subject of the small clause is either a relative clause, which includes an operator-variable relationship, or another clausal constituent with an empty position legitimated by an operator. The functional node I can be filled by the verb *ser* (be) that raises from V, following Ambar (1999). In *é-que* clefts, the expression *é-que* functions as a lexicalizer for C, resulting from a reanalysis process. Hence, the authors proposed the following underlying structures for clefting constructions (21):

(21) a. [$_{IP}$ [$_{vP}$ ser[$_{SC}$ [$_{DP/CP}$ {o que/OP que}   o João comeu]   [$_{DP}$ o bolo]]]]

be   what/OP QUE the João eat.PST.3SG   the cake

b. [$_{IP}$ [$_{vP}$ ser [$_{SC}$ [$_{CP}$ OP [$_{IP}$ o João   comeu]]   [$_{DP}$ o bolo]]]]

be   the João eat.PST.3SG   the cake

c. [$_{CP}$ é que [$_{SC}$ [$_{CP}$ OP [$_{IP}$ o João comeu]]   [$_{DP}$ o bolo]]]

é que   the João eat.PST.3SG the cake

(Costa & Duarte, 2005, examples (9-11))

Costa and Duarte (2001, 2005) held that Spec, IP may be occupied by either a DP or a relative clause. If the relative clause raises to Spec, IP, according to the authors, a basic pseudocleft with the information focus in the embedded DP is derived, and agreement occurs between the post-verbal DP and the verb[9], as shown in (22):

(22) a. [$_{IP}$ [$_{DP}$ O que o João comeu]$_i$ foi$_j$ [$_{vP}$ t$_j$ [$_{SC}$ t$_i$ [$_{DP}$ o bolo]]]]

what the João eat.PST.3SG   be.PST.3SG   the cake

'What João ate was the cake.'

(Costa & Duarte, 2005, example (17))

b. O que o   João   comeu   foram   os bolos.

what the João eat.PST.3SG   be.PST.3PL the cakes

'What João ate were the cakes.'

(Costa & Duarte, 2001, example (25b))

If it is the DP that raises to Spec, IP and hence satisfies the EPP feature of I, an inverted pseudocleft is obtained, with the information focus in the relative clause, and as shown in (23), agreement with the DP in subject position occurs:

(23) a. [$_{IP}$ [$_{DP}$ O bolo]   foi [$_{vP}$ ~~o bolo~~ [$_{SC}$ [$_{DP}$ o que o João comeu]   ~~foi~~]]

the cake   be.PST.3SG   what the João eat.PST.3G

'The cake was what João ate.'

(Costa & Duarte, 2005, example (20))

b. Os bolos   foram / *foi   o que   o João   comeu.

---

[9] In Costa and Duarte (2005), the authors also pointed out that the verb may agree with the preverbal DP/CP, a subject-verb agreement pattern also observed in other null subject Romance languages.

the cakes   be.PST.3PL be.PST.3SG   what   the João   eat.PST.3G

'The cakes was what João ate.'

(Costa & Duarte, 2001, adapted from example (27c))

According to the authors, since clefts are identificational structures instead of a predicative small clause, it is possible for the predicate DP to move by *scrambling* to the left of the small clause; therefore, the following configurations corresponding to a *wh*-cleft (24a) and a standard cleft (24b) are derived.

(24) a. [$_{IP}$ foi [$_{vP}$ [$_{DP}$ o bolo] [$_{vP}$ ~~foi~~ [$_{SC}$ [$_{DP}$ o que o João comeu] ~~o bolo~~]]

    be   the cake   what the João eat.PST.3SG

  b. [$_{IP}$ foi [$_{vP}$ [$_{DP}$ o bolo] [$_{vP}$ ~~foi~~ [$_{SC}$ [$_{CP}$ *Op* que o João comeu] ~~o bolo~~]]

    be   the cake   OP QUE the João eat.PST.3SG

'It was the cake that João ate.'

(Costa & Duarte, 2005, examples (24-25))

As *scrambling* is generally associated with defocalization, the authors maintained that, in *wh*-clefts and standard lefts, the clefted DP does not correspond to the information focus. Reconstruction effects displayed in *wh*-clefts and standard clefts (Modesto, 1995) are compatible with the properties of *scrambling* as an instance of A-bar movement.

Concerning *é-que* clefts, the authors argued that the expression *é-que*, as in *wh*-interrogatives, is a lexicalizer of certain features related to illocutionary conditions of C [10], resulting from a diachronically attested reanalysis process. A representation of the structure under discussion is given below (25):

(25) [$_{CP}$ [$_{DP}$ o bolo] é que [$_{SC}$ [$_{CP}$ *Op* o João comeu] ~~o bolo~~]]

    the cake   é-que   the João eat.PST.3SG

'The cake (is that) João ate.'

(Costa & Duarte, 2005, example (39))

The semipseudocleft is treated by the authors as being associated to the null object construction. In addition, they maintained that the construction obligatorily involves a non-maximal VP and

---

[10] In Costa and Duarte (2001), it was assumed that the expression *é-que* lexicalized I, a case in which there must be a defective CP. However, such hypothesis was refuted by evidence from language acquisition in Soares (2003) and Santos (2003), *apud* Costa & Duarte (2005), in that *é-que* clefting is only acquired when children show evidence for having mastered CP (Costa and Duarte, 2005, footnote 6).

provided the following arguments. Firstly, in opposition to traditional claims, semipseudoclefts cannot be derived from pseudoclefts by deleting the relative pronoun (26).

(26) a. (*Como) ele discursou foi muito bem.

(How) 3SG speak.PST.3SG be.PST.3SG very well

'(*How) he spoke was very well.'

<div align="right">(Costa & Duarte, 2005, example (50a))</div>

Secondly, the construction is only legitimate with constituents smaller than VP (27).

(27) *O João foi leu o livro.

the João be.PST.3SG read.PST.3SG the book

*'João (was) read the book.'

<div align="right">(Costa & Duarte, 2005, example (48a))</div>

Thirdly, the authors claimed that there is an asymmetry between subject and object in the case of semipseudoclefts, that is, a subject can never be clefted in a semipseudocleft (28).

(28) * (Quem) leu o livro foi o João.

(who) read.PST.3SG the book be.PST.3SG the João

*'Read the book was João.'

<div align="right">(Costa & Duarte, 2005, example (51b))</div>

2.1.2.2 Non-unified Analyses: Lobo (2006) & Soares (2006)

In contrast to Ambar (1999) and Costa and Duarte (2001, 2005), Lobo (2006) proposed a non-unified analysis for different clefting strategies of EP. Soares (2006) also proposed a similar analysis differentiating distinct types of clefts. Lobo (2006) suggested different basic syntactic configurations of cleft structures: CP structures selected by the verb *ser* "be"; identificational structures with base-generated focused constituents and free relative clauses; A'-movement of focused constituents to the left periphery. In Lobo, Santos and Soares-Jesel (2016), the authors proposed that *ser* "be" is reanalyzed as a contrastive focus marker.

Having observed the differences between standard clefts and *wh*-clefts, Lobo (2006) concluded that the two types of clefts have distinct structures. Firstly, the agreement patterns of the two types of clefts are distinct. When the subject is clefted, agreement is obligatory between the subject and the embedded verb in a standard cleft, as shown in (29a-b), while in a *wh*-cleft, the embedded verb form

cannot agree with the clefted subject in the standard variety of EP (29c-d).

(29) a. *Foram   os   teus   pais   que   telefonou (telefonaram).

be.PST.3SG the your parents QUE call.PST.3SG (3PL)

'It was your parents that called.'

b. *Fui   eu   que   fez   isso.

be.PST.1SG 1SG QUE do.PST.3SG that

'It was I that did that.'

c. *Foram   os   teus pais quem telefonaram.

be.PST.3PL the your parents who call.PST.3PL

*'It was your parents who called.'

d. *Fui   eu quem   fiz   isso.

be.PST.1SG 1SG who do.PST.1SG that

*'It was I who did that.'

(Lobo, 2006, p. 3)

Secondly, the two clefts behave differently in terms of cleftable constituents. A verbal constituent cannot be clefted in a standard cleft (30a), while a *wh*-cleft clefting on a verbal constituent is legitimate (30b).[11]

---

[11] Also, a standard cleft can cleft adverbial and prepositional constituents (ia,c), whereas in a *wh*-cleft, adverbial and prepositional constituents are not easily clefted (ib,d).

(i) a.   Era   do Rui que ela   gostava.

be.PST.IPFV.3SG of.the Rui QUE 3SG like.PST.IPFV.3SG

'It was Rui that she liked.'

b. ??/*Era   do Rui de quem ela   gostava.

be.PST.IPFV.3SG of.the Rui of who 3SG like.PST.IPFV.3SG

??/*'Rui was who she liked.'

c. Foi   rapidamente que ele   tomou   o   pequeno-almoço.

be.PST.3SG rapidly QUE 3SG have.PST.3SG the   breakfast

'It was rapidly that he had the breakfast.'

d. *Foi   muito lentamente como ele   abriu   os   olhos.

be.PST.3SG very slowly   how 3SG open.PST.3SG the eyes

* 'How he opened the eyes was very slowly.'

(Lobo, 2006, p. 4)

Apart from the aforementioned differences, Lobo (2006) demonstrated that standard clefts and *wh*-clefts differ in other aspects, such as changes in the order of constituents (iia) and the restriction on clefting a pronoun (iib).

(ii) a. O queijo   foi   o que/*que o corvo   comeu.

(30) a. *Foi   telefonar   à   namorada   que   ele   fez.

      be.PST.3SG   call.INF   to.the   girlfriend   QUE   3SG   do.PST.3SG

      * *'It was call the girlfriend what he did.'*

   b.   Foi   telefonar   à   namorada   o que   ele   fez.

      be.PST.3SG   call.INF   to.the   girlfriend   what   3SG   do.PST.3SG

     'What he did was to call his girlfriend.'

<div align="right">(Lobo, 2006, p. 4)</div>

Moreover, the author mentioned that a unified analysis taking standard clefts as relative structures as in Costa and Duarte (2001, 2005) does not provide an account of the ungrammaticality of a headless relative clause introduced merely by *que* (31a); while an analysis like Ambar (1999) taking *wh*-clefts as subcategorized structures involving a reanalysis of the *that*-trace of the Q-form falls short in explaining why such a phenomenon does not occur in other constructions supposedly entailing a similar movement (31b).

  (31) a. *Percebo   que   queres   dizer.

      understand.PRS.1SG   that   want.PRS.2SG   say.INF

      * *'I understand that you want to say.'*

   b. Quem   disse   o   João {*quem/que}   telefonou?

      who   say.PST.3SG the   João   who/that   call.PST.3SG

     'Who did João say that called?'

<div align="right">(Lobo, 2006, p. 5)</div>

Hence, in line with Modesto (1995), Lobo (2006) suggested that a standard cleft involves the movement of the clefted constituent from within the clause introduced by *que*, in which the clefted constituent is an argument or an adjunct in the base. Following Ambar (1999), Lobo (2006) proposed that the verb *ser* (be) selects a CP.

---

      the cheese be.PST.3SG what/*QUE the crow eat.PST.3SG

     Int. reading: 'The cheese was what the crow ate.'

  b.   Fui   eu quem   /*/?? que   o   Pedro   convidou.

    be.PST.1SG 1SG who/ */??QUE the Pedro invite.PST.3SG

     Int. reading: 'It was I who Pedro invited.'

<div align="right">(Lobo, 2006, p. 4)</div>

Soares (2006) independently reached a similar analysis and claims that standard clefts do not share an identical syntactic structure with *wh*-clefts, but correspond to a structure in which a CP is subcategorized by *ser* "be" and the clefted constituent is generated in the embedded clause and displaced to a position in the left periphery[12].

In line with previous studies such as Chomsky (2005), Soares (2006) suggested that C^0 contains a tense feature [T], and that in EP the feature [T] of a C^0 in the embedded domain can be interpretable – [iT], requiring the insertion of a complementizer by external merge -, or uninterpretable – [uT], causing T-to-C movement of the verb. Another feature of C^0 is an uninterpretable "first position" feature with an EPP property [uF$_{EPP}$].

In a standard cleft like (32a), since the embedded C^0 is considered as the complement of *ser* "be", Soares (2006) assumed that this head contains a [uF$_{EPP}$] and an [iT], which is checked by the complementizer *que* by external merge[13] (32a').

(32) a. Foi   a Maria   que   comeu   o   bolo.

be.PST.3SG the Maria QUE eat.PST.3SG the cake

'It was Maria that ate the cake.'

a' [CP [C que$_{[i\,T,\,u\,F]}$] a Maria comeu o bolo].

(Soares, 2006, pp. 224-225)

According to the Single Licensing Condition (SLC, Nash and Rouveret, 2002)[14], as a result of the complementizer *que* checking the [iT] in C^0, the EPP-feature [uF$_{EPP}$] cannot be licensed locally. Consequently, such feature fissions and a proxy category is projected (henceforth XP). The [uF$_{EPP}$] copied onto X^0 thus attracts the clefted constituent, which is capable of checking this feature, to its specifier position. *Ser* "be" is then merged with the embedded CP, as in (33).

---

[12] By comparing data of Portuguese left-dislocated constituents to properties of contrastive focus structures in Italian, Soares (2006) argued that, in EP, there is no strong evidence for the existence of a functional projection like FocusP. Instead, the author suggested that C^0 is the only functional head in the left periphery in EP. See Soares (2006) for a more complete and detailed discussion.

[13] The author held that the C-clause cannot be closed by the external merge of the complementizer and consequently, the feature [iT] must be checked first. See Soares (2006) for discussion on the temporal dependency of C domain and the feature checking of [iT] concerning the Priority Principle (Nash & Rouveret, 2002).

[14] Single Licensing Condition: A functional category can enter into a licensing relation with the feature content of only one terminal node in its checking domain (Nash & Rouveret, 2002, p. 168). Adopting such condition and the Visibility Condition for the C domain of Platzack (1998), Soares (2006) reached the analysis of EP standard clefts and *é-que* clefts.

(33) [TP *pro* [T' foi [VP [v' V⁰ [XP a Maria [X' X⁰[*trF*] [CP [C' que[*iT*] [TP ~~a Maria~~ leu o livro]]]]]]]]]]

be.PST.3SG   the Maria   QUE   read.PST.3SG the book

'It was Maria that read the book.'

According to Lobo (2006), agreement between *ser* "be" and the clefted constituent – only when the latter is a DP -, results from feature matching, i.e. Agree between *pro* and the clefted constituent. The analysis of movement is found to be coherent with the compulsory agreement of a clefted subject and the verbal form in the subordinate clause. Since a *pro*-VP is not available in this hypothetical basic structure, the constraints of clefting a verbal constituent in a standard cleft is also explained.

Regarding *wh*-clefts, as well as other structures involving a *wh*-morpheme, namely basic pseudoclefts and inverted pseudoclefts, the author suggested that they are copular sentences and that the verb *ser* (be) selects a small clause that consists of a focused constituent and an unselected (relative clause) CP, the *wh*-clause (Costa & Duarte, 2001; Heycock & Kroch, 1999). In this structure, and contrasting with standard clefts, the clefted constituent is therefore generated outside of the *wh*-clause. Between the *wh*-constituent and the focused constituent, an anaphoric relation is required to be established (Lobo, Santos & Soares-Jesel, 2016). This hypothesis of base-generation is compatible with the inexistence of agreement between the clefted subject and the embedded verb as well as the possibility of clefting verbal constituents in the aforementioned structures.

Differences also exist amid *wh*-clefts, basic pseudoclefts and inverted pseudoclefts. As pointed out in Heycock and Kroch (1999) and taken up by Lobo (2006), a basic pseudocleft exhibits connectivity effects, for example, the licensing of items of negative polarity (34a), which contrasts with an inverted pseudocleft or a *wh*-cleft (34b,c)[15].

(34) a. O que ele não   faz   é   coisa   nenhuma.

what 3SG NEG do.PRS.3SG   be.PRS.3SG thing   none

'What he does not do is nothing.'

---

[15] In addition, Lobo (2006) mentioned that a basic pseudocleft easily legitimates a clefted prepositional phrase or an adverbial phrase, while the corresponding structure in an inverted pseudocleft and a *wh*-cleft is marginal. Furthermore, different clefts can be associated to different interpretations. Barbosa (2013) also pointed out more differences between a basic pseudocleft and an inverted pseudocleft in that only the former requires conformity of category between XP and the moved constituent inside of *wh*-clause.

b. *Coisa nenhuma   é   o que ele   não   faz.

    thing   none be.PRS.3SG what 3SG NEG   do.PRS.3SG

    *'Nothing is what he does not do.'

c. *É   coisa nenhuma o que ele   não   faz.

    be.PRS.3SG thing   none   what 3SG NEG   do.PRS.3SG

    *'Is nothing what he does not do.'

<div align="right">(Lobo, 2006, p. 10)</div>

Therefore, Lobo (2006) suggested that *wh*-clefts and pseudoclefts do not share a completely identical base structure, more precisely, the position of a relative clause in structures involving a *wh*-morpheme may not always be the same. Quite the opposite, when the clefted XP precedes the relative CP, XP is generated as the subject of the small clause and CP is the predicate, which is coherent with the case of *wh*-clefts and inverted pseudoclefts; when the CP precedes the clefted XP, CP is generated as the subject and XP is the predicate, corresponding to basic pseudoclefts and explaining the effects described in (34). The underlying structures of these three types of clefts are summarized below in (35).

(35) a. Basic pseudocleft:

ser [$_{SC}$ CP-Q$_{SUBJECT}$ XP$_{PREDICATE}$] >> CP-Q$_{SUBJECT\ i}$ ser [$_{SC}$ t$_i$ XP$_{PREDICATE}$]

  b. Inverted pseudocleft:

ser [$_{SC}$ XP$_{SUBJECT}$ CP-Q$_{PREDICATE}$] >> XP$_{SUBJECTi}$ ser [$_{SC}$ t$_i$ CP-Q$_{PREDICATE}$][16]

<div align="right">(Lobo, 2006, p. 12)</div>

As for *wh*-clefts, Lobo (2006) assumed the same underlying syntactic structure as the inverted pseudoclefts, and provided two alternative solutions for the derivation. In the first solution, the clefted CP in a *wh*-cleft raises to Spec, I and ser "be" moves to C. In the second one, the clefted XP remains *in situ* in the subject position of SC.

This bi-clausal analysis solves the problem of categorial restrictions on *wh*-clefts and inverted pseudoclefts, in which clefted PPs and AdvPs are hardly accepted due to the unacceptance of a non-nominal constituent, e.g., a PP or AdvP, in the subject position. Meanwhile, in a basic pseudocleft, as

---

[16] Lobo (2006) also provided an alternative derivation for specificational inverted pseudocleft of a prepositional phrase, in which the clefted PP is generated in the predicate position in the Small Clause and then moved to an A'-position.

the non-nominal constituent functions as a predicate, a PP or AdvP can be easily clefted.

When it comes to agreement patterns, the author followed the proposal of Tavares (2005) for nominal identificational structures and suggested that the verb *ser* agrees with the referentially most prominent constituent. Hence, since one acknowledges that clefts that involve a *wh*-morpheme are identificational structures, by assuming a referential hierarchy of pronoun > DP > CP-Q in the light of Tavares (2005), the puzzle of agreement can be logically solved as the verb always agrees with the element that corresponds to the higher position of the hierarchy, regardless of whether such an element is the subject or the predicate.

As for *é-que* clefts, Lobo (2006) and Soares (2006) maintained that such cleft sentences contain mono-clausal structures, with a functional head in the left periphery lexicalized by the crystalized expression *é-que*, in accordance with various previous studies. Besides, *é-que* clefts share the same kind of agreement pattern with standard clefts (36a) as well as the impossibility of clefting VPs (36b), which helps to distinguish *é-que* clefts from *wh*-clefts and pseudoclefts, by excluding the possibility that any relative clause is involved and consequently sustaining this perspective.

(36)   a. *Estes alunos   é   que   chegou   tarde.

         these student.PL be.PRS.3SG QUE   arrive.PST.3SG   late

       *'*These students (is that) arrived late.'*

   b. *Fumar um cigarro   é   que   o Zé   faz   quando   acorda.

       smoke.INF a cigarette be.PRS.3SG QUE the Zé   do.PRS.3SG when awake.PRS.3SG

     *'*To smoke a cigarette (is that) Zé does when he wakes up.'*

(Lobo, 2006, p. 8)

Lobo (2006) further observed that Portuguese *é-que* clefts have an interpretation that is similar to constructions of contrastive focalization in Italian and Spanish, which are generally analyzed as a result of A' movement of the focused constituent to Spec, CP or Spec, FocP. Given that *é-que* clefts have also been associated with contrastive focus reading in the literature, the author suggested that in an *é-que* cleft the expression *é-que* accounts for the contrastive reading and occupies the C head, and that the clefted constituent is extracted from within IP and placed in Spec, C.

Soares (2006) proposed that the expression *é-que* checks the [uT] feature of C by external merge. The feature [uFEPP] fissions and gives rise to a proxy category XP. The clefted constituent is attracted to the specifier position of XP (37).

(37) [XP a Maria [X' X[*nFEPP*] [CP [C' é   que[*nT*] [TP ~~a Maria~~ leu   o livro]]]]]

       the Maria   be.PRS.3SG QUE   read.PST.3SG the book

'Maria (is that) read the book.'

<div align="right">(Soares, 2006, p. 223)</div>

As argued by Vercauteren (2016), following Soares (2006), in *é-que* clefts, the clefted constituent and the expression *é-que* do not occupy the specifier and the head positions of the same projection, due to the possibility of adjuncts and parentheticals intervening between the clefted constituent and *é-que* (38).

(38) E eu, como sendo o mais velho, é que fui sempre o mais escravo. (AAL35)[17]

  and 1SG as   being the most old É QUE was always the most enslaved

  ''And I, being the oldest, (is that) was always the most hard-working.'

<div align="right">(Vercauteren, 2016, p. 264)</div>

Regarding semipseudoclefts, Lobo, Santos and Soares (2012, 2016) recognized the possibility of clefting a subject, more precisely, a postverbal subject embedded in a position internal to *v*P, "the default position associated to information focus" (Lobo, Santos and Soares-Jesel, 2016, p. 147). As any focused constituent in a semipseudocleft occurs in such a position, the authors assumed that this structure does not involve movements to the left periphery. Furthermore, following Mioto (2012), Lobo, Santos and Soares-Jesel (2016) suggested that the verb *ser* "be" marks the boundary of the left periphery of *v*P and is reanalyzed as a contrastive focus marker that has scope over all the materials in *v*P. According to the authors, elements that are defocalized move out of the *v*P domain by scrambling. The analysis is supported by the following arguments. Firstly, the semipseudocleft is the only cleft structure that can focalize more than one constituent (39b). Secondly, if, as assumed by Costa (1998), the adverb *bem* (well) can mark the leftmost frontier of *v*P, and if it can immediately precede *ser* (be) in semipseudocleft structures (39), this implies that the focalized constituent is within *v*P.

(39) a. Dançou   bem   foi   a Maria.

     dance.PST.3SG well   be.PST.3SG the Maria

     'Danced well (was) Maria.'

---

b. O Pedro   colocou   bem   foi   os pregos na janela.

the Pedro put.PST.3SG well be.PST.3SG the nails at.the window

'Pedro put well (was) the nails at the window.'

(Lobo, Santos & Soares, 2012, p. 3)

In language acquisition, spontaneous production data of clefts in child speech corpora (Santos, 2006; Soares, 2006) and elicited production data from an experiment carried out by Lobo, Santos and Soares-Jesel (2016) were supportive of the hypothesis that different types of clefts have diverse syntactic structures. A clear asymmetry was found between *é-que* and standard clefts on the one hand, and semipseudoclefts and clefts involving a *wh*-constituent on the other, as the former structures emerged earlier in the spontaneous production data and were generally more frequent, while the latter types not only were rare but also emerged comparably later. The differences in frequency and emergence period of different cleft types observed in the data are expected from the perspective of the non-unified analysis. Furthermore, the prevalence of subject clefts over other constituents in both spontaneous and elicited production data suggests a significant subject/non-subject asymmetry, in line with several well-known studies which have found this type of asymmetry in the production and comprehension of other types of A-bar structures (e.g., Friedmann, Belletti & Rizzi, 2009). In the elicited production test, both children and adults hardly produced clefts except in the subject condition, and when the objects or adjuncts were contrastively focused, they preferred other strategies, such as prosodic stress.

2.2 Clefts in Mandarin Chinese

2.2.1 A Descriptive Presentation of Clefts in Mandarin Chinese

MC clefts involve the presence of the copula *shì* "be" and, in some constructions, a particle *de*. In this section, we describe three different types of cleft structures that have been discussed in the literature:

1). *Shì … de* cleft constructions, in which XP (the subject or adjunct), the verb or the object can be clefted, including:

(a) *shì* **XP** V O *de* cleft[18] , a case in which only XP (a subject or an adjunct) can be clefted, but O (or V) cannot be clefted

(b) *shì* **XP** V *de* O / *shì* **V** *de* **O** cleft, in which, according to some authors, O, V or the entire VP can also be clefted;

2). (subject) V (O) *de* (X) *shì* **NP**[19] pseudocleft, in which the NP is the focused element, and an argument X sharing the same features as the NP can optionally occur;

3). Bare-*shì* subject cleft construction[20].

In line with previous studies, the three aforementioned structures, conventionally considered to be clefts, can be discerned in terms of cleftable constituents. The structures all obey the exclusiveness condition, denoting a contrastive and exclusive focus. Some relevant properties of these structures are hereinafter presented.

-Shì … de cleft constructions

a. *shì* **XP** V O *de* cleft

With this configuration, in a sentence, only the subject and adjunct can be focused upon (represented as XP). All kinds of clefted constituents should immediately follow the verb *shì*, and *de* always stays in the sentence-final position. (40a-c) demonstrate the possibility of clefting a subject or

---

[18] It is important to distinguish between *shì...de* cleft and *shì...de* propositional assertion (in the terminology of Paul and Whitman, 2008), with the latter one conveying the implicature of the speaker's certainty about the truth of the proposition and its relevance to the discourse context instead of focalizing any possible singled-out element. This structure cannot be classified as a cleft, as exemplified in (i):

    (i). a. Tā   shì gēn   nǐ   kāi wánxiào de.

          3SG BE   with   2SG open joke   DE

          '(It is the case that) he was joking with you.'

                                (Paul & Whitman, 2008, p11, example taken from Chao, 1968)

[19] We take the position of Aoun and Li (2003) and Huang, Li and Li (2009) acknowledging the existence of a Determiner Phrase (DP) in MC while D can be null. However, we accept the arguments of Aoun and Li (2003, p. 144) in favor of the analysis of relative construction being NP.

[20] There are in fact two configurations related to focus and consisting of bare *shì*, i.e., in the absence of *de*: a bare-*shì*-subject configuration, and a subject-*shì* configuration ("medial bare-*shì* pattern", in the terminology of Paul and Whitman, 2008), the former corresponding to a cleft, and the latter corresponding to a structure of association to focus which is not a cleft.

an adjunct in the configuration schematized in (a).

(40) a. Shì **wǒ** mǎi shū de.   subject cleft

　　　 be 1SG buy book DE

　　　 'It was me that bought the book.'

b. Wǒ shì **zuótiān** mǎi shū de.   adjunct (adverb) cleft

　 1SG be yesterday buy book DE

　 'It was yesterday that I bought the book.'

c. Wǒ shì **zài shūdiàn** mǎi shū de.   adjunct (prepositional phrase) cleft

　 1SG be LOC bookstore buy book DE

　 'It was at the bookstore that I bought the book.'

In this structure, neither can an object be dislocated to the position immediately following *shì*, nor can it be prosodically stressed (41a). On a par with the object, the verb cannot be clefted either (41b).

(41) a. *Wǒ shì mǎi **shū** de/ **shū** mǎi de.   object cleft (intended)

　　　 1SG be buy book DE/book buy DE

　　 **'It was the book that I bought.'*

b. *Wǒ shì **mǎi** shū de.   predicate cleft (intended)

　 1SG be buy book DE

　 **'It was buying the book that I was doing.'*

b. *shì* **XP** V *de* O / *shì* **V** *de* O / *shì* V *de* **O** cleft

What superficially distinguishes configuration (b) from (a) is the relative position of *de* and the object (O). In (a), *de* is in sentence-final position, while in (b), *de* is located between V and O. With configuration (b), one can also cleft a subject or an adjunct (42a-c). According to some studies (Paul & Whitman, 2008; Hole, 2011, among others), the object can also be clefted in the configuration (b), corresponding to V *de* O order (42d).

(42) a. Shì **wǒ** mǎi de shū.   subject cleft

　　 be 1SG buy DE book

　 'It was me that bought the book.'

b. Wǒ shì **zuótiān** mǎi de shū.   adjunct (adverb) cleft

39

1SG be   yesterday buy   DE   book

'It was yesterday that I bought the book.'

c. Wǒ shì **zài   shūdiàn**   mǎi   de   shū.   adjunct (prepositional phrase) cleft

1SG be LOC bookstore buy DE   book

'It was at the bookstore that I bought the book.'

d. Wǒ shì   mǎi   de   **shū**.   object cleft

1SG be   buy   DE book

'It was the book that I bought.'

To test the cleftability of V and O in the configuration (b), we carried out an acceptability judgement survey. 50 native MC speakers were consulted with an online questionnaire which consisted of three types of interrogative sentences like "Who bought what?", "What did Linlin drink?" and "What did Haohao do to the biscuits?" with situations illustrated in pictures, and the task was to choose a correct and natural answer from two options, which corresponded to V O *de* and V *de* O forms. In multiple interrogative sentences, 38 (76%) informants chose V *de* O answers over the V O *de* answers. In object interrogatives, 47 (94%) preferred V *de* O answers, and in the predicate interrogatives the same preference was shown by 43 (87.8%) informants. In sum, the predicate or the object bearing an information (presentational) focus, which does not sound natural in V O *de* order, in V *de* O order is well accepted by the informants. The slightly lower preference rate of V *de* O order in multiple interrogatives than those in other structures can be explained by the fact that the possible pseudocleft reading of V O *de* order could make it sound less odd as an answer. We thus conclude that a genuine predicate or object cleft sentence should take the configuration (b), V *de* O order.

The clefted predicate or object bears a prosodic prominence, which disambiguates the reading since the two types of clefts appear to have the same surface structure, as indicated in (43a,b). The object cleft in (43b) is the only possible case in which the focused constituent is not adjacent to *shì*.

(43) a. Wǒ   shì   **mǎi**   de   shū (, bú shì   jiè   de shū).   predicate cleft

1SG   be   buy   DE book (NEG be borrow DE book)

'I bought the book (, and I did not borrow it).'

b. Wǒ shì   mǎi   de   **shū** (, bú   shì   bǐ).   object cleft

1SG be   buy DE   book (NEG be   pen)

'It was the book that I bought (not the pen).'

Hereinafter, the *shì* XP V O *de* cleft will be referred to as V O *de* cleft, and the *shì* (XP) V *de* O cleft as V *de* O cleft.

- (subject) V (O) *de* (X) *shì* **NP** pseudocleft construction

In subject and direct object pseudoclefts (44a-b), an element X that shares features with the focused element optionally appears immediately following *de*. An indirect object pseudocleft, however, requires the phonetic realization of the element X (44c). When the clefted constituent is an argument, the shift of positions of the focused constituent and the sequence of (subject) V (O) *de* (X) will not affect the interpretation of the sentence (44a-c). An adjunct or a predicate would be nominalized in order to be focused upon (45a-b).[21]

(44) a. Mǎi   shū   de   (rén)   shì   **wǒ.**

buy   book   DE (person) be   1SG

'Who/The person that bought the book is me.'

b. Wǒ   mǎi   de   (dōngxi) shì   **shū**.

1SG   buy   DE (thing)   be   book

'What/The thing I bought is a book.'

---

[21] There is a construction similar to "inverted pseudocleft" that displays more complex properties. When the clefted constituent is the subject, when the element X is omitted, the inverted pseudocleft conveys an identical interpretation to that of a pseudocleft (i).

(i) a. **wǒ**   shì   mǎi   shū   de.
     1SG   be   buy   book   DE
   'Who/The person that bought the book is me.'

However, when the focused element is a constituent other than the subject, e.g., an object (iia), an adjunct (iib), due to the structural similarity to *shì…de* cleft, the omission of the element X entails a focus reading of the element immediately after *shì*, and the intended focus, i.e. the element before *shì*, in turn, becomes the topic:

(ii) a. Shū shì wǒ   mǎi   de
          book be 1SG   buy   DE
        '(As for the book,) it is I that bought the book.'
      b. Lǐsīběn shì Xiǎomíng   qù   de.
           Lisbon be   Xiaoming go DE
        '(As for Lisbon), it is Xiaoming who went there.'

To maintain the pseudocleft reading, the element X cannot be omitted in this construction. As the acquisition of inverted pseudoclefts will not be explored in this work, we will leave such construction out of our discussion.

c. Wǒ jiè gěi tā shū de *(rén) shì Míng.

1SG lend to 3SG book DE (person) be Ming

'The person to whom I lent the book is Ming.'

(45) a. Xiǎomíng qù de (dìfāng) shì **Lǐsīběn**.

Xiaoming go DE (place) be Lisbon

'Where Xiaoming went is Lisbon.'

b. Wǒ zuo de (shìqing) shì **kàn shū**.

1SG do DE (thing) be read book

'What/The thing that I do is reading a book.'

All instances of *shì* in clefts or pseudoclefts behave like a copular verb instead of an emphatic morpheme according to the criteria of Huang, Li and Li (2009). It is not stressed, it can occur in A-not-A questions[22](46a) and be preceded by modal auxiliaries (46b), and the sentential negation markers cannot precede other element than *shì* (46c). *Shì* in a cleft can be omitted in affirmative sentences (46d), while, in a pseudocleft, it is sometimes kept to avoid ambiguity[23].

(46) a. Shì bù shì **nǐ** mǎi de shū/ mǎi shū de shì bù shì **nǐ**?

be NEG be 2SG buy DE book/buy book DE be NEG be 2SG

'Is it you that bought the book?/Who bought the book was you?'

b. Xiǎomíng <u>kěnéng shì qù de</u> / <u>qù de kěnéng shì</u> **Lǐsīběn**.

Xiaoming maybe be go DE/ go DE maybe be Lisbon

'It may be Lisbon that Xiaoming went to./Where Xiaoming went may be Lisbon.'

c. Xiǎomíng <u>bú shì qù de</u> / <u>qù de bú shì</u> **Lǐsīběn**.

Xiaoming NEG be go DE / go DE NEG be Lisbon

'It is not Lisbon that Xiaoming went to./Where Xiaoming went is not Lisbon.'

---

[22] A type of Chinese disjunctive question involving three subtypes: "V-not-VP", "VP-not-V" and "VP-not-VP", as defined by Huang, Li and Li (2009, pp. 242-250).

[23] The potential ambiguity lies in the fact that an NP V *de* O sentence without *shì* has two different readings, a pseudocleft reading and a relative clause reading (i). In contrast, "Wǒ mǎi de shi **shū**" is unambiguously a pseudocleft and could only have the 1st reading in (i).

(i) Wǒ mǎi de **shū.**

1SG buy DE book

1st reading: What I bought was a book.

2nd reading: The book that I bought.

d. Xiǎomíng (shì)   qù de **Lǐsīběn**.

   Xiaoming (be)   go DE Lisbon

   'It was Lisbon that Xiaoming went to.'

-Bare-*shì* subject cleft construction

In line with Paul and Whitman (2008), a sentence-initial bare-*shì* construction is considered as a cleft. This construction focalizes the subject that immediately follows *shì*, without altering the original word order, as shown in (47a). Elements other than the subject cannot immediately follow *shì* and be clefted (47b-c).

(47) a. Shì **Xīqūkēkè** yào   pāi   qúnniǎo.

   be Hitchcock want shoot The Birds

   'It was Hitchcock that wanted to make The Birds.'

   b. *Shì **zài   měiguó** xīqūkēkè   pāi-le   qúnniǎo.

   be   LOC America Hitchcock shoot-PFV The Birds

   *'Hitchcock did make The Birds **in America**.'*

   c. *Shì   **qúnniǎo**   xīqūkēkè   pāi-le.

   be   The Birds   Hitchcock shoot-PFV

   'Hitchcock did make **The Birds.'*

2.2.2 Previous Analyses of MC Clefts

This section is divided into two subsections. The first subsection consists of the main perspectives used to analyze *shì* in clefts. The second subsection describes various proposals regarding *de* and, consequently, analyses of cleft structures in the literature.

2.2.2.1 Analyses of Shì in MC clefts

There have been at least three different kinds of analyses for *shì* in MC clefts*:* a) as a focus marker or operator; b) as an auxiliary verb (or "higher predicate"); and c) as a main verb. Within each of these different types of analyses, there is more than one diverging proposal.

As the perspective (a), i.e. taking *shì* as a focus marker without syntactic functions (Teng, 1979)

or an operator with the status of an adverb (Huang, 1998), neglected the verbal status of *shì*[24], and since arguments for its verbal status were presented by Paul and Whitman (2008) and were previously mentioned in this chapter, see section 2.2.1, this type of analysis will not be presented here in detail. I will therefore consider only the possibility of taking *shì* as a verb, either an auxiliary (hypothesis b) or a main verb (hypothesis c).

Shì as an auxiliary verb

Huang (1988) claimed that *shì* in cleft sentences is an auxiliary verb, which allows the raising of the subject, while Simpson and Wu (2002) considered *shì* to be a higher verbal/auxiliary element V^0, which selects for the TP headed by *de*.

- Raising auxiliary verb: Huang (1988)

Huang (1988) analyzed *shì* in a bare-*shì* cleft sentence as an auxiliary verb having the status of a one-place predicate in that it does not select its external argument and thus permits the raising of the subject[25]. The author considered the focus to be the first constituent following *shì*, in other words, the constituent directly governed by *shì*. In (48a), the subject *wǒ* does not raise and is focused upon as it immediately follows *shì*. In (48b)[26], the subject raises to the specifier of the higher IP and hence, *zuótiān* (yesterday) becomes the clefted constituent.

(48) a. [IP [NP *e* [I' Shì [IP wǒ   zuótiān   dǎ-le   tā]]]].

    be   1SG yesterday hit-PFV 3SG

    'It was me that hit him yesterday.'

   b. [IP [NP Wǒi [I' shì [IP ti   zuótiān   dǎ-le   tā]]]].

---

[24] Teng (1979) considered *shì* as the phonetic spell-out of a diacritic feature [+focus] inserted in a simple sentence. The author also pointed out that his analysis faces an "insurmountable difficulty": the properties of *shì* in simple sentences are identical to those of the main verb in clefts in that it allows negation and can enter the scope of adverbs. Huang (1998) proposed that *shì* in a bare-*shì* cleft is a focus operator having the status of a quantificational adverb on a par and in scope relations with modals, negation, etc., and proposed a LF account of clefts. However, Zhan & Sun (2013) refuted this adverb analysis of *shì* by defending its verbal status. See Huang (1998) and Zhan & Sun (2013) for details.

[25] According to Huang (1988), the raising is not compulsory for some auxiliary verbs, such as *kěnéng* (may) and *yīnggāi* (should), and the subject can occur before or after them, which is also the case of *shì*.

[26] As analyzed by Paul and Whitman (2008), sentences like (48b), namely, medial bare-*shì* pattern, are not considered to be clefts but an association with a focus pattern. This *shì* functions as a focus operator and parallels the behavior of association of focus with the "emphatic" *do* in English.

1SG   be   yesterday hit-PFV   3SG

'It was yesterday that I hit him.'

(Huang, 1988, p. 51)

- Higher verb/auxiliary selecting a TP: Simpson and Wu (2002)

Simpson and Wu (2002) analyzed *de* in past-tense *shì...de* constructions as the head of T (T_2) and treated *shì* as a higher verbal/auxiliary element that selects the TP (TP_2) headed by *de*. The authors compared the *shì...de* structure with the *have -en* perfect tense form in English and, due to the similarity of the time reference of these structures in the two languages, in that the event time is interpreted as being prior to and having a clear relevance to the speech time, Simpson and Wu (2002) assumed that the past tense (event time) is encoded in *de*, while the present tense (speech time) is encoded in *shì*. The presence of *shì*, as analyzed by the authors, would be "similar to the common (optional) use of the English auxiliary verb 'do' to cause readings of emphatic focus" (Simpson & Wu, 2002, p. 198). *Shì* would be located in the V^0 in the higher TP, as illustrated in (49).

(49) $[_{TP} \text{Wǒ}_k [_{T'1} T^0{}_k [_{VP} [_{V'} \text{shì} [_{TP2} < \text{zuótiān pro}_k \text{lái}>_i [_{T'2} \text{ de } [_{AspP} t_i]]]]]]]]$

1SG   be   yesterday   come   DE

"It was yesterday that I came."

(Simpson & Wu, 2002, p. 197)

Shì as a main verb

Shì being a main verb is reckoned in different analyses, including those that treat *shì* as a copular verb and those that consider it to be a main verb other than the copula in declarative sentences[27]. Here, we present more recent analyses within a generative framework: Huang (1998), Cheng (2008), Paul and Whitman (2008) and Hole (2011).

-Huang (1998)

Huang (1998) considered *shì* in pseudocleft sentences as a copular verb. The author stated that

---

[27] In traditional analyses, proposals by scholars like Chao (1968) took *shì* as the copula and the cleft sentence as a derivation from an equation sentence; Zhu (1978) recognized the predicational and identificational function of *shì*; Paris (1979) proposed that *shì* is the copular verb in the underlying structure of the *shì...de* construction and takes the remaining part of the sentence, which is nominalized by the nominalizer *de*, as its subject; Tang (1980) argued that *shì* in a cleft sentence is not a copula but rather a judgment verb, due to its assertive function. See their works for details.

pseudoclefts involve structural gaps and that there is an overt structural dependency between the focus and the gap.

-Cheng (2008)

Cheng (2008) argued that there is no actual "*shì…de*" construction and sentences containing both *shì* and *de* can have distinct structures, but all involve the copula taking a small clause. For the general base-structure involving both *shì* and *de*, Cheng (2008) considered *shì* as the copular verb taking the whole *de*-clause as its subject and a null pronominal *pro* as its predicate. In bare-*shì* sentences, *shì* is also considered as a copula taking a small clause. There are, as the author assumed, two sources of focus in *shì…de* sentences, one being the copula *shì*, and the other the phonological prominence.

Cheng (2008) proposed that sentences containing both *shì* and *de* and pseudoclefts are derived from the same copular sentence. The author argued that *shì*, as a copula, always takes a small clause consisting of a subject-predicate structure, whereby the canonical predication structure is shown in (50a), and subsequently, the subject raises to the matrix Spec, IP, deriving (50b).

(50) a. shì [SC [SUBJ Zhāngsān] [PRED xuéshēng]]

      COP   Zhangsan   student

    b. Zhāngsān shì xuéshēng.

     Zhangsan COP student

    'Zhangsan is a student.'

(adapted from Cheng, 2008, p. 15)

The author considered focusing as identification and that a *pro*-predicate goes through predication inversion in simple identification sentences. In broad *shì…de* configurations, the *de*-clause is hence treated as the subject of the small clause, while the null pronominal *pro* is the predicate. The broad focus of a broad *shì…de* sentence is expected, as shown in (51), with the whole *de*-clause being focused upon, and with a base-structure as shown in (51a), a sentence like (51b) is derived. Narrow focus, according to the author, is indicated by phonological prominence. What remains to be solved is the impossibility of changing the canonical word order in broad *shì…de* cases and the potential focus reading on the subject within the *de*-clause.

(51) a. *pro*ᵢ shì [SC [SUBJECT *de*-clause] [PREDICATE tᵢ]]

    b. *pro*ᵢ shì [SC [SUBJECT Zhangsan came by train yesterday DE] [PREDICATE tᵢ]]

(adapted from Cheng, 2008, p. 21)

As for a "'pseudocleft' sentence"[28], Cheng (2008) again took up Moro's (1997) analysis and claimed that it is also derived from the same base small clause as the canonical predication sentences (52a), and that, in a "pseudocleft", it is the predicate that undergoes movement and raises to a pre-copular position, exhibiting "inverse" predication (52b). Moreover, the author stated that the inverted subject-predicate order yields focus on the element immediately following the copula.

(52) a. shì [[SUBJ Xiǎowáng] [PRED zuótiān wǎnshàng lái de]]

　　　 COP　　　 Xiaowang　　　 yesterday night come DE

　　 b. [PRED zuótiān wǎnshàng lái de] shì [SUBJ Xiǎowáng]

　　　　 'Who came last night is Xiaowang.'

(adapted from Cheng, 2008, p. 19, p. 17)

Nonetheless, the author left the reason for the unique interpretation in an inverse predication unexplained. Regarding focus, Cheng (2008) reckoned that in a canonical predication, a phonological prominence stresses a particular element, otherwise the whole post-copular constituent is focused. Again, the author suggested that in diverse cases containing both *shì* and *de*, a phonological prominence can shift the focus.

-Paul and Whitman (2008)

Paul and Whitman (2008) analyzed *shì* in *shì…de* cleft sentences as the matrix copular verb in V. This *shì* can be negated, preceded by auxiliaries, and form an A-not-A question in the manner of the copular verb *shì* in other contexts. Hence, the authors consider it as a copula, as instantiated in (53). In a sentence like (53) where *shì* is in the sentence-initial position, only the subject can bear the focus reading.

(53)　 *shì…de* cleft:

　　 [TP [vP Shì [AspP jiějie [Asp' kāi+de [vP tjiejie [v' V [VP tv　 mén]]]]]]]

　　　　 be　 elder.sister　 open DE　　　　　 door

　　 'It was the elder sister who opened the door.'

(Paul & Whitman, 2008, p. 18)

---

[28]　 The author doubted the existence of pseudoclefts in MC. Cheng (2008) considered that the pivots of pseudoclefts cannot be a *wh*-phrase, as in English, however, in MC in such sentences a *wh*-phrase as the pivot is allowed. Meanwhile, these sentences exhibit connectivity effects.

The sentence-initial bare-*shì* pattern[29] i.e. the bare-*shì* subject clefts, can either strongly assert the entire proposition or focalize the subject. The whole sentence may be negated or converted into a question, including an A-not-A one, and *shì* is thus reckoned as a copular verb. When a sentence in this pattern is interpreted as a subject cleft with a phonological stress on the subject, it obeys the exclusiveness condition and focalizes the constituent in the position immediately following *shì*, evidencing that this pattern qualifies as a cleft structure. This *shì*, as claimed by the authors, selects a finite TP, to which the focused subject is dislocated through A-movement. The finiteness of the TP accounts for the inexistence of further subject raising, as shown in (54).

(54) sentence-initial *shì* pattern (subject cleft only):

$[_{TP} [_{vP}$ shì $[_{TP}$ dìdi$_t$ $[_{T'}$ T^0 $[_{vP}$ t$_{didi}$ xǐ pánzi$]]]]]$

be younger.brother wash dish

'It was the younger brother that did the dishes.'

(Paul & Whitman, 2008, p. 23)

-Hole (2011)

Hole (2011) categorized *shì* in *shì…de* clefts as a copular verb in V^0 selecting a CP, considering the restricted tense interpretation and capacity to host TAM elements of *shì* V *de* O clefts. The author reckoned *shì* in such pattern as a copula with restricted default TAM values [+anterior, -irrealis, +terminative], with present-perfect-like implications for the overall structure.

2.2.2.2 Analyses of the de and the shì…de construction

The nature of *de* in the *shì...de* cleft construction[30] is strictly related to the derivation of cleft

---

[29] As analyzed by Paul and Whitman (2008), we maintain that in the medial bare-*shì* pattern (in the terminology of the authors), in which the subject precedes *shì*, *shì* functions as a focus operator. When *shì* does not bear a phonetic stress, any element to the right of *shì* can be interpreted as the focus provided that it is phonetically stressed; when *shì* is phonetically prominent, what the authors defined as the veridicality of the whole proposition is asserted. Additionally, this pattern is not subject to the exclusiveness condition. This pattern thus cannot be treated as a cleft as in the analysis of Huang (1988), but rather a strategy of association with focus.

[30] It is conventionally accepted that the morpheme *de* can subordinate modifiers to the nominal head (e.g., Paul, 2005) and hence appears in possessive constructions, attributive constructions, and relative clauses, among others. This property of *de* also causes the ambiguity mentioned in Footnote 23. As analyzed by Paul & Whitman (2008), we take the position that *de* in *shì...de* clefts differs from *de* in the other aforementioned structures.

sentences in the literature.[31] Here, we present analyses in more recent frameworks of generative grammar that classified *de* as different functional heads (the Minimalist Program: from D^0 to T^0 - Simpson & Wu, 2002; Assertion^0 - Cheng, 2008; Asp^0 - Paul & Whitman, 2008; Distributed Morphology: C^0 - Hole, 2011).

a. -Transitional state from D^0 to T^0: Simpson and Wu (2002)

Simpson and Wu (2002) concentrated on the nature of *de* and argued that *de* is undergoing a D-to-T conversion, changing its category from an original source as a D^0 element to a new past tense morpheme in T^0 and consequently cliticizing to the verb, as an instance of "lateral grammaticalization".

Simpson and Wu (2002) claimed that, in clefts, the *shì* V *de* O order is derived diachronically from the *shì* V O *de* order.[32] The authors proposed that the derivation is caused by *de* moving leftwards from a base-generated clause-final position, which is considered as an instance of cliticization gradually narrowing the target for attachment.

Following previous analyses (e.g., Li & Thompson, 1981; Paris, 1979), Simpson and Wu (2002) suggested that the *shì…de* sequence forms an original Complex Noun Phrase (CNP) or a nominalization structure, in which *de* precedes a phonetically null head-noun (N).

Simpson and Wu (2002) claimed that, in V O *de* patterns, with a non-past reading, *de* is an enclitic determiner that occupies D^0 and selects a rightward clausal complement as in all CNP type structures (see 55). In line with Kayne (1994), in relative clauses, the relativized noun/NP raises to Spec, CP, followed by the remnant IP raising higher to Spec, DP, and such a remnant movement is triggered by the clitic property of *de*. Hence, the authors proposed the representation of Chinese CNPs as in (55a,b). As the clausal complement could contain aspectual markers, it can arguably be an AspP. The non-past *shì…de* structure is suggested (55b).

(55) a. [$_{DP\ IP}$ <wǒ   zuótiān mǎi t$_i$>$_k$ [$_{D'}$ de [$_{CP}$ shu$_i$ [$_{C'}$ [$_{IP}$ t$_k$]]]]]

             1SG   yesterday buy   DE   book

---

[31] Although Chao (1968) left the nature of *de* undefined, this morpheme has been analyzed as a particle (Hashimoto, 1966; Zhu, 1978; Tang, 1980), a past tense marker (Teng, 1979), a perfective suffix (e.g., Huang, 1998), a nominalizer in the terminology of transformational grammar (Paris, 1979; Zhan & Sun, 2013). See their work for detailed discussions.

[32] However, this analysis of diachronic derivation of Simpson and Wu (2002) was refuted by the diachronic evidence in Long and Xiao (2011), which indicated that the first emergence of (b) order in the Song Dynasty (11th -13th C.E.) precedes that of (a) order in the Yuan Dynasty (13th -14th C.E.).

'The book I bought tomorrow.'

b. [IP NP tā [I' [VP [V' shì [DP AspP/IP< míngtiān cái huì qù Běijīng>i [D' de

        3SG   be   tomorrow only will go   Beijing   DE

[NP [N' N ∅ [AspP/IP ti]]]]]]]]]]]]

    'It's only tomorrow that I will go to Beijing.'

<div align="right">(Simpson & Wu, 2002, p. 186, p. 189)</div>

In both V *de* O and V O *de* patterns, only when *de* has a "past time interpretation", in the terminology of the authors, can adverbs and *wh*-adjuncts appear after *shì*, as shown in (56).

(56) a. Zuótiān   nǐ   shì wèishěnme   lái xúexiào de?

       Yesterday 2SG   be   why   come school DE

    'Why is it that you came to school yesterday?'

  b. Nǐ   qùnián   shì wèishěnme/zěnme qù de Běijīng?

      2SG last.year be   why/how   go DE Beijing

'Why/how did you go to Beijing last year?'   (Simpson & Wu, 2002, p. 184)

The authors, therefore, assumed that in V *de* O patterns there is a simple TP, which shows no island effect for *wh*-adjuncts nor restrictions on external adverbial modification. They also stated that, in the absence of any non-past/future/modal elements, the occurrence of *de* in *shì…de* forms makes the past time interpretation the only possible reading, and that when a past time reading is necessary, the occurrence of *de* is mandatory. Hence, they proposed a reanalysis of *de* as an enclitic currently undergoing a transformation process into a type T^0 past tense morpheme, and that, in a V *de* O pattern, *de* cliticizes to the verb as a more specific phonological host.

When the V O *de* form obligatorily has a past interpretation, *de* goes through a categorial switch from the CNP head D^0 to T^0 and directly selects for a clausal complement. Such clausal complement, arguably an AspP, raises to the specifier position of TP (TP₂) headed by *de* to provide phonetical support, as in (57).

(57) [TP IP<wǒ zuótiān mǎi piào>i [T' de [AspP ti]]]

      1SG yesterday buy ticket   DE

   'I bought the ticket yesterday.'

<div align="right">(Simpson & Wu, 2002, p. 190)</div>

As for the V *de* O form, the authors suggested that there is no difference from V O *de* form in the

underlying structure and that *de* alternatively targets the verb and moves leftwards. No representation was provided in Simpson & Wu (2002). Additionally, the authors assumed that *shì* projects a present tense, which licenses the occurrence of a TP headed by *de*. The subject then raises to the specifier position of a higher TP (TP1), as shown in (58).

(58) [TP Wǒ$_k$ [T'1 T^0_k [VP [V' shì [TP2 < zuótiān pro$_k$ lái>$_i$ [T'2 de [AspP t$_i$]]]]]]]

        1SG   be   yesterday   come   DE

        'It was yesterday that I came.'

(Simpson & Wu, 2002, p. 197)

b.-Assertion^0, Indicator of the presence of non-overt operators: Cheng (2008)

Cheng (2008) claimed that *de* can indicate the presence of two non-overt operators, namely, the λ-abstraction operator and an assertion operator. In this proposal, there is no extraction of clefted constituents in syntactic structures.

The author maintained that in cleft sentences *de* is on a par with the *de* associated with relative clauses[33] and is a realization of the null generalized λ-abstraction operator, which can bind an argument variable, an adjunct variable or an event variable.

The author proposed a canonical predication for cases in which both *shì* and *de* are involved, including relative clauses in which *de* marks the null generalized λ-operator and yields the predicate, the *de*-clause, as shown in (59).

(59) shì [[SUBJ Zhāngsān] [PRED *e* tiāntiān dōu chídào de]]

       be   Zhangsan   everyday all be.late *de*

       'Zhangsan is late everyday.'

Cheng (2008, p. 15)

In the meantime, according to the author, in bare-*de* (60a) and broad *shì-de* structures (60b), *de* marks the presence of an assertion operator related to sentential emphasis/focus, which is also not

---

[33] Chinese relative clauses, as summed up by Cheng (2008), are prenominal with *de* appearing between the relative clause and the head noun, and there is no relative pronoun present in Chinese relative clauses, as in (i):

(i) tā   chàng gē   de shēngyīn. (event variable)   (Cheng 2008, p. 7)

   3SG sing song DE   voice

   'the voice that he has while singing'

spelled out. The assertion operator takes the whole proposition as an argument and is hosted in AssertionP, of which *de* may be the head.

(60) a. tā   míngtiān huì   lái   zhǎo   wǒ de.

he tomorrow will come look.for me DE

'(It is the case that) He will come to see me tomorrow.'

(Cheng, 2008, p. 6)

b. shì   Xīlà   rén   zuì xiān kāishǐ niàng jiǔ de.   (all-new cleft)

be Greece person most first begin brew wine DE

'It is the Greek that first started to produce wine.'

(Cheng, 2008, p. 24)

As for the "broad *shì-de*" cases, in which *shì* and *de* are placed, respectively, in the initial and final position of the sentence, as it is the case in (59) and (60b), the author considered that they are derived distinctively from the canonical predication cases, in that broad *shì-de* sentences must have the canonical word order and the predicate type is similar to those of bare-*de* cases, which cannot be questions. Hence, the author suggested that *de* in broad *shì-de* sentences also marks the presence of an assertion operator. The post-copular clause of a broad *shì-de* sentence, as the author proposed, is expected to have two interpretations: the interpretation that the element immediately following *shì* is the focus, as in (59), and a broad focus interpretation, in which the whole clause after *shì* can provide new information, as in (60b).

Given this, the author stated that the two mechanisms that can lead to contrastive focus, movement and stress result in the same structure at LF (Cheng, 2008, p. 37). The disallowance of marking a narrow focus, except for the subject focus in broad *shì-de* sentences, was explained by the author in the following way: the *de*-clause is the subject of the small clause, which leads to the difficulty of moving across a subject island. The reason why the subject can have a focus interpretation is left unexplained.

The analysis of *de* as Assertion[0] also faces a problem. The argument Cheng (2008) provided for the presence of the assertion operator indicating sentential emphasis is that bare-*de* sentences cannot be questions; however, a bare-*de* question was well accepted by many MC speakers that we consulted, as in (61):

(61) shéi   xiě   nèi-běn   shū   de?

who   write that-CLF book   DE

'Who was it that wrote that book?'

<div align="right">(Cheng, 2008, p. 9)</div>

c. C⁰: Hole (2011)

Within the framework of Distributed Morphology, Hole (2011) suggested two operations for the derivation of V O *de* and V *de* O clefts, namely a PF movement of the complement of *de* for both structures, and an object shift preceding such PF movement to derive V *de* O clefts. According to the author, only V *de* O clefts have stricter temporal/aspectual/modal restrictions, which can be explained by the PF-true linearization constraints[34]. As for *de*, the author argued that *de* always heads a CP.

The derivation of V *de* O clefts is accounted for by object shift combined with remnant movement/prosodic inversion at PF[35].

The PF movement assumed here is a linearization operation that occurs after late vocabulary insertion, motivated by the enclitic nature of *de*, in which, X, part of the complement of *de*, moves around *de*, with a phonetically empty head Y_{empty} intervening between *de* and X, as shown in (62).

(62) *de* + Y_{empty}+X→X +*de* + Y_{empty} (Hole, 2011, p. 1720)

De in clefts is considered by the author as a C head and a presupposition marker that cliticizes to post-focal deaccented material[36]; thus, the PF-movement of the verb, together with preverbal adverbials, takes place. The author assumed the phonetically empty head to be $C^{\emptyset}$, which specifies

---

[34] In the light of Fox and Pesetsky's (2004) assumption that object shift may not alter the order regarding V and O, Hole (2011) assumed the following MC PF-true linearization constraints: (a) a TAM words>V order, which restates the generalization that TAM words precede full verbs in Chinese (Li & Thompson, 1981), and (b) a V>O ordering constraint employed for information-structurally neutral objects in clefts ('>' symbolizing linear PF precedence). As these orders cannot be altered, after the object shift, there is no position preceding V for TAM words and, therefore, these materials are excluded from V *de* O clefts.

[35] The object shift proposed by the author consists of 2 steps. In the first step, the object raises to Spec, AspP. Subsequently, the remnant string underneath the object moves around *de*, as in (i).

    (i) a. [$_{AspP}$ shī$_i$...[$_V$ xiě t$_i$]]

          poem   write

      b. [(xiě)$_{PF}$ [C-de ...[$_{Asp}$ shī$_i$ ... (xiě)$_{spell-out}$]]]

        write   DE   poem

<div align="right">(Hole, 2011, p. 1721)</div>

[36] In Hole & Zimmermann (2013), the authors explicitly stated that a V *de* O structure cannot cleft the verb, and it must be deaccented to be the potential host for the cliticization of *de*, which contradicts the judgement of our informants.

de's categorial features[37] and is right adjacent to another C domain head, C*, which is occupied by *de*. *De* thus links the presupposition X to the C$^\emptyset$, yielding a [C*P [C*$^{0'}$ *de* [C$^\emptyset$P [C$^{\emptyset0'}$ Ø [X(P)]]]] structure. As *de* heads a CP, when the object shift does not occur, pronounced TAM material is allowed in its scope.

Hole (2011) proposed that the focus, PP *cóng déguó* (from Germany) in (63), has already moved to the specifier of *de* in the overt syntax to check its exhaustiveness feature. The copula takes the C*P as an argument. The structure at spell-out and also at LF is shown in (63).

(63) Zhāngsān shì [C*P [PP cóng déguó]i [C*' de [CP   Ø [TP ti [T' proj lái]]]]]]

　　　Zhangsan be   from Germany   DE   come

　　'As for Zhangsan, it was Germany where he came from.'　　　　(Hole, 2011, p. 1724)

Subsequently, only the embedded verb, *lái* (come) in (63), undergoes reordering at PF and lands before *de*. In a subject cleft, it is the subject that raises to the specifier of *de*. However, the author suggested no explanation for the impossibility of focusing an object in the V O *de* pattern.

d. – Asp^0: Paul and Whitman (2008)

Paul and Whitman (2008) analyzed *de* in *shì…de* clefts as the head of AspP selected by the copular verb *shì*.

Concerning *shì…de* proper (in the authors' terminology), they took up Chao's (1968) generalization that the *shì* … O *de* order is more preferable for southern speakers of MC, while northern speakers prefer to use the *shì* … *de* O order. The proposal of Paul and Whitman (2008) on the *shì…de* cleft construction is based on northern dialects due to this distinction; namely, they only tackled the *shì* XP V *de* O clefts.

Paul and Whitman (2008) proposed three descriptive generalizations that distinguish the *shì…de* proper from other patterns involving bare-*shì* or *shì…de*. First, materials above vP are banned in the presupposition of a *shì…de* cleft, such as negation, modal auxiliaries, tense/aspect markers (perfective -*le*, experiential -*guò* or durative -*zhe*…) or the quantificational adverb *dōu* (all). Second, only subjects and adjuncts can be focused by occupying the position immediately following *shì*. Third, a *shì…de*

---

[37] This C$^\emptyset$ belongs to a paradigm of C categories heading conditional, temporal, reason and other clauses (Hole & Zimmermann, 2013).

cleft proper only has a past tense reading as *de* is associated with the past tense. Hence, future-oriented temporal adverbs are excluded in *shì…de* clefts (Paul & Whitman, 2008, p. 17). Moreover, it is observed that, in a cleft, only activity verbs are allowed.

Such restrictions are not found in another *shì…de* configuration that does not contain any focused element but expresses the assertion of the whole proposition, corresponding to a "broad focus" in Cheng (2008)'s term. This pattern is not considered as a cleft but in fact a "propositional assertion" by Paul and Whitman (2008)[38].

A *shì…de* cleft in MC, according to the authors, is derived by an A-type movement instead of the A-bar type, contrasting with the *it*-cleft in English. The authors insisted on a bipartitioning analysis of the focus and the presupposition and suggested identifying *de* as the head of an AspP, a position immediately above the base position of the subject in Spec, vP of the main verb, below *shì* and certainly below TP, by virtue of the temporal properties of *de* (64a) and the incompatibility with negation or auxiliaries (64b,c).

(64) a. Shì tā   dǎpò (*-le)   de bēizi.

      be 3SG smash -PFV DE cup

      'It was him who smashed the cup.'

   b. * Shì   dìdi   bù/méi   xǐ   de pánzi.

      be younger.brother NEG   wash DE plate

      'It was younger brother who didn't do the dishes.'

   c. * Shi Lisi neng/yinggai kai de men.

      be Lisi   can / must open DE door

      'It was Lisi who could/had to open the door.'

(Paul & Whitman, 2008, pp. 13-14)

The authors proposed that *shì* in VP selects this AspP, and the lexical verb raises through v, left adjoining to *de* in Asp^0. The proposed representations of *shì…de* clefts are shown in (65a,b).

---

[38] Paul and Whitman (2008) stated that, in the propositional assertion *shì…de* pattern, *de* has a function and syntactic position that are both distinct from those of *de* in cleft sentences. Given that this pattern does not face the constraints faced by a *shì..de* cleft, the projection headed by this *de* should be higher than TP. The authors also argued that *de* cannot be a head of a complex NP as analyzed by Simpson and Wu (2002), on the grounds that TP allows extraction, which is not possible in a complex NP. Paul and Whitman (2008) thus proposed that this *de* in propositional assertion is the head of *De*P in split CP following Rizzi (1997).

(65) a. [TP [VP Shì [AspP jiějie [Asp' kāi+de [VP tjiějie [v' v [VP tv mén]]]]]]]

        be   elder.sister   open DE   door

"It was the elder sister who opened the door."

b. [TP Lǔ Xùn [VP shì [AspP tLu Xun [AspP shěnme shíhou [Asp' xiě+de [VP tLu Xun [v' v [VP tv

    Lu Xun   be   what   time   write DE

  A Kiu?]]]]]]]]

  AQ

'When was it that Lu Xun wrote A Q?'

<div align="right">(Paul & Whitman, 2008, pp. 18-19)</div>

To trigger the obligatory raising of the subject to Spec, AspP, the Asp^0 must bear an EPP or OCC feature (Chomsky, 2004). In an adjunct cleft, the subject agrees with and raises to T in order to check the EPP feature, leaving its adjunct *shěnme shíhou* (when) in Spec, AspP[39] (see (65b)). In the case of subject clefts (65a), the subject is case-licensed by *shì* and hence stays in AspP. The Spec, AspP, hence, can either be an A-position or an A'-position. The authors also argued that, in MC, the landing site for topicalization is TopP, and that the landing site for object fronting is a position below TP but above negation (Paul, 2002, 2005), neither of which is in AspP. Therefore, the object cannot be clefted in this pattern.

Here, we summarize some essential aspects of the above presented analyses.

Shì is agreeingly considered to be a copular verb in all MC cleft structures, with two exceptions: Huang (1988) analyzed it as an auxiliary verb in bare-*shì* subject clefts, and Simpson and Wu (2002) took it as a higher verbal or auxiliary element in *shì...de* clefts.

The derivation of bare-*shì* subject clefts, in Huang (1988), was based on the optional subject raising to higher IP, and in Paul and Whitman (2008) was claimed to result from A-movement of the subject to Spec, TP selected by *shì*.

The structure of pseudoclefts was considered by Huang (1998) as involving a free RC, and by Cheng (2008) as an inverse predication.

---

[39] Adopting the assumption that "the position that intermediate traces are deleted (Lasnik & Saito,1992) or that multiple specifiers occupied only by traces are deleted (Moro, 2000)" (Paul & Whitman, 2008, p. 19), the authors stated that the focused constituent occupies the outmost Spec, AspP that is PF-adjacent to *shì*.

Various proposals have been put forward on the status of *de* in *shì…de* clefts.

Simpson and Wu (2002) considered *de* as an enclitic determiner in a diachronic process of becoming a past tense morpheme and argued that *shì* V O *de* clefts can have past or non-past interpretation. In non-past *shì* V O *de* clefts, the authors claimed that *shì* selects a DP headed by *de*, which in turn selects a null-head CNP, and that, from within such CNP, an IP/AspP containing the focused constituent raises to Spec, DP. In past *shì…de* clefts, according to the authors, *shì* selects a TP headed by *de*, and *de* selects an AspP, which raises to Spec, TP. The authors proposed a re-analysis of *de* from a D^0 element to a T^0 element cliticizing to the verb, resulting in the *shì* V *de* O form. No partitioning of the focus and the presupposition was suggested in this analysis.

Cheng (2008) took *de* as a λ-abstraction operator and assertion operator marker ($Assertion^0$), whilst left the question of yielding the sentence-final position of *de* as well as the properties and nature of this operator open. The author maintained that *shì* V O *de* clefts involve a canonical predication, in which the subject raises from a small clause selected by *shì*. No analysis of *shì* V *de* O clefts was proposed.

Hole (2011) considered *de* as a C head, being a presupposition marker that cliticizes to post-focal material. The focus was claimed to raise to the specifier position of such CP selected by *shì*. The derivation of *shì* V *de* O clefts, according to the author, results from object shift combined with remnant movement or prosodic inversion at PF.

Paul and Whitman (2008) analyzed *de* as an aspectual head with temporal properties and claimed that, in a *shì* V *de* O cleft, the focus raises to the specifier of this AspP headed by *shì*. As for the sequence of *de* and the verb, the authors proposed that the lexical verb left-adjoins to *de* in Asp^0. The analysis of *shì* V O *de* clefts was not explored by the authors.

2.2.3 Analysis of cleft structures in MC

In accordance with proposals such as Paris (1979), Huang (1998), Paul and Whitman (2008), among others, we assume that the three cleft structures at stake, bare-*shì* subject clefts, pseudoclefts, and *shì … de* clefts, are derived in distinct ways due to their diverse syntactic properties.

Apart from the distinctive restrictions on which constituents can be clefted, the three cleft

structures diverge in terms of TAM constraints. Certain TAM material, such as future-oriented adverbs, atelic verbs, as well as negation, which is closely related to tense, are banned in *shì…de* clefts, as shown in the following (a) examples, while such constraints are absent in the formation of bare-*shì* subject clefts, as in (b) examples below, and in pseudoclefts, as in (c) examples:

(66) Future-orientation:

    a. *Shì Xiǎowáng míngtiān   qù   shànghǎi de/qù   de shànghǎi.

       be Xiaowang tomorrow go Shanghai DE/go DE Shanghai

     'It is Xiaowang that will go to Shanghai tomorrow.'

    b. Shì Xiǎowáng míngtiān qù shànghǎi.

       be Xiaowang tomorrow go Shanghai

     'It is Xiaowang that will go to Shanghai tomorrow.'

    c.   Míngtiān qù Shànghǎi de shì Xiǎowáng.

       tomorrow go Shanghai DE be Xiaowang

     'Who will go to Shanghai tomorrow is Xiaowang.'

(67) Atelicity:

    a. *Shì Xiǎowáng huì tán gāngqín de/huì   tán de gāngqín.

       be Xiaowang can play piano   DE/can play DE piano

      'It is Xiaowang that can play piano.'

    b. Shì Xiǎowáng huì tán gāngqín.

       be Xiaowang can play piano

      'It is Xiaowang that can play piano.'

    c. Huì tán gāngqín de shì Xiǎowáng.

      can play piano   DE be Xiaowang

      'Who can play piano is Xiaowang.'

(68) Negation:

    a.*Shì Xiǎowáng méi   qù-guò   Shànghǎi   de/qù-guò de Shànghǎi.

       be Xiaowang NEG go-EXP Shanghai DE/go-EXP DE Shanghai

     'It is Xiaowang that has never been to Shanghai.'

    b. Shì Xiǎowáng méi   qù-guò   Shànghǎi

       be Xiaowang NEG   go-EXP Shanghai

'It is Xiaowang that has never been to Shanghai.'

　　c. Méi　qù-guò　　Shànghǎi de shì Xiǎowáng.

　　　NEG. go-EXP Shanghai DE be Xiaowang

　　　'Who has never been to Shanghai is Xiaowang.'

Hence, in spite of the seeming resemblance, a bare-*shì* subject cleft cannot be considered as a *shì…de* cleft omitting *de*, nor can a pseudocleft be taken as a variation of a *shì…de* cleft. Furthermore, we will argue that the nature of *de* in pseudoclefts is different from that of *de* in *shì…de* clefts. In what follows, we present the syntactic analyses for each structure adopted in this work.

Bare-*shì* subject cleft (sentence-initial bare-*shì* pattern)

　　A sentence in which *shì* occupies the sentence-initial position and *de* is absent can convey either the focalization of the subject when it is prosodically stressed (69a), or the assertion of the truth value of the entire proposition if the intonational prominence is on *shì* (69b), conveying a broad focus reading. No element other than the subject and *shì*, however, can receive any phonological stress in such pattern. In the second case, as in (69b), there is no singled-out element that could bear the focus, nor is there any presupposition part. This structure will be left out of the discussion of cleft structures.

　　(69) a. Shì **Xiǎomíng**　　xǐ-le　　　pánzi.

　　　　　be　Xiaoming wash-PFV　dish

　　　　　'It was Xiaoming that did the dishes.'

　　　b. **Shì** Xiǎomíng　　xǐ-le　　　pánzi.

　　　　　be　Xiaoming wash-PFV　dish

　　　　　'It is true that Xiaoming did the dishes.'

　　When the subject is intonationally prominent, in line with Paul and Whitman (2008), one can consider this structure (hereinafter designated as *bare-shì* subject cleft) as a cleft structure, as it (i) obeys the exclusiveness condition (70a) and (ii) exhibits the adjacency effect, given that the focus is only assigned to the position immediately following *shì*, namely the subject (70b).

　　(70) a. *Shì **Xiǎomíng**　xǐ-le　pánzi,　Xiǎohóng yě　　xǐ-le.

　　　　　 be Xiaoming wash-PFV dish　Xiaohong also wash-PFV

　　　　 **'It was Xiaoming that did the dishes, Xiaohong also did.'*

　　　b. *Shì Xiǎomíng　　xǐ-le　　　pánzi, méi　xǐ　wǎn.

be Xiaoming wash-PFV   dish   NEG wash bowl

*'It was Xiaoming that did the dishes, not the bowls.'

As Paul and Whitman (2008) observed and as presented above, *bare-shì* subject clefts allow tense or aspectual markers (e.g., perfective particle *-le*), modal auxiliaries, negation (71a) and future-oriented temporal adverbs in the part corresponding to the presupposition (71b). Additionally, as shown in (47b,c) and below in (71c), no other element, e.g., the object or the adjunct, can bear the focus.

(71) a. Shì   dìdi   bù   kěn   xǐ   pán-zi.

    be younger.brother NEG want wash   dish

    'It was the younger brother who didn't want to do the dishes.'

(Paul & Whitman, 2008, (66))

    b. Shì Xiǎomíng míngtiān kěnéng huì mǎi shū.

    be Xiaoming tomorrow maybe will buy book

    'It is Xiaoming that may buy a book tomorrow.'

    c. *Shì **pánzi/*zuótiān**   Xiǎomíng   xǐ-le.

    be dish/yesterday   Xiaoming wash-PFV

    'It was the dishes/yesterday that Xiaoming washed.'

To accommodate such elements generated in or below TP, *shì* must take a complement that has a size no smaller than TP. However, considering that only the subject can be focalized, it seems unmotivated to assume that the complement of *shì* is as large as a CP.

Taking into account the fact that the prosodic difference between two bare-*shì* initial patterns yields discrepancy in their interpretation, we assume that the head lexicalized by *shì* bears a focus feature and the scope of focus is determined and disambiguated by prosody.

(subject) V (O) *de* (X) *shì* **NP** Pseudocleft

Following Huang (1998), we maintain that the presupposition in a pseudocleft corresponds to a relative clause, due to the possibility and restrictions on the omission of the element X, which pattern with those of a null head NP in a relative clause in MC, namely, in a relative clause of indirect object (72a) or adjunct of [+reason]/[+manner] (72b,c), the head cannot be phonetically null.

    (72) a. Wǒ   jiè gěi   tā   shū   de*(rén)   shì Míng.

60

1SG lend to 3SG book DE (person) be Ming

'The person to whom I lent the book was Ming.'

b. Tā-men chǎojià de *(yuányīn) shì qián.

3-PL quarrel DE (reason) be money

'The reason why they quarrelled was money.'

c. Tā-men qù shànghǎi de *(fāngshì) shì zuò chuán.

3-PL go Shanghai DE (way) be take ship

'The way how they went to Shanghai was by ship.'

Consequently, the morpheme *de* in the presupposition of a pseudocleft is the same as *de* in relative clauses, more precisely, a head-final complementizer. According to Aoun and Li (2003), an MC relative clause is left-adjoined to the nominal which it modifies, this complex nominal can be projected as an NP, forming such a structure of the relative construction: [$_{NP}$ CP NP]. Supports for such an analysis come from the free ordering of relative clauses among themselves and the permission of NP, but not DP or CP, reconstruction effects. There are two types of relative constructions in MC. The first type is derived by movement, raising the nominal. The legitimacy of a head movement process is justified by satisfying the extension requirement, a derivational notion[40]. Such an argument NP hence raises to the head position to be relativized, leaving a gap in the relative clause, and this head of the relative construction can be null. The second type contains two constructions, adjunct relatives, and argument relatives in which a resumptive pronoun appears instead of a gap. Both constructions involve a relative operator and are derived by base-generation of the head, which cannot be null[41].

Furthermore, as Cheng (2008) observed, MC specificational pseudoclefts exhibit connectivity effects like specificational pseudoclefts in English. Reflexives (73a) and Negative Polarity Items (73c)

---

[40] Aoun and Li (2003) claimed that adjunction is no exception to extension (cf. Chomsky, 1995), which does not distinguish adjunction and substitution, and that a phrase marker is labeled after the creation of a position Ø, external to C' or C, inserted by generalized transformation (GT). Ø is then substituted by NP, a copy of a phrase inside C' or C, after which NP projects and the higher projection is determined, forming an adjunction structure. See Aoun and Li (2003, pp. 160-162) and Chomsky (1995/2015, pp. 171-175) for details.

[41] While the derivation of each type of relative clauses in MC is not the major concern of this work, the adopted analysis patterns with those of Aoun & Li (2003), Huang, Li & Li (2009), i.a.. The distinction of two types of relatives is based on the relevance of island conditions and the possibility of reconstruction. The categorial status of a relative construction, according to Huang, Li & Li (2009), can be an NP or a D(emonstrative)P, and in the presence of a number or a classifier, the projection must be larger than an NP. See Aoun & Li (2003) and Huang, Li & Li (2009) for detailed discussions.

are licensed, while coreference is prohibited (73b).

(73) a. Tā$_i$ xiāngxìn de   (rén)   shì zìjǐ$_{i/*j}$.

3SG believe DE (person) be   self

'(The person) who she$_i$ believes is herself$_i$.'

b. *Tā$_i$ hē   de (dōngxi) shì Hóng$_i$ mǎi de   jiǔ.

3SG drink DE (thing)   be Hong buy DE wine

'The thing that/ she$_i$ drinks is the wine that Hong$_i$ bought.'

c. Wǒ   méi   mǎi   de (dōngxi) shì rènhé yì-běn shū.

1SG NEG. buy DE (thing)   be any one-CL book

'What I did not buy was any book.'

Hence, in line with Heycock and Kroch (1999, 2002), one can assume that a MC specificational pseudocleft is a type of equative involving a small clause, which is selected by the copula *shì* and composed of a focalized element as its predicate and a relative clause as its sentential subject. Thus, the relative clause raises to Spec,T, leaving the predicate in the sentence-final position. Since a TP in MC is head-initial (Paul, 2002, 2005), the most deeply embedded node is in the sentence-final position, a default position for focus, according to Nuclear Stress Rule in the sense of Cinque (1993). The representation is presented in (74).

(74) a. [$_{TP}$ [$_{NP}$[$_{CP}$ [$_{TP}$ [$_{NP}$ rén/Ø] Mǎi   shū   de]] [$_{NP}$ rén/Ø]] [$_{T'}$ T^0 [$_{VP}$ shì [$_{SC}$ [$_{NP}$[$_{CP}$ [$_{TP}$ [rén/Ø]

buy   book   DE   person   COP

Mǎi   shū   de]] [$_{NP}$ rén/Ø]] **Xiǎowáng**]]]]

Xiaowang

'The person that/Who bought the book was Xiaowang.'

b. [$_{TP}$[$_{CP}$ Op$_i$ [$_{TP}$ Wǒ jiè gěi   tā$_i$ shū de]] [$_{NP}$ rén]$_i$] [$_{T'}$ T^0 [$_{VP}$ shì [$_{SC}$ [$_{NP}$[$_{CP}$ Op$_i$ [$_{TP}$ Wǒ jiè gěi

1SG lend to 3SG book DE   person   COP

tā$_i$ shū de]] [$_{NP}$ rén]$_i$] **Míng**]]]]

Ming

'The person to whom I lent the book was Ming.'

Shì … de cleft

62

Following Paul and Whitman (2008), we assume that *shì* in *shì...de* clefts is a copula verb, and *de* is the head of AspP. Differently from the authors, we propose a Focus projection below the VP headed by *shì*. This Focus projection is selected by *shì*, and dominates an AspP. In this structure, we therefore find two positions for focus: Spec, FP and the sentence-final position, the latter interpreted as the default focus position.

In both V O *de* and V *de* O orders, when the sentence is classified as a cleft, the verb in the presupposition part can only indicate a past event. The possible non-past interpretation of certain sentences with a seemingly identical superficial structure to a cleft may be due to the structural ambiguity of such sentences. A "NP *shì* (…) V O *de*" sentence, as noted by Paris (1979), Simpson and Wu (2002), among others, could be a) a cleft focalizing the constituent immediately following *shì*, or b) a generic proposition in the terminology of Paris (1979) or propositional assertion as labeled by Paul and Whitman (2008). When a non-past event occurs, the sentence cannot convey a cleft reading, as shown in (75a-b).

(75) a. Wáng shì míngtiān qù de.

Wang be tomorrow go DE

Int.:'It is the case that Wang is going tomorrow.'

Impossible int.:'It is Tomorrow that Wang is going.'

b. Wáng shì tán   gāngqín de.

Wang be play   piano   DE

Int.:'It is the case that Wang plays / can play the piano./Wang is a piano player.'

Impossible int.: 'It is playing piano that Xiaowang does.'

However, a "*Shì* NP (…) V O *de*" sentence only permits the cleft reading, focalizing the NP following *shì*, and, in this case, only a past interpretation is plausible and future-oriented adverbs are banned (76).

(76) a.*Shì Wáng míngtiān qù   de.

be Wang tomorrow go DE

'It is Wang that will go tomorrow.'

a'. Shì   Wáng zuótiān   qù   de.

be   Wang yesterday go   DE

'It was Xiaowang that went yesterday.'

Thus, we claim that the only possible reading for both V *de* O and V O *de* clefts in terms of temporal orientation is a past-tense reading.

Apart from the temporal constraint, as noted by Paul and Whitman (2008), in a *shì…de* cleft, any elements generated above vP are prohibited to occur between *shì* and *de*, such as negation (77a), modal auxiliaries (77a), and tense/aspectual markers (77b).

(77) a. *Shì Xiǎowáng yīnggāi / méi qù de fǎguó / fǎguó de.

     be Xiaowang should / NEG go DE France / France DE

     'It was Xiaowang who should go/ has never been to France.'

    b. Shì Xiǎowáng qù (*-guò/-le) de fǎguó/fǎguó de.

     be Xiaowang go -EXP/-PFV DE France/France DE

     'It was Xiaowang who went to France.'

The verb itself in a *shì…de* cleft faces constraints as well. Paris (1998) stated that such construction is sensitive to the aspectual value/semantic properties of the VP. Paul and Whitman (2008) pointed out that this structure only selects activity verbs, and Mai and Yuan (2016) maintained that the verb phrase carries a telicity feature, as shown above in (67a) and repeated in (78). Such properties would be expected under the assumption that *de* is an Asp^0 instead of a C^0.

(78) *Shì Xiǎowáng huì tán gāngqín de/huì tán de gāngqín. (= (67a) above)

    be Xiaowang can play piano DE/can play DE piano

    'It is Xiaowang that can play piano.'

Moreover, a *shì…de* cleft sentence allows extraction from the complement of *shì*, as demonstrated in (79). If *de* was a nominalizer heading a DP/NP, as claimed by Simpson and Wu (2002) and the sequence that follows *shì* formed a nominal structure, i.e. the complement of *shì* was a relative clause, the extraction of a constituent, e.g., the object in (79), would not be legitimate giving rise to an island effect.

(79) Lí$_i$, wǒ shì [jīntiān mǎi t_i de].

    pear 1SG be today buy DE

    '(As for) pears, it is today that I bought them.'

To sum up, given the TAM constraints on the cleft clause and aspectual restriction on the verb, and the island effects left unexplained by a nominal structure analysis, we assume that, in accordance with Paul and Whitman (2008), it is more plausible to consider *de* in *shì…de* clefts as an Asp head.

Paul and Whitman (2008) argued that only subjects and adjuncts can be clefted in *shì...de* clefts; however, empirical facts can justify that, in a V *de* O configuration, when there is no subject or adjunct immediately following *shì*, either the verb or the object will be clefted, when these constituents show intonational prominence. Examples (80-82) demonstrate the possibility of assigning focus to the verb or the object in *shì* V *de* O clefts, rather than in *shì* V O *de* clefts.

(80) – Línlin   hē-le   shěnme?

    Linlin drink-PFV what

'What did Linlin drink?'

a. – Línlin shì   hē   de kělè.

        Linlin be drink DE cola

b. */?? - Línlin shì hē   kělè   de.

        Linlin be drink cola DE

'It's cola that Linlin drank.'

(81) – Hàohao dùi bǐnggān zuò-le shěnme?

    Haohao to   biscuit do-PFV what

'What did haohao do to the biscuit?'

a. – Hàohao shì chī de bǐnggān.

    Haohao be   eat DE biscuit

b. */?? - Hàohao shì chī bǐnggān de.

        Haohao be   eat biscuit DE

    'It was eating the biscuit that Haohao did.'

(82) -Shéi   mǎi-le   shěnme?

    Who buy-PFV what

'Who bought what?'

a. -Shì Xiǎowáng mǎi de   huā.

    be Xiaowang buy DE flower

b. */?? - Shì Xiǎowáng mǎi huā   de.

        be Xiaowang buy flower DE

    'It is Xiaowang that bought the flowers.'

Hence, we conclude that in *shì* V O *de* clefts only the subject and the adjunct can be clefted, while

in *shì* V *de* O clefts, not only can the subject or the adjunct bear the focus, but also the verb itself or the object can be clefted.

We thus suggest a lower Focus projection selected by *shì* in line with Belletti (2004) for clefted subject and adjunct in both V O *de* and V *de* O configurations, as well as for some cases in which the verb itself is clefted, and we assume that the sentence-final position, coinciding with the most embedded position, would correspond to the default focus position and allows the object in V *de* O clefts to be interpreted as focused.

The lower FocP selects the AspP headed by *de*, which in turn selects a vP. As this FocP is a projection that needs to be activated once a focus is to be assigned, the embedded verb raises to Foc^0, whether focalized or not. Since we assume that no θ-role is assigned in this projection, Spec, FocP is an A'-position. In a subject cleft, the subject raises to Spec, AspP, an A-position, and then to Spec, FocP. The nominative case of the subject is then licensed via long-distance agreement with T. When an adjunct is focalized, it raises to Spec, FocP. The subject raises to Spec, AspP, and then to Spec, TP.

We should now explain the difference between *shì* XP V *de* O and *shì* XP V O *de* clefts. We take as a departure point the fact that only when the object remains in the most embedded position, following *de*, it can be clefted.

When the subject/adjunct bears the contrastive focus, only in *shì* XP V *de* O cleft can the object bear the informational focus, as shown in (83a,b). (83b), a *shì* XP V O *de* cleft, in which the object is not in the sentence-final position, is not a felicitous answer to the question. Also, in (83a), *zuótiān* "yesterday" should be phonologically more salient than other elements in the sentence, including the informational focus *lí* "pear", satisfying the Contrastive Focus Prominence Rule (Truckenbrodt 1995, Rooth 1996, *apud* Selkirk, 2008)[42]. Hence, only a *shì* XP V *de* O cleft can convey a contrastive focus reading on XP and, at the same time, an informational focus reading on the object.

(83) – Nǐ jīntiān mǎi-le shěnme shuǐguǒ?

2SG today buy-PFV what fruit

'What kind of fruits did you buy today?'

a. - Jīntiān méi mǎi. Wǒ   shì   zuótiān   mǎi de   lí.

---

[42] Contrastive Focus Prominence Rule (Truckenbrodt 1995, Rooth 1996b): Within the scope of a focus interpretation operator, the corresponding F-marked [contrastive focus] constituent is the most metrically prominent (Selkirk, 2008, p. 4).

today NEG buy 1SG COP yesterday buy DE pear

b. *⁄?? - Jīntiān méi mǎi. Wǒ   shì   zuótiān mǎi   lí   de.

today NEG buy 1SG COP yesterday buy pear DE

'I didn't buy any fruit today. It was yesterday that I bought some pears.'

In a subject or adjunct cleft, when the object does not bear either the contrastive/corrective focus or the information focus, it raises to a position of adjunction to AspP by scrambling, yielding the V O *de* order, as represented in (84a-b). This movement of the object is explained by defocalization.

(84) *shì* **XP** V O *de* cleft

a. XP = subject:

[TP T⁰ [VP Shì [FocP **Wáng** [Foc' mǎi [AspP shū [AspP ~~Wáng~~   [Asp'   de [vP ~~Wáng~~ [v' ~~mǎi~~ [NP ~~shū~~]]]]]]]]]]

COP   **Wang**   buy   book   DE

'It was Wang that bought the book.'

b. XP = adjunct:

[TP Wáng [T' T⁰ [VP shì [FocP **zuótiān** [Foc' mǎi [AspP shū [AspP ~~Wáng~~ [Asp' de [vP [AdvP ~~zuótiān~~] [vP ~~Wáng~~ [v' ~~mǎi~~ [NP ~~shū~~]]]]]]]]]]]]]

Wang   COP   **yesterday**   buy   book   DE

'It was yesterday that Wang bought the book.'

When the object does bear the information focus, since its original position, sentence-final, is a position for focus by default, constrained by the Economy of derivation, it obligatorily stays in this position and receives the focus, yielding the V *de* O structure, as in (85a-b).

(85) *shì* **XP** V *de* O cleft

a. XP = subject:

[TP [VP Shì [FocP **Wáng** [Foc' mǎi [AspP ~~Wáng~~ [Asp' de [vP ~~Wáng~~ [v' ~~mǎi~~ [NP shū]]]]]]]]]

COP   **Wang**   buy   DE   book

'It was Wang that bought the book.'

b. XP = adjunct:

[TP Wáng [VP Shì [FocP **zuótiān** [Foc' mǎi [AspP ~~Wáng~~ [Asp' de [VP [AdvP ~~zuótiān~~] [vP ~~Wáng~~ [v' ~~mǎi~~ [NP

Wang. COP   **yesterday** buy   DE

shū]]]]]]]]]]

book

    'It was yesterday that Wang bought the book.'

When no other candidate for a focused element intervenes between FocP and VP, both V and O can be interpreted as the focus. The whole VP forms a syntactic object bearing the focus, and the object thus stays in the original position, subject to the principle of economy of derivation (Chomsky, 1995). A V *de* O order is then shared by both a VP cleft and an object cleft. A focus shift hence occurs for disambiguation, and the verb or the object, once focalized, receives intonational prominence, represented in (86a,b) in capital letters.

(86) a. *shì* **V** *de* O cleft

    [$_{TP}$ Wáng [$_{VP}$ Shì [$_{Foc'}$ MǍI [$_{AspP}$ ~~Wáng~~ [$_{Asp'}$ de [$_{vP}$ ~~Wáng~~ [$_{v'}$ ~~mai~~ [$_{NP}$ shū]]]]]]]]

      Wang   COP   buy   DE   book

    'It was buying the book that Wang did.'

b. *shì* V *de* **O** cleft

    [$_{TP}$ Wáng [$_{VP}$ Shì [$_{Foc'}$ mǎi [$_{AspP}$ ~~Wáng~~ [$_{Asp'}$ de [$_{vP}$ ~~Wáng~~ [$_{v'}$ ~~mai~~ [$_{NP}$ SHŪ]]]]]]]]

      Wang   COP   buy   DE   book

    'It was a book that Wang bought.'

2.4 A note on English clefts

Since English is acquired as a foreign language by the majority of Chinese university students, as explained in the opening chapter (1.3), it is important to take into consideration the properties of English cleft structures, namely, the *it*-cleft and the pseudocleft.

The following description of English *it*-clefts is based on Dékány (2010). An English *it*-cleft sentence mainly consists of *it*, a "cleft pronoun", in the author's terms, a copula *be*, the clefted constituent, and an embedded clause. Similar to an EP standard cleft, the clefted element in an *it*-cleft must form a syntactic constituent, and generally, only DPs and PPs are cleftable. However, unlike the obligatory presence of the complementizer *que* in EP standard clefts or the *wh-* morpheme in EP *wh*-clefts, a cleft clause in English *it*-clefts can be headed by a null complementizer. The null

complementizer is nevertheless limited in use, occurring only in object *it*-clefts in informal situations[43]. As far as semantic properties are concerned, English *it*-clefts bear a presuppositional nature and the clefted constituent conveys an exhaustive identification.

The syntactic derivation of English *it*-clefts has been widely studied by linguists and various proposals have been made, such as Meinunger's (1998) monoclausal analysis, the specificational analysis (e.g., Hedberg, 2000), and Reeve's (2011) analysis of the cleft clause as a restrictive relative clause adjoined to the clefted constituent. Studies such as Chomsky (1977), Heggie (1993), Kayne (1994) and Kiss (1998) consensually maintained that the clefted constituent is situated in a position related to focus in the left periphery. Kiss (1998) assumed that the clefted constituent is an identificational focus associated with [+exhaustive, +contrastive] features in the left periphery. The cleft pronoun *it* is analyzed as an expletive and occupies the position of Spec, IP. I^0 selects a FP, in which the clefted constituent occupies the specifier position. F^0 is filled by the copula, which then raises to I^0. FP, in turn, selects an embedded CP, namely the cleft clause. The clefted constituent can be moved from the embedded VP across CP, as in (87a). When the clefted constituent is the subject, it must be base-generated in Spec, FocP and, at LF, be linked to a *wh*-pronoun that moves into the embedded CP to avoid ECP violation, and either the wh-phrase or the complementizer can regularly undergo deletion, as in (89b-b').

(87) a. $[_{CP} [_{IP}$ It $[_{I'}$ I$_{PAST}$ $[_{FP}$ to John$_i$ $[_{F'}$ be $[_{CP}$ t$_i$ $[_{C'}$ that $[_{IP}$ I $[_{I'}$ I$_{PAST}$ $[_{VP}$ speak $[_{PP}$ t$_i$]]]]]]]]]]]]

    b.   It is me$_i$ $[_{CP}$ who$_i$ Ø $[_{IP}$ t$_i$ is sick]]

    b'. It is me$_i$ $[_{CP}$ Ø$_i$ that $[_{IP}$ t$_i$ is sick]]

<div align="right">(Kiss 1998, pp. 258-259)</div>

One can thus notice the resemblance of the deep structures and derivations in the English *it*-cleft and the EP standard cleft, which are quintessentially the extraction of the clefted constituent and its dislocation to the left periphery.

Another English cleft structure that should also be taken into consideration in the L3 acquisition

---

of EP clefts is the pseudocleft structure. A basic English pseudocleft consists of a free relative, a copula verb *be*, and a focused element, which can be a noun phrase, adjectival phrase or verb phrase. Linguists such as Higgins (1979) have noticed that basic (or canonical) pseudoclefts exhibit connectivity effects in the same pattern that simple sentences do, in that they show binding effects (88a) and license Negative Polarity Items (88b). As mentioned in the previous chapter, Lobo (2006) observed the same behaviors in EP pseudoclefts.

(88) a. What Mary$_i$ was was proud of herself$_{i/*j}$.

b. He bought lots of textbooks; what he didn't buy was any good novels.

(Heycock and Kroch, 1999, p. 1)

Heycock and Kroch (1999) considered specificational pseudoclefts to be a case of a specificational copular sentence, more precisely, an equative sentence[44]. In Heycock and Kroch (2002), the authors maintained that an inverted pseudocleft, in contrast to the canonical pseudocleft, has an information structure of a topicalization, and the initial phrase is typically a topic but can also alternatively be a focus. Using this approach, not only the connectivity effect but also the anaphora resolution with *it/that* and the mandatory presence of the copula through gapping in pseudoclefts can be explained. An inverse predication analysis as in Moro (1997) was proven by Heycock and Kroch (1999) to be unnecessary and problematic. Hence, it would be more favourable to assume that it is always the subject that raises. A canonical pseudocleft and an inverted pseudocleft, before reaching the interface and LF, would have the representation as shown in (89a) and (89b), respectively, which is in line with Lobo's (2006) analysis of EP pseudoclefts.

(89) a. [[$_{CP}$ What Jane bought]$_i$ is [SC t$_i$ [a book]]].

b. [[A book]$_i$ is [$_{SC}$ t$_i$ [$_{CP}$ what Jane bought]]].

---

[44] The authors held that the equative semantic interpretation implies an interpretive procedure, ι-reduction, to take place at the syntactic derivation of the logical form of a pseudocleft, which reduces the Russellian operator and assigns the focus to the variable introduced by the free relative, yielding a logical form representation on a par with that of a canonical sentence. It is at the level of LF, as proposed by the authors, that the connectedness effect, as well as other discourse-related constraints, e.g., binding constraints, apply.

In Table I, we present a comparison and the central questions relevant to this study of derivation of cleft structures in the three languages under investigation, MC, English and EP.

MC		English		EP	
Shì…de cleft	Movement of clefted constituent to focus position: Subj., ~~Obj.~~, …	*It*-cleft	Movement of clefted constituent to focus position: Subj., Obj, ….	Standard cleft	Movement of clefted constituent to focus position: Subj., Obj., …
	[+focus] in the lower FocusP		[+focus] in the left periphery		[+focus] in the left periphery
	Null C: C domain not involved		Null C: possible in object cleft		Null C: impossible
Pseudocleft	Cleftable: Subj., Obj., …	Pseudocleft	Cleftable: Subj. Obj., …	Pseudocleft	Cleftable: Subj., Obj., …
	Small Clause		Small Clause		Small Clause
Bare-*shì* subject cleft	Cleftable: Subj. ~~Obj.~~			Semipseudo-cleft	Cleftable: Subj. (post-verbal), Obj., …
	movement of the clefted subject to the TP	-			*Ser* (be) as contrastive focus marker; no movement to the left periphery
-		-		*É-que* cleft	Movement of clefted constituent to focus position: Subj., Obj., …
					Lexicalization of [+focus] in the left periphery

Chapter 3 Acquisition of clefts in L3 European Portuguese by L1-Mandarin L2-English speakers

In this chapter, we will present the theoretical framework of generative L2 acquisition adopted in this study and the working hypotheses underlying the experimental task. The notion of L2 acquisition (SLA) assumed here refers, in a broad sense, to the acquisition of any non-native language subsequent to that of the native language, L1. In this regard, we approach the L2 acquisition of cleft structures with the Feature Reassembly Hypothesis (Lardiere, 2008, 2009) in 3.1.1. As EP is in fact the L3 of the overwhelming majority of the Chinese students that we tested, who, apart from L1 MC, have acquired English as a foreign language before learning EP, we will also refer to some recent studies of third or additional language acquisition (L3A) in 3.1.2. In Section 3.2., we present the working hypotheses formulated in the light of the analyses of different clefts presented in the previous chapter and the SLA theories introduced in 3.1.

3.1 Feature Reassembly Hypothesis and third/additional language acquisition

3.1.1 Feature Reassembly Hypothesis

From an innatist perspective, any particular natural language is acquired by a speaker through the interaction of linguistic input, i.e. "experience", with an innate component genetically determined to yield the acquisition, "the faculty of language" of the human mind, or the Universal Grammar (henceforth UG) (Chomsky, 1986). According to generative studies on second language acquisition since the 1980s (e.g., White, 2003), second language grammars are constrained by the UG as well. More recently, it is assumed that three factors interacting with one another, genetic endowment, experience and principles that are not specific to the faculty of language, determine all (I-)languages attained (Chomsky, 2005).

Within this perspective, the Principles-and-Parameters approach has profoundly investigated the relevance of Universal Grammar in language acquisition, including L2 acquisition. For example, according to the Full Transfer/Full Access Model (FT/FA) (Schwartz & Sprouse, 1994, 1996), at the

initial stages, all the knowledge of syntactic properties of the L1 is fully imposed on the input of L2 excluding the language-specific morpholexical items, and when the L1 grammar cannot accommodate the L2 input, learners take recourse to UG, to which they have full access, and build their L2 grammar.

However, in current SLA studies, parameter resetting theories, which lead to expect a drastic change in the grammar of an L2 learner once he/she simply resets one or more parameters accommodating the L2 input, cannot explain evidence of morphological variability, i.e. "the variable omission, underspecification, overreliance on default forms, and/or apparent optionality vs. obligatoriness of the morphophonological expression of grammatical properties" (Lardiere, 2008, p. 2). In the framework of the Minimalist Program (Chomsky, 1995), phonological, semantic and grammatical features, as the primitive units that compose the lexical items of each language, constitute the functional categories and trigger movements. Functional categories, in turn, are viewed as assembled bundles of features, or "feature-matrices". As a result of the fragmentation of parameters into features, the process of parameter-setting has been broken up into the process of selecting and assembling features into language-specific lexical items. Hence, the core question of SLA task is considered nowadays as the acquisition of features and their configuration.

Building on the FT/FA Hypothesis, Lardiere (2008, 2009) proposed the Feature Reassembly Hypothesis, which assumes that the second language acquisition essentially consists of the mapping and reconfiguration of features into L2 specific lexical items[45]. Three main challenges that learners encounter are: (1) the association of relevant features with the exact functional categories in L2 and the potential divergence from the exact configuration of feature bundles in L1; (2) the combination and expression of features in the correct lexical items in L2; (3) the optionality or obligatoriness of the realization of certain forms conditioned by phonological, morphosyntactic, semantic or discourse-linked factors in L2.

As in the process of the acquisition of L1 the learners have already assembled the selected features into language-specific items, when learners are exposed to a new language, the acquisition initiates from the fully assembled bundles of features, i.e. the whole grammatical categories of L1. Naturally,

---

[45] In the Feature Reassembly Hypothesis, the term "(morpho)lexical items" refers to "language-specific morphemes with functional features (e.g., *the*, *-s*, *who*, etc.)." (Lardiere, 2009, footnote 1)

by contrasting the differences of sound-meaning pairs in L1 and L2, learners tend to seek the morpholexical items in L2 which are the most approximate correspondence to those in L1, "presumably on the basis of semantic meaning or grammatical function (the phonetic matrices will obviously differ)" (Lardiere, 2009, p. 191). Once the features or the set of features encoded in the L1 and in the L2 are equivalent, learners do not need to repeatedly access the features inventory but map the established configurations onto the probably different morpholexical items of the target language, L2. Within a separationist model of grammar, the author distinguished the syntactic component of grammar from a morphological (or phonological) component, where the source of morphological variability is located, and assumed an indirect mapping of the output of syntactic computation "via morphological (or phonological) module-specific 'translation' procedures to actual phonological forms" (Lardiere, 2008, p. 25). When the language-specific morphosyntactic feature contrasts between L1 and L2 are detected by adult L2 acquirers - even for uninterpretable features, such contrasts seem to be detectable -, they enter the (re-)assembly stages of acquisition, and the true learning problems and great difficulties lie in how to discern and disentangle the configuration of features from the placement and properties of specific lexical items in their linguistic input, build the right representation, reassemble such features in the L2 lexical items and determine the proper environments, e.g., diverse contextual conditions for such expression. In principle, any detectable feature contrast, according to the author, is ultimately acquirable, in other words, the UG is fully accessible to the (re-)assembly of features in specific L2 lexical items. Following models such as Distributed Morphology, Lardiere (2005, 2008) maintained that the problem of morphological variability displayed in the performance of second language acquisition lies in a post-syntactic spell-out or selection of morphological or phonological feature, in which the "closest-matching *vocabulary entry*" (Lardiere, 2008) is selected and inserted into the nodes packed with already-assembled morphosyntactic feature bundles. Such processes of selecting and inserting the vocabulary entry consist of the issue of morphological competence, which is to be acquired by L2 learners separately from the mental representation of syntactic structures.

The Feature Reassembly Hypothesis would predict that the essence of the acquisition of cleft structures lies in the association of focus features to different functional categories in each language, and, apparently, to distinct lexical items. As presented in the previous section, in MC, focus features

are associated to functional categories different from the ones of EP and English. We assume that, in MC *shì…de* clefts, the [+focus] feature is associated to the head of a low FocP. To ultimately acquire the EP cleft structures, apart from the mastery of the lexical items, an L1-MC learner would have to untangle the [+focus] feature from the low Foc^0 in MC and reassembly it onto different functional categories of different cleft structures according to their EP input. In *é-que* clefts and standard clefts, the same feature should be associated to a functional category in the left periphery; in semipseudoclefts, it should be attached to the projection lexicalized by the form *é* "be"; in *wh*-clefts, basic and inverted pseudoclefts, i.e. identificational structures involving a *wh*-morpheme, a learner would be expected to exhibit apparent knowledge of the target grammar at the *mapping* stage, resulting from mapping of MC pseudoclefts.

3.1.2 Third/additional language acquisition

In the last two decades, generative acquisition studies have extended the scope of inquiry to complementary populations to adult L2 acquisition, for instance, third or additional language acquisition (L3A). In L3A, the debate on interlinguistic influence or language transfer is more complex due to the eventual impact from two or more previous linguistic systems that learners have already acquired. Studies like Na Ranong and Leung (2009) and Hermas (2014) presented evidence that support the L1 transfer scenario, claiming the privileged role of the L1 as the source of both facilitative and non-facilitative transfer in the initial stages, overriding factors such as the (psycho-)typological proximity. The L2 Status Factor (Bardel and Falk, 2007, 2012; Falk, Lindqvist & Bardel, 2015) argues that, due to the cognitive and situational similarities that a formally learned L2 shares with an also formally learned L3, the L2 has a privileged status in morphosyntactic transfer over the L1. The Cumulative-Enhancement Model (CEM, Flynn et al. 2004; Berkes and Flynn, 2012) claims that language learning is a cumulative process, and that all previously acquired linguistic systems provide facilitating effects or remain neutral in the subsequent acquisition of an L3, while negative or non-facilitative transfer is not predicted. The Typological Primacy Model (TPM, Rothman, 2010, 2011, 2013, 2015) holds that either actual structural proximity or learners' unconscious perception of structural proximity between the target L3 and the previously acquired linguistic systems is what

determines which system, taken as most similar by the internal parser, will provide transfer at initial stages. Both facilitative and non-facilitative transfer are predicted to occur. The Scalpel Model (SM, Slabakova, 2017) argues that there is no wholesale transfer at the initial stages and that the already acquired "L1-plus-L2 grammar" acts with a "scalpel-like precision" to extract the L1 and/or L2 options. According to this model, however, the transfer is not always facilitative.

As explored above, a L1-MC L2-English learner of EP as an L3 would face the task of disentangling the [+focus] feature from the low Foc^0 as in MC *shì...de* clefts, and associating the same feature to the left periphery, which may be accomplished in the process of the acquisition of English *it*-clefts and have a facilitative effect on the acquisition of Portuguese, according to some L3A models such as CEM, TPM and SM. Alternatively, as predicted by the L1 transfer scenario, the learner may be forced to start from scratch due to the inexistence of focus in the left periphery in L1 MC.

3.2 Working hypotheses

The aforementioned distinctive properties of clefts in European Portuguese and Mandarin Chinese explored in previous studies, as well as the eventual transfer from L2 tackled in the literature of L3 acquisition, lead us to speculate that L1-MC learners acquiring EP as an additional language may exhibit the acquisition of four cleft structures – *é-que* clefts, standard clefts, pseudoclefts, and semipseudoclefts – in diverse stages due to their distinct syntactic structures, based on the Feature Reassembly Hypothesis (Lardiere, 2008, 2009).

This section will state the working hypotheses, formulated taking into account our research questions (see section 1.1). The specific assumptions underlying each hypothesis will be explained.

Hypothesis 1. *É-que* clefts display a delayed development.

As an *é-que* cleft is a unique configuration of L3 compared to both L1 and L2, learners cannot find the exact mapping source from the two previously acquired language systems. The difficulties encountered by a learner when acquiring the structure consist of both the lexicalization of the *é-que* expression and the association of the [+focus] feature to such an expression in a higher domain.

a. Learners could lack sufficient knowledge about the lexicalization of the *é-que* expression, taking the copula *ser* in the expression as an inflected verb and *que* as the head of a relative clause, and may subsequently accept phi-feature agreement between the copular verb *ser* and the main verb in the cleft sentence.

b. If *é que* is acquired as a lexicalized expression, the association of the [+ focus] feature with the *é-que* expression can still be hard to acquire, as neither the L1 nor the L2 provides an equivalent expression bearing such a feature. In addition, although the *é-que* cleft is a cleft structure frequently produced by native speakers (in spontaneous and elicited production; Lobo, Santos & Soares, 2016, a.o.), empirical observation and some PFL (Portuguese as a Foreign Language) materials suggest that the expression is more probably taught to learners as a strategy for the formation of *wh*-interrogatives, rather than as a clefting strategy. In this case, learners will reject the appearance of the *é-que* expression in clefts as well as the inflected copula *ser* combining with *que*.

Hypothesis 2. Standard clefts are more precociously acquired.

Due to the relatively high frequency of standard clefts in the L1 (Lobo, Santos & Soares, 2016), learners are predicted to have sufficient exposure to the structure. The remaining difficulty would lie in the reassembly of the focus feature, disentangling it from the lower Foc^0 as in L1 grammar and realizing it in the head position of a C projection, in line with analysis of MC clefts suggested in this dissertation for MC and the analysis of EP clefts presented by Lobo (2006).

a. If the starting point of the acquisition of EP as an additional language is solely or favorably the L1 grammar, as claimed by Na Ranong and Leung (2009) and Hermas (2014), a subject-object asymmetry could be attested at the earlier stages of acquisition in the learners' grammar of standard clefts, given that in the L1 the object can never be fronted in either *shì…de* clefts or in bare-*shì* clefts. This discrepancy is caused by the divergence in the syntactic derivation between the two languages. As explored in the previous chapter, neither of these two clefting strategies in MC involves the association of the [+focus] feature to a certain C head, and thus, learners would face the difficulty of disengaging the [+focus] feature from a lower functional head as in L1 and associating it to a head in the left periphery in EP subject or object standard clefts. Additionally, in the L1, the impossibility of

clefting an object by dislocating it to a higher position, i.e. the prohibition of associating the [+focus] feature to a functional projection in object clefts, could cause even more difficulties in the acquisition of object standard clefts. Consequently, if learners have not completely acquired the target grammar but simply transfer the superficial structures of *shì...de* clefts or bare-*shì* subject clefts from the L1, they will not find object standard clefts acceptable, even though this structure is perfectly natural in EP.

b. If the transfer effect from the L2 does somehow play a part in L3 acquisition, as argued by the L2 Status Factor (Bardel and Folk, 2007, 2012), the Cumulative-Enhancement Model (Flynn et al. 2004; Berkes and Flynn, 2012) and the Typological Primacy Model (Rothman, 2011, 2013, 2015), amongst other proposals, the superficial similarity between the English *it*-cleft and the EP standard cleft could facilitate the acquisition of the latter by providing clues for the mapping to happen at the initial stages. As it is perfectly acceptable to focus the object in English *it*-clefts and *that* is generally agreed to be a complementizer, the mapping from English can help overcoming the obstacle of acquiring object standard clefts. In other words, there would be no subject-object asymmetry.

Hypothesis 3. Learners of less advanced levels may find a standard-cleft-like structure with an empty C acceptable.

a. Neither *shì...de* clefts nor bare-*shì* subject clefts in Mandarin involve the C domain in their syntactic derivation as the clefted constituent never raises to C and consequently, a complementizer is absent in both constructions. The central question lies again in whether learners can untangle the [+focus] feature from a lower Foc^0 or T^0 and associate the same feature to a projection in the C domain. If the C domain in learners' L3 clefts remains "inactive", and thus the association of the [+focus] feature to a C head is not acquired, learners will not reject a configuration like "*ser* + focus + presupposition", namely a standard-cleft-like structure with an empty C head. Such a configuration is similar to bare-*shì* subject clefts and *shì...de* clefts in L1 MC. However, MC bare-*shì* subject clefts only allow the focalization on subjects. When the association of the feature to the left periphery fails, if the L1 transfer is not blocked, and learners at the initial stages exactly map the structure of bare-*shì* subject clefts onto lexical items in Portuguese, they will consider the aforementioned configuration as

grammatical when a subject is clefted, while rejecting the "*ser* + clefted object + presupposition" configuration, since, in Mandarin, there are no *shì* + object clefts.

b. If the learners accept the "*ser* + clefted object + presupposition" configuration but reject its counterpart, namely, a focalized subject in this string without a complementizer, it will be implicated, as predicted by TPM or SM, that the possibility of having a null complementizer in English object *it*-clefts is carried into learners' L3 grammar as a negative transfer, and that L2 English is taken as the only source for mapping in terms of this structure.

Hypothesis 4. Semipseudoclefts are a difficult type of cleft structure to acquire.

As an idiosyncratic clefting strategy of Portuguese with unique syntactic constraints (Costa & Duarte, 2001; Lobo, 2006, i.a.), there is no source for mapping from either L1 or L2 for the acquisition of semipseudoclefts. In simple descriptive terms, in simple sentences in Mandarin Chinese, the copular verb *shì* (be) cannot immediately follow another verb as it does in such a structure in EP; in English, this configuration also leads to ungrammaticality. In more precise terms, in MC, contrary to EP, a form of BE does not lexicalize the [+focus] feature. The reassembly of the focus feature to a lexicalized form of the verb *ser*, which in these circumstances is located at the edge of the vP in EP as a contrastive focus marker, is then the only problem, yet a brand-new problem, for learners.

Hypothesis 5. Learners have better performance on pseudoclefts, whether the clefted constituent is a subject or an object.

In both MC and English, the focalization of either a subject or an object in pseudocleft structures is felicitous. In line with the analyses for each language that we adopt in this study, the pseudoclefts in all three languages involve a small clause containing an embedded *wh*-clause. Therefore, the acquisition of EP pseudoclefts, should find sufficient sources for mapping from either of the previously acquired linguistic systems.

Hypothesis 6. Even if learners can accept standard clefts and pseudoclefts in the earlier stages, they may not be able to fully acquire the target grammar, more specifically, the agreement patterns of the

two constructions.

In subject standard clefts, the verb in the cleft clause always agrees with the clefted subject, which is explained by Lobo's (2006) analysis, assuming that the clefted constituent is extracted from the inner VP position of the embedded clause. When it comes to subject pseudoclefts, as the verb in the cleft clause, namely the *wh*-clause, is headed by the *wh*-word, it is ungrammatical for it to agree with the clefted subject, with which the *wh*-clause forms a small clause. To avoid agreement mismatches, learners should first obtain the mental representation of the relevant syntactic structures.

a. If learners consistently produce agreement between the clefted subject and the embedded verb irrespective of the type of cleft, they may be processing both standard clefts and pseudoclefts as a structure involving extraction. If, conversely, such agreement never happens, learners may be assimilating the complementizer *que* in standard clefts to *wh*-words in pseudoclefts and taking both types of cleft as identificational structures.

b. Alternatively, learners may systematically produce agreement between the embedded verb and the clefted subject, as a result of semantically motivated agreement, if they comprehend the anaphoric relation between the clefted constituent and the wh-word, which in turn c-commands the embedded verb.

Chapter 4 Methodology

To evaluate the acquisition of cleft structures in EP, an acceptability judgment task was designed and applied. The task comprised two subtests, namely an acceptability test of different cleft structures (*é-que* cleft, standard cleft, pseudocleft, semipseudocleft, and a cleft without realization of the element(s) in C^0) and a test of person agreement patterns between a clefted subject and the embedded verb in such structures.

Section 4.1 describes the experimental task and its application, whereas section 4.2 presents the type of treatment and statistical analysis of the data obtained in the experimental task. In section 4.3, we describe a complementary corpus search centered on cleft structures produced by L1-MC learners.

4.1 The Experimental Task

Participants

Sixty Chinese students aged 20 to 31 years old whose L1 is MC and who were attending L2 Portuguese courses provided by the Institute of Portuguese Language and Culture (ICLP) or were enrolled in a master program at the Faculty of Arts at the University of Lisbon (FLUL) were tested as participants in this study.[46] The participants were divided into three groups according to their Portuguese proficiency level. There were 23 learners of level B1 (20-24 years old, mean age 21.5), 20 of level B2 (20-25 years old, mean age 23.4), and 17 of level C1 (21-31 years old, mean age 24.2)[47]. The self-reported English levels of all Chinese participants range from B1 or CET4 (see section 1.3) to C2. Twenty-one monolingual European Portuguese speakers from continental Portugal, aged 20 to 69 years old (mean age 27.7), also took part in the task as a control group. None of the participants

---

[46] In total, 127 participants, who were students at ICLP, were tested; 67 of these were excluded for not being Mandarin Chinese speakers or for being potentially influenced by dialects structurally distinct from Mandarin with respect to cleft structures.

[47] Participants' level of Portuguese corresponded to the level of CAPLE exams (Centre for Evaluation of Portuguese as a Foreign Language) they had achieved. In the case that the participants had not participated in the exams, they were grouped according to the highest level of an annual course that they had finished and in which they had passed the corresponding evaluation exams.

from either group had a relevant background in linguistics[48].

Data collection and design of the experiment

With the written and informed consent of each participant, the data were mainly collected at FLUL as a classroom task. Both tests were presented in printed form and were to be completed in writing. In the task, the participants were asked to read each given context and evaluate the corresponding sentence according to how it sounded grammatically; if they found a sentence to be ungrammatical, they were asked to correct it to a grammatical one. They evaluated the acceptability of the test items on a four-point Likert scale ranging from 0 (totally ungrammatical) to 3 (totally grammatical). An option of *não sei* "I do not know" was also included. They were explicitly told to ignore any punctuation and orthographic problems, which did not exist in the test, but which the participants might think were at stake.

The task included items from two subtests, with a total of 48 test items, and 66 fillers. The first subtest was designed to evaluate the participants' acceptance of different types of subject and object cleft structures. The second subtest was designed to evaluate participants' acceptance of agreement patterns in subject clefts.

In all the test items, the verbs were designed to be uniformly in the past tense, since tense agreement is beyond the interest of the present thesis; the verbs were transitive, except for the items presenting subject semipseudoclefts, in which the verbs were unaccusative to ensure the legitimacy of the structure. To avoid intervention effects in sentence processing, the subjects and objects were always kept prototypical, i.e. [+ human] subjects and [- animated] objects. For the second test, which evaluated the agreement patterns accepted by learners, all the clefted constituents were set as plural subjects. In order for the task to be accessible for participants with a less advanced level, namely B1 learners, we used frequent and accessible lexical items.

Since all participants were also L2 English speakers, to test the potential influence of English on the acquisition of Portuguese clefts, especially on standard clefts and pseudoclefts, a similar test of English clefts was also conducted with the same requirements. It was applied after the Portuguese test,

---

[48] Some of the Chinese and Portuguese participants were enrolled in linguistics courses. However, their response patterns in the tasks did not differ from those of the others.

so as to avoid a possible priming effect.

All the items were always preceded by a sentence that provided an adequate discourse context, assuring that all cleft sentences were associated to a contrastive focus reading, since this reading is appropriate with all EP cleft structures. The presentation order of the test items and of the distractors in all tests and in both languages was semi-randomized. For each language, subtest 1 and subtest 2 were mixed (also mixed with the fillers) and presented as one whole task. All the items of the first and second tests were first randomized in Excel and then manipulated to avoid the presentation of items from the same condition in a sequence, an uneven distribution of the test items and distractors, and to exclude any test item from the first position of the task. In what follows, each subtest is presented with an example for each condition.

Test 1

Test 1 manipulated two variables in a 2 x 5 design: i) type of cleft (with 5 levels: é-que cleft, standard cleft, semipseudocleft, pseudocleft, and a structure which would correspond to a standard cleft without a lexicalized C); ii) syntactic function (with 2 levels: subject; object).

The test conditions are presented with details in Table II.

Condition	Type of Cleft	Clefted Constituent	Number of Items
1	é-que cleft	Subject	3
2	standard cleft	Subject	3
3	*standard-cleft-like structure with empty C	Subject	3
4	semipseudocleft	Subject	3
5	Pseudocleft	Subject	3
6	é-que cleft	Object	3
7	standard cleft	Object	3
8	*standard-cleft-like structure with empty C	Object	3
9	semipseudocleft	Object	3
10	Pseudocleft	Object	3

Conditions 1 and 6 enabled us to ascertain the participants' judgement regarding *é-que* clefts when clefting a subject or an object, respectively.

(1) Test item 1, condition 1 (subject *é-que* cleft):

CONTEXTO: *A Vera não sujou a roupa.*

CONTEXT: Vera did not dirty the clothes.

Test item: O Vítor   é que   sujou   a   roupa.

**   the Vítor is that dirtied the clothes**

Vítor **dirtied the clothes.**

(2) Test item 16, condition 6 (object *é-que* cleft):

<small>CONTEXTO:</small>*A Antónia não visitou o museu.*

<small>CONTEXT:</small> Antónia did not visit the museum.

Test item: O castelo é que   a   Antónia visitou.

       the castle   is that the Antónia visited

       Antónia visited *the castle*.

Conditions 2 and 7 tested the participants' intuition about subject and object standard clefts.

(3) Test item 6, condition 2 (subject standard cleft):

<small>CONTEXTO:</small>*A Rafaela não cancelou a reunião.*

<small>CONTEXT:</small> Rafaela did not cancel the meeting.

Test item: Foi   o   Santiago que cancelou   a   reunião.

       was the Santiago that canceled   the meeting

       It was *Santiago* that canceled the meeting.

(4) Test item 19, condition 7 (object standard cleft):

<small>CONTEXTO:</small>*A Sónia não estudou um romance realista.*

<small>CONTEXT:</small> Sónia did not study a realist novel.

Test item: Foi um poema modernista que   a   Sónia estudou.

       was   a   poem   modernist that the Sónia studied

       It was *a modernist poem* that Sónia studied.

Conditions 3 and 8 enabled us to assess whether the participants accepted the standard-cleft-like structure without an empty C head, which is ungrammatical in European Portuguese.

(5) Test item 7, condition 3 (*subject standard-cleft-like structure with empty C)

<small>CONTEXTO:</small>*O Gilberto não partiu a janela.*

<small>CONTEXT:</small> Gilberto did not break the window.

Test item: Foi   a Margarida partiu   a   janela.

      was the Margarida broke the window

      It was *Margarida* that broke the window.

(6) Test item 22, condition 8 (*object standard-cleft-like structure with empty C)

CONTEXTO: *A Joana não tomou chá.*

CONTEXT: Joana did not have tea.

Test item: Foi   café   a Joana tomou.

      Was coffee the Joana had

      It was *coffee* Joana had.

Conditions 4 and 9 assessed the participants' acceptance of subject and object semipseudoclefts.

(7) Test item 10, condition 4 (subject semipseudocleft):

CONTEXTO: *A Liliana não chegou antes da aula.*

CONTEXT: Liliana did not arrive before class.

Test item: Chegou   foi   a Teresa.

      Arrived was   the Teresa

      Teresa **arrived.**

(8) Test item 25, condition 9 (object semipseudocleft)

CONTEXTO: *A Lorena não comprou um quadro.*

CONTEXT: Lorena did not buy a painting.

Test item: A Lorena comprou foi um CD.

      the Lorena bought was a   CD

      Lorena bought *a CD*.

Conditions 5 and 10 allowed us to test whether the participants accepted subject and object pseudoclefts.

(9) Test item 15, condition 5 (subject pseudocleft):

CONTEXTO: *A Fabiana não aprendeu francês.*

CONTEXT: Fabiana did not learn French.

Test item: Quem aprendeu francês foi   o   Alberto.

           who   learnt   French was the Alberto

           Who learnt French was *Alberto*.

(10) Test item 29, condition 10 (object pseudocleft):

CONTEXTO: *A Frederica não vendeu os óculos.*

CONTEXT: Frederica did not sell the glasses.

Test item: O que   a Frederica vendeu foi   o casaco.

           the what the Frederica sold   was   the coat.

           What Frederica sold was *the coat*.

<u>Test 2</u>

The second test manipulated two variables, in a 2 x 3 design: i) type of cleft (with 3 levels: *é-que* cleft, standard cleft and pseudocleft); ii) agreement pattern (with 2 levels: with and without agreement) – 6 conditions. In this case, and since subject – verb agreement was at stake, only subject clefts were tested, and the clefted subjects were always plural.

In Table III and the presentation of examples of the test items, Condition 2, *é-que* cleft with verbal inflection of *ser* "be", will be designated in short as "*é-que* cleft with inflection", to be differentiated from Condition 1, *é-que* cleft. Condition 3 will be labeled as "standard cleft", and Condition 4, standard cleft without agreement in the embedded verb, will be labeled as "standard cleft without agreement". Condition 5, pseudocleft without agreement in the verb of the relative clause, which corresponds to the standard EP, will be designated as "pseudocleft without agreement". Condition 6, pseudocleft with agreement in the verb of the relative clause, will be shorten as "pseudocleft with agreement".

The test conditions are presented in detail in Table III.

Condition	Type of cleft	Agreement	Number of Items
1.é-que cleft	é-que cleft	-	3
2. *é-que cleft with inflection	é-que cleft	+	3
3. standard cleft	standard cleft	+	3
4. *standard cleft without agreement	standard cleft	-	3
5. Pseudocleft without agreement	Pseudocleft	-	3
6. *Pseudocleft with agreement	Pseudocleft	+	3

As *é* and *que* constitute a frozen expression not exhibiting any type of agreement, conditions 1 and 2 evaluated the mastery of this property of *é-que* expression. Condition 2 tested the *é-que* cleft of a plural subject with inflection on *ser* "be".

(11) Test item 31, condition 1 (*é-que* cleft)

CONTEXTO: *A Anita não preparou sobremesas.*

CONTEXT: Anita did not prepare dessert.

Test item: Os irmãos   é que prepararam sobremesas.

            the brothers is that prepared-3PL dessert

***The brothers* prepared dessert.**

(12) Test item 34, condition 2 (*é-que cleft with inflection)

CONTEXTO: *O rapaz não corrigiu o texto.*

CONTEXT: The boy did not correct the text.

Test item: As raparigas foram que   corrigiram   o texto.

        the   girls   were   that corrected-3PL the text

        ***The girls* corrected the text.**

Conditions 3 and 4 separately tested the participants' acceptance of a standard cleft without or with agreement between the clefted subject and the verb form in the clefted clause.

(13) Test item 38, condition 3 (standard cleft)

CONTEXTO: *O Diogo não fez o teste.*

CONTEXT: Diogo did not do the test.

Test item: Foram os amigos   que fizeram o teste.

        Were the friends that did-3PL the test.

        It was *the friends* that did the test.

(14) Test item 41, condition 4 (*standard cleft without agreement)

CONTEXTO: *O João não recebeu o prémio.*

CONTEXT: João did not receive the prize.

Test item: Foram as meninas que   recebeu   o   prémio.

        were   the   girls   that received-3SG the prize.

        It was *the girls* that received the prize.

Conditions 5 and 6 respectively assessed whether the participants accepted the subject pseudocleft without or with standard agreement between the clefted plural subject and the verb form in the relative clause. This is because in standard European Portuguese, the embedded verb agrees with the *wh*-word and not with the clefted subject and hence it always occurs in a singular form.

(15) Test item 44, condition 5 (pseudocleft without agreement)

CONTEXTO: *O dono não fechou a loja.*

CONTEXT: The boss did not close the shop.

Test item: Quem   fechou   a   loja foram os filhos.

who   closed-3SG the shop were the sons

Who closed the shop were *the sons*.

(16) Test item 47, condition 6 (*pseudocleft with agreement)

CONTEXTO: *A Rosa não elaborou esta teoria.*

CONTEXT: Rose did not elaborate this theory.

Test item: Quem   elaboraram esta teoria foram as amigas.

who elaborated-3PL this theory were the friends

Who elaborated this theory were *the friends*.

Complementary English Test

To examine the possible influence of L2 English, two tests focusing on the acceptability and the agreement patterns of English clefts were also applied to the same groups of participants.

The English test also contained two subtests, totaling 16 test items as well as 24 distractors. Subtest 1 examines the acceptability of English clefts and is focused on three types of cleft structures: the *it*-cleft, the pseudocleft, and a structure that is superficially identical to the *it*-cleft but without any element in C; there were two items for each condition defined by cleft types, one in which the subject was clefted and the other in which the object was clefted. In subtest 2, the agreement pattern test, only the agreement of the subject and the matrix verb "be" was manipulated in *it*-clefts, since English does not mark person agreement with the subject in the simple past tense with most verbs.

The test conditions of the English subtest 1 and 2 are presented in Table IV and V.

Condition	Type of Cleft	Clefted Constituent	Number of Items
1	*it*-cleft	Subject	2
2	**it*-cleft-like structure with empty C	Subject	2
3	Pseudocleft	Subject	2
4	*it*-cleft	Object	2
5	*it*-cleft-like structure with empty C	Object	2
6	Pseudocleft	Object	2

Table V: Test conditions of the English subtest 2

Condition	Type of cleft	Agreement	Number of Items
1 plural subject *it*-cleft	plural subject *it*-cleft	-	2
2 * plural subject *it*-cleft with agreement	plural subject *it*-cleft	+	2

Instances of the two English tests are given below:

(17) Test item 1, condition 1 of Test 1 (subject *it*-cleft)

CONTEXT: *Levin didn't drop his key.*

Test item: It was *Peter* that dropped his key.

(18) Test item 42, condition 2 of Test 1 (*subject *it*-cleft with empty C)

CONTEXT: *Teddy didn't break the window.*

Test item: It was *Fredric* broke the window.

(19) Test item 20, condition 3 of Test 1 (subject pseudocleft)

CONTEXT: *David didn't drink the milk.*

Test item: Who drank the milk was *Paul.*

(20) Test item 26, condition 4 of Test 1 (object *it*-cleft)

CONTEXT: *Cristine didn't write a post card.*

Test item: It was *a letter* that Cristine wrote.

(21) Test item 9, condition 5 of Test 1 (object *it*-cleft with empty C)

CONTEXT: *John didn't borrow a dictionary.*

Test item: It was *a book* John borrowed.

(22) Test item 36, condition 6 of Test 1 (object pseudocleft)

CONTEXT: *Adam didn't pick an apple.*

Test item: What Adam picked was *a pear*.

(23) Test item 34, condition 1 of Test 2 (plural subject *it*-cleft without agreement)

CONTEXT: *The singers didn't practice the musical.*

Test item: It was *the dancers* that practiced the musical.

(24) Test item 16, condition 2 of Test 2 (*plural subject *it*-cleft with agreement)

CONTEXT: *The police didn't search the whole town.*

Test item: It were *the volunteers* that searched the whole town.

4.2 Statistical Treatment of the Data

Upon completion of the task, all answers and corrections made by the participants were transcribed and saved in digital form in a Microsoft Excel file.

In the quantitative analysis, all the classifications made by the participants in the Likert Scale were taken into account, and entered the statistical analysis, with the exception of responses with *não sei* "I do not know", which were coded as n.a. (no answer). In the qualitative analysis, where we concentrated on the modifications made by the participants, a lack of correction for responses with score inferior to 3 (totally grammatical) were coded as n.a. (no answer) in the Excel file.

The corrections of the test items made by the participants were transcribed, and those made regarding the cleft structures were considered as relevant corrections and served as the sources for further observation. Other kinds of corrections, e.g., the substitution of a complete noun phrase for a pronoun, were noted but not taken into consideration for this work.

The numeric responses were later introduced into SPSS for an inferential statistical analysis. We adopted an independent sample non-parametric Kruskal-Wallis test for each experimental condition – this test allowed us to maintain the treatment of the answer variable as ordinal and permitted comparisons for more than 2 independent samples. When differences between groups were found, a post-hoc Dunn's test was also performed for multiple pairwise comparisons, using a Bonferroni correction to adjust the *p* value.

4.3 Complementary corpus search

To complement the experimental task, we also conducted a corpus search centered on cleft structures produced by L1-MC learners in COPLE2 - Corpus de Português Língua Estrangeira/Língua Segunda (Mendes et al., 2016). The cleft structures encountered in the corpus were counted and saved in TXT files.

COPLE2 is a collection of written and oral texts produced by learners of Portuguese as a second/foreign language in the Institute of Portuguese Culture and Language (ICLP) at the School of Arts and Humanities of the University of Lisbon (FLUL) and by candidates for examinations of the

Center for Evaluation and Portuguese Certification Foreign Language – CAPLE– FLUL.[49] Amongst learners with 15 different native languages, Chinese native speakers, regardless of their dialect, contributed 323 files, in which only 15 sentences involving a cleft structure were found, whereby one was a citation and was hence not considered. Table VI summarizes the spontaneous production data and presents the number of cleft structures found in the corpus, considering type of cleft structure, syntactic function of clefted constituent and proficiency group. We will make reference to the results of the corpus search, when relevant, in Chaper 6. Discussion.

---

[49] COPLE2. Retrieved from http://beta.clul.ul.pt/teitok/learnercorpus/pt/index.php in 11/3/2019.

Table VI: Spontaneous production data of cleft structures in COPLE2

Cleft type \\ Level Production	A	B1				B2				C1				Total
		sub	obj	adj[50]	pred	sub	obj	adj	Pred	sub	obj	adj	pred	
standard cleft	-	-	-	2	-	3	-	1	-	-	-	-	-	6
é-que cleft		-	-	-	-	-	-	-	-	-	-	-	-	0
pseudocleft		-	1	-	-	2	-	-	1	1	-	-	2	7
semipseudocleft		-	-	-	-	-	-	-	-	-	-	-	-	0
cleft without C element		-	-	-	-	-	-	1	-	-	-	-	-	1
"*foi+ é que*" cleft		-	-	-	-	-	-	1	-	-	-	-	-	1

---

[50] Adjunct cleft.

Chapter 5 Description of the experimental results

In this chapter, we present the results of the experimental task.

The results are introduced by types of structures. For each structure, a table and a figure are presented, showing the percentages of responses obtained and corresponding to each Likert scale score, namely, 0, 1, 2 and 3 (from totally ungrammatical to totally grammatical). Thereafter, the results of statistical treatment are introduced, followed by the description of relevant corrections made by the participants. Such corrections can be consulted in Appendix IV.

5.1 Results of Test 1

Test 1 examined the acceptability of diverse types of Portuguese cleft structures, namely the *é-que* cleft, standard cleft, semipseudocleft, pseudocleft, and an ungrammatical standard-cleft-like structure with an empty C, when clefting a subject and an object.

É-que Cleft: Condition 1 and Condition 6

The results of the acceptability test of subject and object *é-que* clefts are presented in Table VII and Figure I.

Group	Clefted Constituent	0	1	2	3
Control	Subject	0.0%	0.0%	7.9%	92.1%
	Object	38.1%	22.2%	22.2%	17.5%
B1	Subject	36.4%	13.6%	30.3%	19.7%
	Object	27.0%	15.9%	19.0%	38.1%
B2	Subject	31.7%	11.7%	16.7%	40.0%
	Object	32.8%	19.0%	12.1%	36.2%
C1	Subject	3.9%	7.8%	17.6%	70.6%
	Object	15.7%	13.7%	33.3%	37.3%

Figure I: Acceptability rates for é-que clefts

98
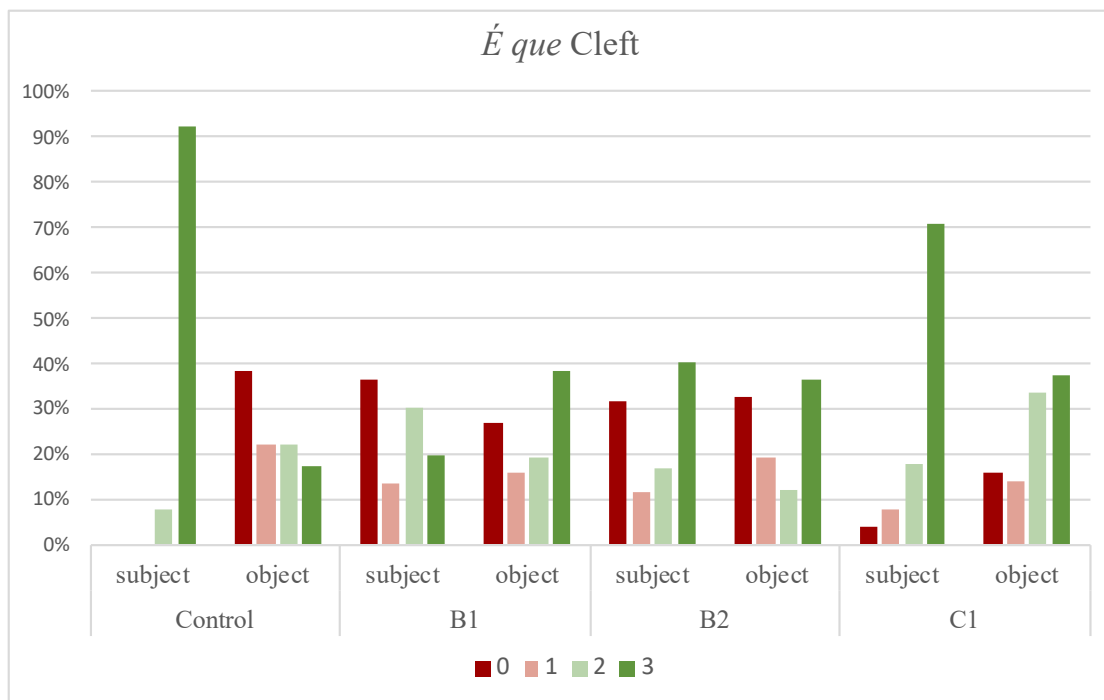
The subject *é-que* cleft appears to be well accepted at an advanced acquisition stage, C1, whereas both the B1 and B2 learners showed low acceptance rates of the structure in contrast with the control group. Conversely, the object *é-que* cleft was rejected not only by the control group but also by the B1 and B2 groups; however, the responses of C1 learners did not show a similarly strong rejection. Given the fact that the control group rejected object clefts at a relevant rate, only the subject cleft may be used to establish the acquisition of the *é-que* cleft structure by L2 learners.

Despite the high acceptance of the subject *é-que* cleft by the control group (92.1%), only 19.7% of the B1 learners and 40% of the B2 learners considered sentences with the subject *é-que* cleft to be perfect. Regarding the C1 group, 3.9% considered these sentences ungrammatical, and the complete acceptance rate was 70.6%.   In terms of the object *é-que* cleft, the control group showed the strongest rejection amongst all groups, with merely 17.5% considering the sentences completely grammatical. Groups B1 and B2 also manifested a strong rejection, with 27.0% and 32.8%, respectively, choosing to score the items as completely ungrammatical. In contrast, although the C1 group had a moderate complete acceptance rate (37.3%), it had less difficulties accepting the object *é-que* cleft sentences compared to the other groups, with only 15.7% marking the response "0" and 13.7% marking "1".

The result of the Kruskal-Wallis test shows that there is a statistically significant difference between groups in the subject *é-que* cleft condition ($H(3) = 82.078$, $p < .001$). In terms of pairwise comparisons, the responses of the B1 group are significantly different from those of the control group ($p < .001$) and from those of the C1 group ($p < .001$). Significant differences also exist between the B2 group and the control group ($p < .001$) and between the B2 group and the C1 group ($p = .001$). Meanwhile, there is no statistically significant difference between groups B1 and B2 ($p = 0.340$) or between the control group and the C1 group ($p = 0.268$). A leap in the acquisition of the subject *é-que* cleft is therefore attested between levels B2 and C1. In the object *é-que* cleft condition, the general outcome of the Kruskal-Wallis test ($H(3) = 10.880$, $p = .012$) shows a statistically significant difference between groups. The pairwise comparisons tests prove that there is a statistical significance between the control group and the C1 group only ($p = .010$).

As implied by the high acceptance rate, only a few responses from the native speakers in the control group changed the subject *é-que* cleft to a simple sentence or an inverted pseudocleft, on a par with two incidences of the substitution of the repeated noun phrase with a null object. In all the relevant

99

corrections made by the B1 learners, 52% (22 out of 42) used a simple sentence with an SVO word order to substitute the original subject *é-que* cleft by deleting the *é-que* expression. Furthermore, 19% (8 out of 42) turned it into a standard cleft, while another 19% chose an inverted pseudocleft. Starting from the B2 level, learners seemed to prefer a standard cleft (46%, 16 out of 35) to a simple sentence (28.5%, 10 out of 35) while correcting. Two participants, one from each group at the B level, substituted *é* in the *é-que* expression with *foi* (simple past tense of the verb *ser* "be", third-person singular). While few corrections were made at the C1 level, 9 out of 12, namely 75%, resulted in standard clefts, and in other three cases, the *é-que* clefts were converted into simple sentences.

Contrary to the subject *é-que* cleft, the object *é-que* cleft was not welcomed by the control group, and in half of the relevant corrections made by the native speakers (24 out of 48), a simple sentence with SVO order replaced the structure, followed by the corrections resulting in a pseudocleft (22.9%, 11 out of 48). The B1 learners also preferred correcting such a structure into a simple sentence (37.5%, 12 out of 32) or a standard cleft (28%, 9 out of 32), and five cases were found in which the structure was replaced by a standard-cleft-like structure without any element in C (15.6%). The B2 learners corrected an object *é-que* cleft into an object standard cleft (55%, 21 of 38), followed by an inverted pseudocleft (16%) and a basic pseudocleft (10.5%). There were also three corrections into a standard-cleft-like structure with an empty C. The substitution of *é* in the *é-que* expression with *foi* (simple past tense of the verb *ser* "be", third-person singular) also occurred in the responses of both the B1 and B2 groups, for one participant per group. The C1 group mostly corrected the structure into an object standard cleft (69%, 18 out of 26). Five sentences of a mixed structure "*Foi* + clefted constituent + *é que...*", e.g., *Foi o castelo é que a Antónia visitou* "It was the castle that Antónia visited", also appeared in the relevant corrections of the C1 group (19%). Such a mixed structure also occurred, albeit with fewer instances (3 in B1, 1 in B2) in the responses of both groups at the B level.

Standard Cleft: Condition 2 and Condition 7

The acceptance rates of subject and object standard clefts are shown in Table VIII and Figure II.

Table VIII: Acceptability rates for standard clefts

Group	Clefted Constituent	0	1	2	3
Control	Subject	4.8%	4.8%	20.6%	69.8%
	Object	19.0%	4.8%	23.8%	52.4%
B1	Subject	2.9%	0.0%	11.8%	85.3%
	Object	1.5%	0.0%	7.4%	91.2%
B2	Subject	6.7%	0.0%	1.7%	91.7%
	Object	0.0%	0.0%	1.7%	98.3%
C1	Subject	0.0%	0.0%	0.0%	100%
	Object	0.0%	0.0%	0.0%	100%

Figure II: Acceptability rates for standard clefts
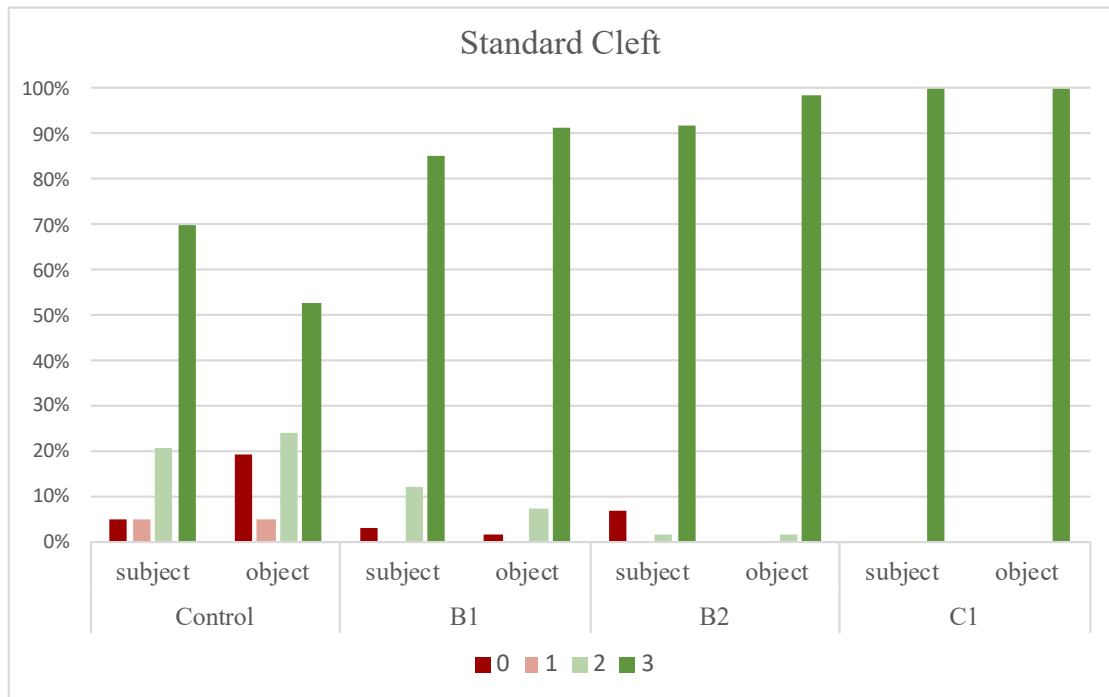
As shown in Table VIII and Figure II, the acceptance rates, i.e. the rates of full mark 3, by the

control group (subject standard cleft: 69.8%; object standard cleft: 52.4%) were unexpectedly lower than those of all three learners' groups. However, both subject and object standard clefts got high acceptance rates from the learners at level B1, and a progressive increase of the acceptance rate was seen from B1 to C1. While 85.3% of the responses of the B1 learners fully accepted the subject standard cleft, their complete acceptance rate of the object standard cleft was 91.2%. As for the B2 group, the acceptance rate of the subject standard cleft was 91.7%, and that of the object standard cleft, 98.3%. All C1 learners considered sentences with this structure completely grammatical irrespective of the clefted constituent (100.0%).

The results of the Kruskal-Wallis test on the items of condition 2, subject standard cleft, show a statistically significant difference between groups (H(3) = 22.505, p < .001). The multiple pairwise comparisons of groups B1, B2 and C1 display no statistical significance, whereas the B2 group and the C1 group both differ from the control group (p = .005, p < .001, respectively). The difference between the B1 group and the control group is marginally significant (p = .068). The general outcome of the Kruskal-Wallis test on the items of condition 7, object standard cleft, shows the existence of a statistical significant difference (H(3) = 71.900, p < .001). All the pairwise comparisons between the control group and B1, B2 or C1 show statistically significant differences (for all groups - control: p < .001), indicating that the learners, in general, behaved more distinctly from the native speakers in accepting the object standard cleft compared to the subject standard cleft.

All 12 relevant corrections of the subject standard clefts made by the control group uniformly converted the sentences into *wh*-clefts, while their modifications of object standard clefts were more diverse. In 20 corrections of the object standard clefts, besides *wh*-clefts (45%, 9 occurrences), the native speakers also turned the structure into a pseudocleft (30%, 6 occurrences), a semipseudocleft (10%, 2 occurrences) and a simple sentence (10%, 2 occurrences). The learners' groups showed a different correction pattern. The B1 learners also used a *wh*-cleft (57.1%, 4 out of 7) to alter the subject standard cleft structure. The standard-cleft-like structure with an empty C also occurred in the responses of the B1 group when correcting both the subject standard cleft (42.9%, 3 out of 7) and the object standard cleft (100.0%, 2 out of 2). In the only 3 corrections of a subject standard cleft made by the B2 learners, the structure was uniformly transformed into a *wh*-cleft. No correction was made to the object standard cleft. The C1 learners also made no alterations to either the subject or the object

standard cleft.

Standard-cleft-like Structure with Empty C: Condition 3 and Condition 8

Table IX and Figure III introduce the acceptance rates of a standard-cleft-like structure with an empty C when the subject or object are clefted.

Group	Clefted Constituent	0	1	2	3
Control	Subject	33.3%	28.6%	34.9%	3.2%
	Object	42.9%	28.6%	28.6%	0.0%
B1	Subject	51.5%	11.8%	8.8%	27.9%
	Object	37.9%	18.2%	19.7%	24.2%
B2	Subject	53.3%	8.3%	18.3%	20.0%
	Object	56.7%	18.3%	13.3%	11.7%
C1	Subject	49.0%	21.6%	23.5%	5.9%
	Object	49.0%	17.6%	23.5%	9.8%

Figure III: Acceptability rates for standard-cleft-like structures with empty C
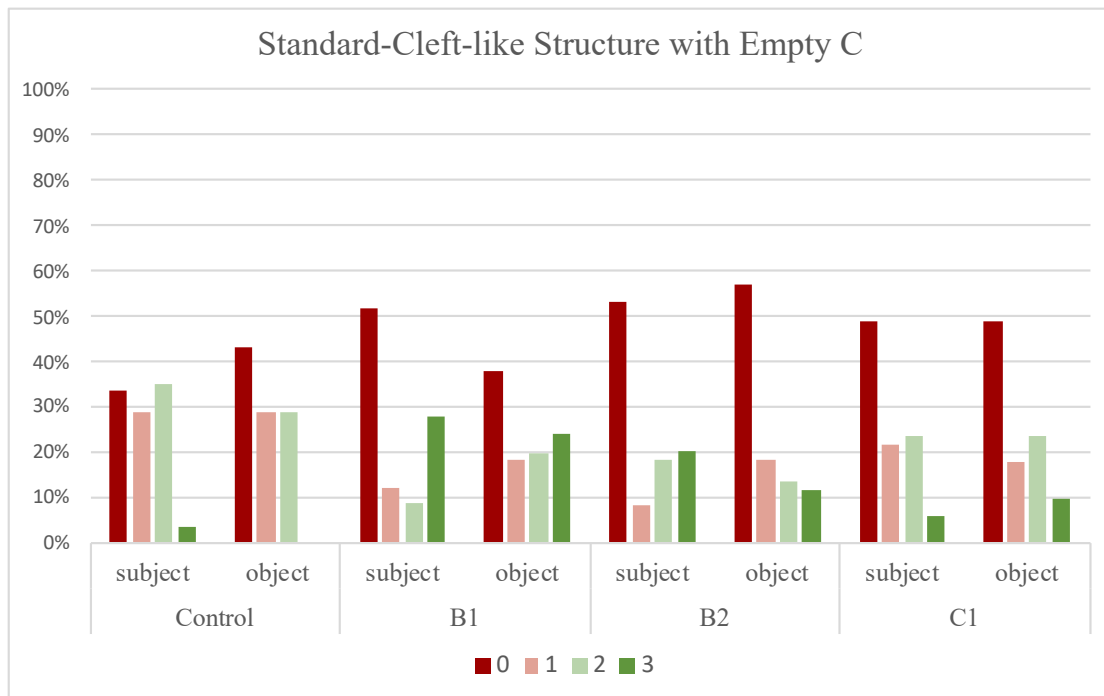
The control group showed the lowest full acceptance rate amongst all groups (3.2% for the subject

"cleft" and 0% of the object "cleft"). Although the complete acceptance rates of all the learner groups were very low, a slight and progressive decline was observed from B1 (27.9%, 24.2%) to B2 (20.0%, 11.7%) and then to C1 (5.9%, 9.8%). B1 learners fully rejected the subject standard-cleft-like structure with empty C more strongly than the object counterpart (51.5% vs. 37.9%), while in the responses of all other groups there was no such phenomenon.

All groups are almost at a ceiling level, particularly with respect to the subject standard-cleft-like structure with an empty C, as no statistical significance is shown in the outcome of the Kruskal-Wallis test ($H(3) = 1.612$, $p = .657$). As for the object standard-cleft-like structure with an empty C, the result of the Kruskal-Wallis test shows that differences between the groups are marginally significant ($H(3) = 7.142$, $p = .068$).

For the subject standard-cleft-like structure, the native speakers mainly corrected the ungrammatical structure into a standard cleft (63.3%, 38 out of 60) or a *wh*-cleft (35%, 21 out of 60). The 56 modifications to an object standard-cleft-like structure by the natives were more diverse, converting the sentence into a standard cleft (51.7%, 29 occurrences), a *wh*-cleft (19.6%, 11 occurrences), a simple sentence (12.5%, 7 occurrences) or a semipseudocleft (5.3%, 3 occurrences). Amongst the 45 relevant corrections made by the B1 learners on this subject "cleft" structure, 71.1% (32 occurrences) replaced it with a standard cleft and 22.2% (10 occurrences) with a *wh*-cleft. As for this object "cleft" structure, 80.0% (36 out of 45) of the corrections turned the sentence into a standard cleft, and 6.7% (3 out of 45) turned it into a *wh*-cleft. Four occurrences (8.9%) were found in which the learners did not change the ungrammatical structure but rather changed the word order or inserted a comma immediately after the clefted constituent, which would be filled by *que* "that" or *o que* "what" if corrected into a standard cleft or *wh*-cleft, respectively. In the responses of the B2 learners, 91.6% (44 out of 48) of the corrections of the subject standard-cleft-like structure produced a standard cleft, while the rest were *wh*-clefts. For the object standard-cleft-like structure, the substitution with a standard cleft also prevailed (98.1%, 52 out of 53), and there was only one correction to a pseudocleft. The C1 learners also manifested a strong preference to adapt the structure into a standard cleft, at 93.3% when clefting the subject and 93.1% when clefting the object. Two and three incidences of the structure *Foi…é que* are also observed in the corrections of the subject cleft-like structure and object cleft-like structure, respectively.

Semipseudocleft: Condition 4 and Condition 9

The acceptance rates of subject and object semipseudoclefts are presented in Table X and Figure IV.

Table X: Acceptability rates for semipseudoclefts

Group	Clefted Constituent	0	1	2	3
Control	Subject	12.7%	23.8%	38.1%	25.4%
	Object	6.3%	19.0%	34.9%	39.7%
B1	Subject	64.7%	11.8%	16.2%	7.4%
	Object	69.6%	13.0%	13.0%	4.3%
B2	Subject	66.7%	18.3%	15.0%	0.0%
	Object	63.3%	21.7%	11.7%	3.3%
C1	Subject	52.1%	20.8%	25.0%	2.1%
	Object	68.6%	19.6%	11.8%	0.0%
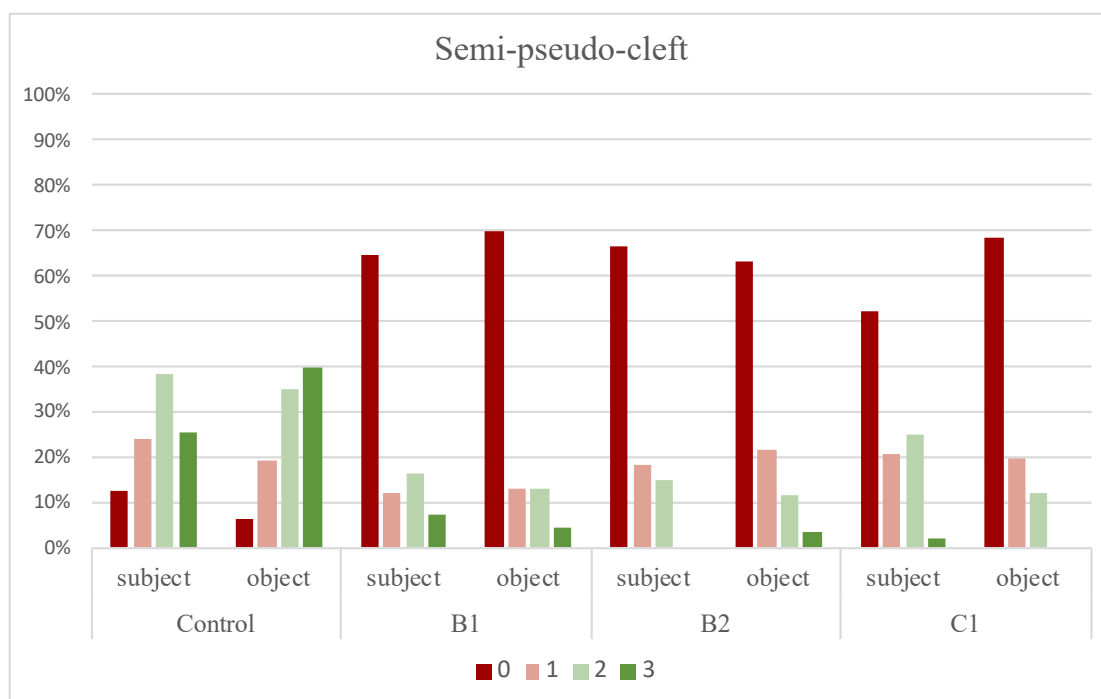
The semipseudocleft does not seem to have been acquired even when the learners have reached level C1, as sentences involving this structure were uniformly rejected by the learners at all levels regardless of which constituent was clefted. Meanwhile, the native speakers' acceptance of the semipseudocleft, though different from that of the learners, was not substantially high.

The control group partially accepted the subject semipseudocleft (25.4% marked 3), which always appeared with an unaccusative verb in all the items, while they accepted the object semipseudocleft slightly more often (39.7% marked 3). All learners showed a very low acceptance of both the subject and object semipseudoclefts (full acceptance responses from 0.0% to 7.4%).

The result of the Kruskal-Wallis test in the subject semipseudocleft condition manifests a statistically significant difference between groups (H(3) = 56.040, $p < .001$). The responses of the control group to the items of both conditions are different from the answers of all three learner groups, as shown by the results of multiple pairwise comparisons (each of the L2 groups-control: $p < .001$). Between the three learner groups, no statistically significant differences is found (B2-C1: $p > .05$; B2-B1: $p > .05$; B1-C1: $p > .05$).

The corrections of the subject semipseudocleft made by the participants from all groups were found to predominantly convert the structure into a pseudocleft by adding a *wh*-word to the initial

position of the sentence, such as in 81.4% (35 out of 43) of the modifications of the control group. The B1 learners displayed more complicated solutions. For the subject semipseudocleft, the replacement with a pseudocleft occupied 40% of all corrections (24 out of 60), followed by substitution with a standard cleft (20.0%, 12 occurrences) and alteration into a standard-cleft-like structure without any element in C (16.7%, 10 occurrences). Regarding the B2 learners, 63.2% (36 out of 57) of the corrections substituted the subject semipseudocleft with a pseudocleft, and 21.1% (12 out of 57) with a standard cleft, with another 5 occurrences (8.78%) of the standard-cleft-like structure with an empty C head. The C1 learners also mainly preferred turning the subject semipseudocleft into a subject pseudocleft (61.9%, 26 out of 42) or a subject standard cleft (31.0%, 13 out of 42).

In terms of the object semipseudocleft, 50.0% (14 out of 28) of the corrections made by the control group corresponded to the addition of a *wh*-word to the initial position of the sentence. The native speakers also used a simple sentence (35.7%, 10 out of 28) to replace the structure in the tested items. This solution also prevailed (55.4%, 36 out of 65) in the corrections made by the B1 group, while the replacement with a standard cleft only occupied 12.3% of the 65 corrections (8 occurrences), following the alteration into a simple sentence (15.4%, 16 occurrences). Among the corrections made by B2 learners, 77.6% (45 out of 58) were into a pseudocleft, 12.1% (7 out of 58) were into a standard cleft, and 10.3% (6 out of 58) were into a simple sentence. 77.6% of corrections made by the C1 learners replaced the structure with a pseudocleft, and 14.3% with a standard cleft.

Pseudocleft: Condition 5 and Condition 10

Table XI and Figure V display the results of the acceptability test of subject and object pseudoclefts.

Table XI: Acceptability rates for pseudoclefts

Group	Clefted Constituent	0	1	2	3
Control	Subject	0.0%	0.0%	3.2%	96.8%
	Object	0.0%	1.6%	3.2%	95.2%
B1	Subject	13.4%	6.0%	17.9%	62.7%
	Object	2.9%	5.9%	8.8%	82.4%
B2	Subject	3.3%	0.0%	3.3%	93.3%
	Object	3.3%	0.0%	3.3%	93.3%
C1	Subject	0.0%	0.0%	3.9%	96.1%
	Object	0.0%	0.0%	0.0%	100%
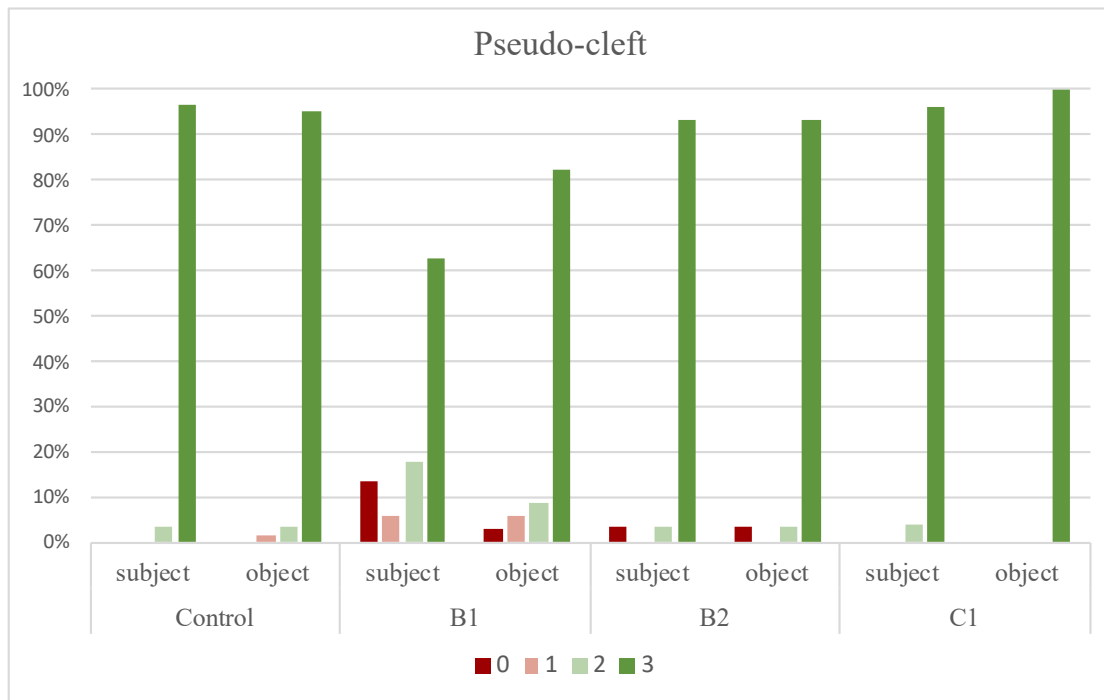
The pseudocleft seems to have been acquired relatively early by the learners, with a leap in the acceptance rate seen from the B1 group to the B2 group.

The full acceptance rates of the two conditions by the control group were both also above 95%. The B1 learners found 62.7% of the sentences involving a subject pseudocleft totally grammatical and 82.4% of those with an object pseudocleft were equally considered completely acceptable. However, the B2 learners displayed an even higher acceptance of the structure (93.3% marked 3 for both subject and object pseudoclefts). The complete acceptance rates of the subject and object pseudoclefts by the C1 group were also both above 95% and at the ceiling level.

The results of the Kruskal-Wallis test for the subject pseudocleft condition demonstrate a statistically significant difference between groups (H(3) = 44.839, $p < .001$). As for multiple pairwise comparisons, the B1 group displays a significant difference from each group at higher levels (B1-B2: $p < .001$; B1-C1: $p < .001$) and from the native speakers (B1-control: $p < .001$). Concerning the object pseudocleft, there is also a statistical significant difference between groups (H(3) = 14.254, $p = .003$). The pairwise comparisons show that the B1 group behaved differently from the C1 group ($p = .002$) and the control group ($p = .036$) in this condition. No statistically significant difference is found

between the B2, C1 and control groups for either the subject pseudocleft or the object pseudocleft ($p > .05$).

The control group made no relevant corrections to the pseudocleft, except for one incidence of an object pseudocleft in which the *wh*-word *o que* "what" was substituted with a demonstrative pronoun plus the complementizer: *aquilo que* "the one that". Among the 20 corrections to the subject pseudocleft, the B1 learners replaced the structure with a *wh*-cleft (35%, 7 occurrences) or a standard cleft (30%, 6 occurrences). Furthermore, 50% of the corrections to the object pseudocleft by the B1 group replaced the structure with a *wh*-cleft (5 out of 10). Neither the B2 nor the C1 learners made any relevant modifications to sentences with either clefted constituent.

5.2 Results of Test 2

Test 2 evaluates the judgment on agreement patterns between the plural clefted subject and the embedded verb in the clefted clause in *é-que* clefts, standard clefts, and pseudoclefts.

É-que cleft: Condition 1 and Condition 2

The acceptance rate of plural subject *é-que* cleft with and without the verbal inflection of *ser* "be" in agreement with the plural subject are presented in Table XII and Figure VI.

Group	*Ser* Inflection	0	1	2	3
Control	-	0.0%	1.6%	12.7%	85.7%
	+	44.4%	23.8%	31.7%	0.0%
B1	-	53.1%	7.8%	29.7%	9.4%
	+	33.3%	18.2%	21.2%	27.3%
B2	-	38.3%	13.3%	20.0%	28.3%
	+	47.5%	15.3%	11.9%	25.4%
C1	-	25.6%	6.7%	20.2%	47.5%
	+	41.2%	29.4%	23.5%	5.9%
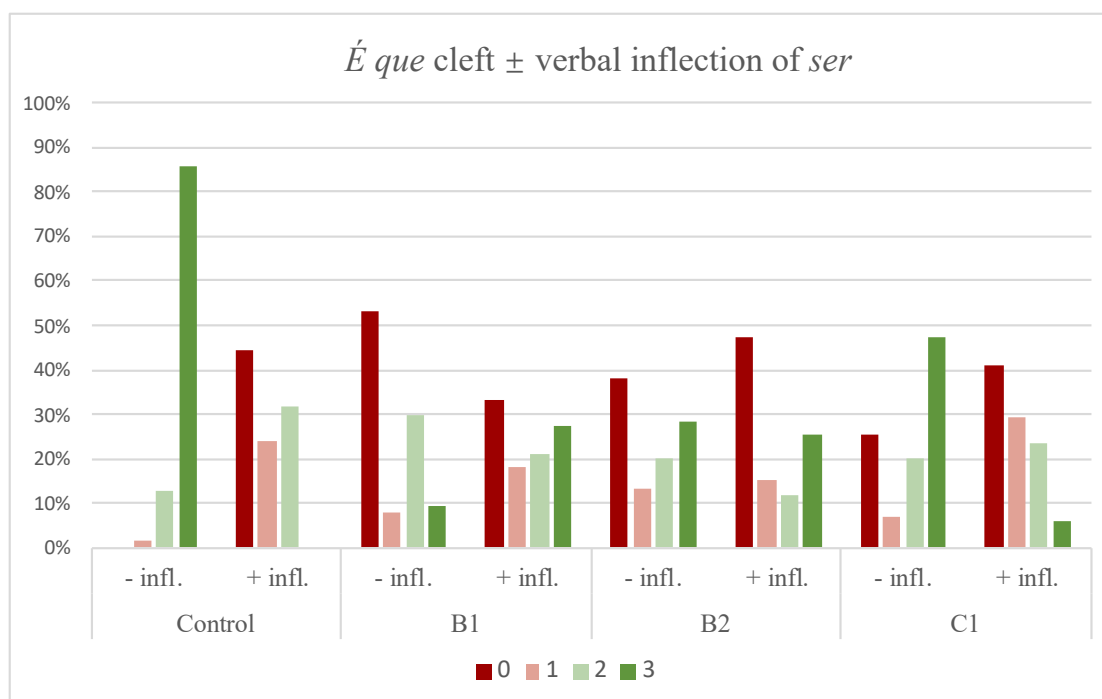
The control group accepted the plural subject *é-que* cleft and rejected the ungrammatical configuration, as expected. The plural subject *é-que* cleft was not highly accepted by the learners at the B1 and B2 levels, whereas the C1 learners found it more acceptable. When the verb *ser* "be" in the *é-que* expression agreed with the plural subject, the sentences were rejected by all groups of participants, even though not at the same levels.

Compared to the high acceptance of the plural subject *é-que* cleft by the control group (85.7% were fully accepted), it was possibly harder for the learners to completely accept the absence of the subject-verb agreement (B1: 9.4%; B2: 28.3%; C1: 47.5%). In turn, the plural subject *é-que* cleft with verbal inflection was completely rejected by the native speakers but accepted by a few learners, whereby the B1 learners showed a slightly higher acceptance rate (27.3%) than groups B2 (25.4%) and clearly higher that the C1 group (5.9%).

The results of the Kruskal-Wallis test conducted on condition 1 evidence the existence of a statistical significance between groups (H(3) = 101.552, $p < .001$). The B1 and B2 learners performed distinctly from the native speakers and the C1 learners in that the pairwise comparisons between groups B1 and control, B1 and C1, B2 and control and B2 and C1 all identified statistically significant differences ($p < .001$). As for the plural subject *é-que* cleft with verbal inflection, the overall Kruskal-

113

Wallis test does not show a significant difference between groups (H(3) = 7.378, *p* = .061).

All 7 corrections to the plural subject *é-que* clefts made by the native speakers only involved articles and the substitution of pronouns for complete noun phrases; in other words, these corrections did not concern the structure itself. The B1 learners made more corrections to the plural subject *é-que* clefts compared to native speakers (48 occurrences in total) and showed a tendency to modify the sentence into a simple sentence (56.3%, 27 occurrences) or a standard cleft (33.3%, 16 occurrences). Nine corrections (18.8%) of *é-que* in the structure into *são que* (third-person plural inflection of the verb *ser* "be") were also observed, showing that the learners were making agreement in person but not tense agreement. The B2 group, likewise, preferred a standard cleft (48.7%, 19 out of 39) and a simple sentence (38.5%, 15 out of 39), while three occurrences of the structure were modified into an *é-que* cleft with agreement between the subject and the verb *ser* "be" in the *é-que* expression. Among the 12 corrections made by the C1 learners, 75% (9 occurrences) turned the structure into a standard cleft and 25% (3 occurrences) turned it into a simple sentence.

When the *é* in the *é-que* expression was ungrammatically inflected, the native speakers tended to adapt the structure into a standard cleft (36.4%, 20 out of 55), an *é-que* cleft (32.7%, 18 out of 55), or an inverted pseudocleft (25.5%, 14 out of 55), among which 42.9% (6 occurrences) unconventionally showed agreement between the clefted subject and the embedded verb in the clefted clause. The B1 group replaced the structure with a standard cleft (43.2%, 19 out of 44) or a simple sentence (22.7%, 10 out of 44) as well as 6 occurrences (13.6%) of an inverted pseudocleft with subject-verb agreement and 4 corrections irrelevant to the structure. Furthermore, 84.4% of the corrections (38 out of 45) made by the B2 learners used a standard cleft. The C1 learners continued to prefer a standard cleft (70.5%, 31 out of 44), whereas 15.9% of the corrections (7 out of 44) changed the ungrammatical structure into an *é-que* cleft.

Standard Cleft: Condition 3 and Condition 4

Conditions 3 and 4 tested the acceptability of the standard cleft with and without the agreement in person between the plural subject and the embedded verb: the results obtained are presented below in Table XIII and Figure VII.

Table XIII: Acceptability rates for standard clefts ± agreement

Group	Agreement	0	1	2	3
Control	+	1.6%	1.6%	11.3%	85.5%
	-	38.1%	22.2%	34.9%	4.8%
B1	+	2.9%	0.0%	13.2%	83.8%
	-	30.9%	5.9%	23.5%	39.7%
B2	+	3.3%	0.0%	0.0%	96.7%
	-	44.1%	10.2%	15.3%	30.5%
C1	+	5.9%	0.0%	0.0%	94.1%
	-	49.0%	15.7%	21.6%	13.7%

Figure VII: Acceptability rates for standard clefts ± agreement
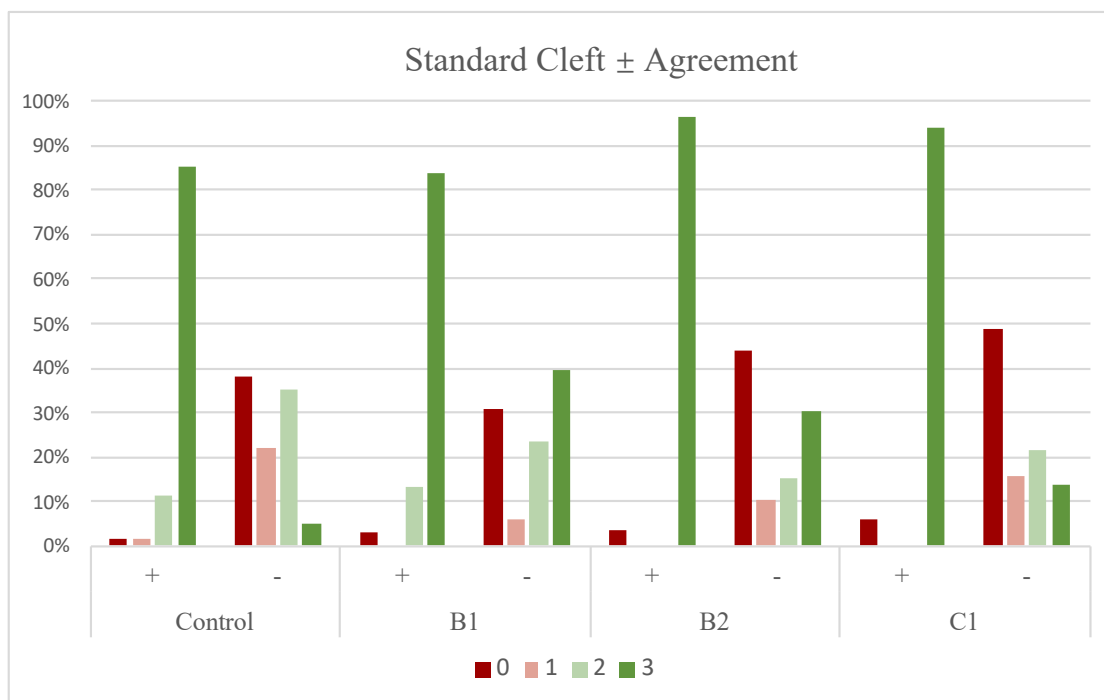
As expected, the control group clearly accepted the plural subject standard cleft with grammatical agreement (85.5% were fully accepted) and rejected the structure when such an agreement was absent (4.8% marked 3). The plural subject standard cleft with agreement seemed to be clearly accepted by the learners of all groups, with acceptance rates ranging from 83.8% for B1 learners to 96.7% for B2 learners. When the subject-verb agreement no longer occurred, these sentences had a considerably low

acceptance rate by the C1 group (13.7%), whereas the B1 and B2 learners accepted them relatively more (39.7% and 30.5%, respectively).

The overall result of the Kruskal-Wallis test evidences that, for the plural subject standard cleft with subject-verb agreement, no statistical significance can be seen across groups (H(3) = 7.084, p = .069). On the plural subject standard cleft without subject-verb agreement, the general outcome of the Kruskal-Wallis test shows a statistically significant difference between groups (H(3) = 13.113, p = .004). Significant differences are observed between groups B1 and C1 (p = .010), as well as between the groups B1 and control (p = .016).

Of the 9 corrections of a plural subject standard cleft with agreement made by the native speakers, 77.8% were a *wh*-cleft with an unconventional subject-verb agreement (7 occurrences). The other corrections focused on the pronoun. The B1 learners also corrected the standard cleft into a *wh*-cleft with subject-verb agreement (66.7%, 4 out of 6). The B2 and C1 learners made few corrections.

The ungrammatical structure was mainly corrected into a grammatical standard cleft, namely, with agreement, occupying 80% of the corrections made by the control group (48 out of 60), 66.7% made by the B1 group (24 out of 36), 95.1% made by the B2 group (39 out of 41) and 94.9% made by the C1 group (37 out of 39). In addition, 10% of the corrections introduced by the native speakers (6 occurrences) also adapted the structure into a *wh*-cleft with a non-standard subject-verb agreement. Also, 19.4% of the corrections (7 out of 36) made by the B1 speakers replaced the structure with a *wh*-cleft, which only occurred twice, respectively, in the corrections of learners of both B2 and C1 groups.

Pseudocleft: Condition 5 and Condition 6

Conditions 5 and 6 evaluated the judgement on the agreement pattern of the pseudocleft, namely, the standard agreement pattern, in which the embedded verb does not agree with the clefted plural subject in person but rather with the *wh*-word, and the non-standard agreement with the verb agreeing with the plural subject. The acceptance rates are shown in Table XIV and Figure VIII.

Table XIV: Acceptability rates for pseudoclefts ± agreement

Group	Agreement	0	1	2	3
Control	-	0.0%	0.0%	0.0%	100.0%
	+	33.9%	22.6%	37.1%	6.5%
B1	-	7.5%	13.4%	13.4%	65.7%
	+	15.9%	10.1%	20.3%	53.6%
B2	-	16.9%	1.7%	6.8%	74.6%
	+	11.7%	1.7%	6.7%	80.0%
C1	-	29.4%	0.0%	23.5%	47.1%
	+	28.0%	6.0%	24.0%	42.0%

Figure VIII: Acceptability rates for pseudoclefts ± agreement
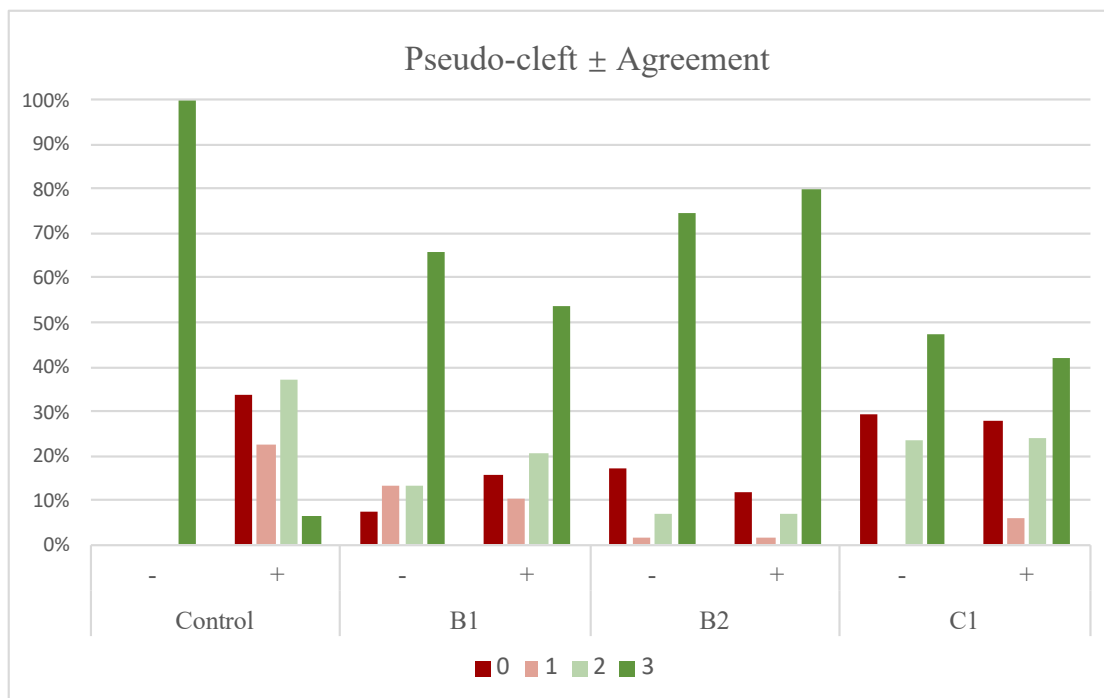
When the clefted subject was plural, a pseudocleft seemed to be more challenging for the learners at all levels with the agreement pattern.

While it was 100% accepted by the native speakers, a plural subject pseudocleft without

117

agreement did not appear to be favored by the learners at levels B1 (full acceptance: 65.7%) and B2 (full acceptance: 74.6%), while the complete acceptance of C1 learners was surprisingly low (47.1%). The complete acceptance of a plural subject pseudocleft with a non-standard agreement pattern between the clefted subject and the embedded verb was observed, although at a much smaller scale, in the responses of the native speakers (6.5%), and was persistent across all learner groups (B1: 53.6%; B2: 80%; C1: 42.0%).

For the plural subject pseudocleft without agreement, the Kruskal-Wallis test shows a statistically significant difference across groups (H(3) = 40.981, p < .001). The pairwise comparisons indicate that the learners significantly differ from the native speakers (B1-control: p < .001; B2-control: p = .007; C1-control: p < .001); meanwhile, a significant difference is also attested between the B2 learners and C1 learners (p = .012). When non-standard agreement between the clefted plural subject and the verb in the clefted clause occurred, there is also a significant difference between groups (H(3) = 54.784, p < .001). All the learner groups diverged significantly from the control group (B1-control: p < .001; B2-control: p < .001; C1-control: p = .008). C1 and B2 learners also contrasted with each other (p = .002).

The native speakers made no corrections to the plural subject pseudocleft without agreement. In 17 corrections, the B1 learners changed the same structure into a *wh*-cleft with agreement (23.5%) or a pseudocleft with non-standard agreement (17.6%), whereas 17.6% only modified the structure by adding a comma after the clefted clause. Nine out of the 14 corrections made by the B2 speakers (64.3%) substituted the structure with a pseudocleft with non-standard subject-verb agreement, and in another four cases (28.6%) the verb *foram* "were" was replaced by *foi* "was". The same type of corrections prevailed in the responses of the C1 learners (63.0%, 17 out of 27), who also corrected the structure into a pseudocleft with non-standard agreement (22.2%, 6 out of 27).

As for the pseudocleft with agreement between the clefted subject and the embedded verb, amongst the 59 corrections of the native speakers, 57 (96.6%) transformed the structure into a pseudocleft without such agreement, corresponding to the standard EP grammar, while the other two cases (3.4%) turned it into a *wh*-cleft. In the 27 corrections made by the B1 learners, 37.0% (10 occurrences) adapted the structure into a pseudocleft without agreement, among which there were two incidences of changing the verb in past and plural form into present and singular form. Eight corrections (29.6%) changed the pseudocleft into a *wh*-cleft with subject-verb agreement, which is also

not standard EP grammar. Six out of the eight corrections (75%) made by the B2 learners corrected the structure into a pseudocleft without agreement. Meanwhile, 44.4% (12 out of 27) of the corrections of the C1 learners make the same adaptation, while 18.5% (5 occurrences) were changed into a *wh*-cleft with non-standard agreement; additionally, there were 10 occurrences (37.0%) in which the verb *foram* "were" was replaced by *foi* "was".

5.3 Results of the Tests on English Clefts

The results of complementary English tests – the acceptability test of the English *it*-cleft, *it*-cleft like structure with empty C and pseudocleft and the agreement pattern test of the *it*-cleft -, are presented below in Table XV and XVI.

Note that the labels of the groups do not correspond to the English levels of the participants but indicate the same groups of participants as in the Portuguese experiment.

Table XV: Tests on English Clefts – Test One

Condition	B1				B2				C1			
	0	1	2	3	0	1	2	3	0	1	2	3
Subject *it*- cleft	10.9%	2.2%	8.7%	78.3%	7.5%	0.0%	0.0%	92.5%	5.9%	0.0%	0.0%	94.1%
Object *it* -cleft	0.0%	2.2%	8.9%	88.9%	5.0%	0.0%	2.5%	92.5%	0.0%	0.0%	2.9%	97.1%
Subject *it*- cleft with empty C	44.4%	15.6%	13.3%	26.7%	57.5%	12.5%	15.0%	15.0%	47.1%	26.5%	17.6%	8.8%
Object *it*- cleft with empty C	44.4%	26.7%	6.7%	22.2%	60.0%	15.0%	15.0%	10.0%	41.2%	35.3%	14.7%	8.8%
Subject pseudocleft	0.0%	4.3%	13.0%	82.6%	5.0%	0.0%	0.0%	95.0%	0.0%	0.0%	5.9%	94.1%
Object pseudocleft	0.0%	4.3%	8.7%	87.0%	2.5%	0.0%	2.5%	95.0%	0.0%	0.0%	0.0%	100.0%

Table XVI: Tests on English cleft – Test Two

Condition	B1				B2				C1			
	0	1	2	3	0	1	2	3	0	1	2	3
Plural subject *it*- cleft	0.0%	0.0%	8.7%	91.3%	7.5%	2.5%	0.0%	90.0%	9.1%	0.0%	0.0%	90.9%
Plural subject *it*- cleft with agreement	8.7%	2.2%	10.9%	78.3%	23.7%	0.0%	5.3%	71.1%	23.5%	2.9%	17.6%	55.9%

The results of the two English tests, which were only applied to learners of Portuguese, showed that in L2 English, both the *it*-cleft and the pseudocleft when clefting subject and the object were well accepted since the B1 level. Although some of the B1 learners accepted the *it*-cleft structure without a filled C, the complete acceptance rate was above 20% and the percentage consistently dropped and stayed below 10% in the responses of the C1 learners. However, the agreement of the clefted plural subject and the copular verb in the *it*-cleft sentences, which is ungrammatical in the language, seemed to be persistently fully accepted by the learners of all levels (B1 - 78.3%, B2 - 71.1%, C1 - 55.9%).

Chapter 6 Discussion

This chapter discusses the results of the experimental task and of the corpus search while revisiting the working hypotheses proposed in Chapter 3.

6.1 Discussion of the data

The experimental results show the panorama of the acquisition of the EP *é-que* cleft, standard cleft, semipseudocleft, and pseudocleft by L1-MC learners. The data indicate that, generally, the L1-MC learners were already aware from the B1 level that the EP standard cleft and pseudocleft constructions can be used to focalize both the subject and the object. However, the learners encountered persistent difficulties regarding the *é-que* cleft and semipseudocleft constructions, even at a high level of proficiency (C1). The results of Test 1 demonstrate that the acceptance of all types of cleft structures, except the semipseudocleft and object *é-que* cleft, had a certain divergence amongst learners of different levels. Progress was seen in the acceptance of the *é-que* cleft, the standard cleft, and the pseudocleft; however, the results from Test 2 show that the learners had not completely acquired the grammar of the pseudocleft constructions of the target language.

Hypothesis 1 predicted a delayed acquisition of the *é-que* cleft due to (a) the existence of a crystallized *é-que* expression associated to C or (b) difficulties in the association of the [+ focus] feature to such an expression. The learners' performance on the *é-que* cleft by the groups, in general, deviated considerably from that of the native speakers. In Test 1, the learners of levels B1 and B2 uniformly rejected both the subject *é-que* cleft and the object *é-que* cleft. Meanwhile, the advanced learners from the C1 group not only accepted the subject *é-que* cleft in a native-like way, but they also showed a slightly higher acceptance rate of the object *é-que* cleft compared to the control group, who welcomed the subject *é-que* cleft but strongly rejected the object *é-que* cleft, mainly changing it into a simple sentence or an object pseudocleft.

The acceptance rates for the grammatical plural subject *é-que* cleft by the three groups of learners showed the same increasing tendency from B1 to C1 as for the single subject *é-que* cleft. When the verb *ser* "be" in the lexicalized expression *é-que* ungrammatically agreed with the clefted plural

subject in person, as shown by the results of Condition 2 of Portuguese Test 2, although the complete acceptance rates declined from B1 to C1, the acceptance rates of this configuration by the B1 and B2 learners were not substantially high (27.3% and 25.4%, respectively).

Hence, for L1-MC speakers, it seems possible to acquire the subject *é-que* cleft when they reach level C despite the delayed development at an intermediate stage, as predicted in Hypothesis 1. The test results show that a leap in acquisition is seen from level B2 to level C1. The non-native like acquisition of the object *é-que* cleft, however, may suffer from an overgeneralization when learners reach level C and recognize the usage of the subject *é-que* cleft.

Given the distinct results for the subject *é-que* cleft of groups B1 and B2 in comparison with those of the control group in both tests, we suggest that learners at level B do not reject the *é-que* cleft, either that of the subject or of the object, for the same reason as the native speakers. As speculated in Hypothesis 1, the rejection of the *é-que* cleft by the B1 and B2 learners could be due to a lack of knowledge of the crystallization of the *é-que* expression or a failure to reassemble the [+ focus] feature to the expression. If the learners had failed to acquire *é-que* as a lexically filled C, they would have accepted the *é-que* cleft with verbal inflection due to the subject-verb agreement established between the verb *ser* "be" and the clefted plural subject; however, the results of Test 2 showed the opposite. Although in Test 1 the substitution of *é* in the *é-que* expression with *foi* "was" occurred twice in the subject cleft and twice in the object cleft, we do not consider it a relevant pattern of the learners. Moreover, Kou (submitted) showed that learners from the B1 level have been conscious of the lexicalization of the *é-que* expression. In her experimental work, root questions with *wh*-movement that involved the lexicalized expression *é-que*, in which the main verbs were in the simple past tense, showed considerably high complete acceptance rates by the participants of all levels tested (B1: 95.6%, B2: 97.5%, C1: 98.5%). In addition, the occurrences of the mixed structure "*foi…é que*" in the corrections, persistent from level B1 to level C1, seem to imply that in the attempt to produce a standard cleft, the intact *é-que* expression was considered to be the lexicalizer of the C head and that, consequently, such an expression was already associated to C.

In accordance with the Feature Reassembly Hypothesis (Lardiere, 2008, 2009), given the inexistence of an equivalent expression bearing the [+ focus] feature in both L1 MC and L2 English, learners cannot resort to mapping when encountering the *é-que* cleft. From level B1, the learners

acquired *é que* as a lexically filled C in *wh*-interrogatives, and thus they were able to associate the expression to C. Therefore, one can deduce that the delayed acquisition of the *é-que* cleft is due to the difficulty in assembling the [+ focus] feature with the *é-que* expression. Hypothesis 1 (b) is borne out.

The unexpected rejection by native speakers of the object *é-que* cleft leads us to suspect that the contexts in which such a structure naturally occurs may be playing a more critical role. Vercauteren's (2010) corpus study of nonstandard EP varieties of CORDIAL-SIN also documented a massive contrast between the occurrences of the subject *é-que* cleft (628 times) and that of the object *é-que* cleft (71 times). According to Vercauteren (2010), the clefted constituent in the *é-que* cleft can not only bear the informational focus, but also bears certain characteristics of topic and, in such topic structures, the *é-que* expression marks the value of contrast and exhaustivity. Lobo, Santos & Soares-Jesel (2016) also reported a clear subject-object asymmetry of *é-que* clefts in the elicited production by both adults and children and in the spontaneous production by adults in Santos corpus (Santos, 2006). Hence, the test results of this work continue to arouse a certain interest in a further investigation into the syntactic nature and semantic/pragmatic properties of the *é-que* cleft.

In the second language acquisition corpus COPLE2, no *é-que* cleft was spotted, except in one occurrence of the mixed structure "*foi…é que*". As presented in the previous section, although with a limited number of tokens, the mixed structure "*Foi… é que*" was also found in the corrections made by the non-native participants in the experimental task.

Interestingly, there is a considerable presence of the mixed structure "*Foi…é que*" in non-standard varieties of EP. Costa and Lobo (2009) noted that in the annotated dialect corpus, CORDIAL-SIN (Syntax-oriented Corpus of Portuguese Dialects), this structure not only occurs in non-standard varieties, but also is not geographically circumscribed. Vercauteren (2010) reported 122 occurrences of the mixed structure in the same corpus, and stated that it is widespread and only absent in the data of several locations (out of more than 200) in the Portuguese territory.

Hypothesis 2 expected that standard clefts are more precociously acquired. The results of Test 1 indicate that both the subject standard cleft and the object standard cleft are well accepted at least since level B1, which confirms this hypothesis. The standard cleft is also one of the most frequently produced cleft structures by L1-MC learners in COPLE2 (6 out of 15 occurrences). As conjectured in

Hypothesis 2, this early development of the standard cleft is due to the abundance of this strategy in the EP input and the formal instruction that learners receive in the classroom, which was confirmed by some participants in verbal communication after the test.

No subject-object asymmetry was found in the responses of the learners. In fact, the learners from B1 and B2 even showed a higher acceptance of the object standard cleft than its counterpart, the subject standard cleft, and C1 learners rated both structures as 100% grammatical. Hypothesis 2 (a), arguing that the appearance of a subject-object asymmetry could occur during the earlier stages of acquisition according to the L1 transfer scenario (Hermas, 2014), is thus refuted.

The results of the complementary English test showed that the subject *it*-cleft and object *it*-cleft were well acquired by the learners, while the object it-cleft was also better received. Such results are not only congruent with the results of the EP test, but also help to reason that it is possible that the learners had solved an essential problem of acquiring the EP standard cleft when acquiring their L2. In other words, the [+ focus] feature appears to have already been successfully disentangled from a lower functional head in the L1 and associated to the C head during the acquisition of L2 English. Subsequently, in the acquisition of EP as the L3, such reassembly of the feature or the bundle of features in C could be mapped directly from the L2. Therefore, Hypothesis 2 (b) is borne out, in that the results do not question a certain L2 facilitation effect or the possibility of the L2 having the potential to be at least one of the sources of transfer, if not the only one, as predicted by the L2 Status Factor (Bardel and Folk, 2007, 2012), the CEM (Flynn et al. 2004; Berkes and Flynn, 2012), the TPM (Rothman, 2011, 2013, 2015), and the SM (Slabakova, 2016).

Interestingly, the native speakers demonstrated a relatively low level of acceptance of both the subject standard cleft and the object standard cleft, partially attributed to problems of pronouns or articles. The prevailing modification made by the native speakers on the subject standard cleft was transforming it into a *wh*-cleft, while the results of their corrections of the object standard cleft were much more diverse, including *wh*-clefts, pseudoclefts, semipseudoclefts and simple sentences. The learners' few corrections, meanwhile, only showed a prevalence of the *wh*-cleft and the standard-cleft-like structure with an empty C by the less advanced (B1) group. Such a discrepancy implies that the native speakers are more sensitive to the association of different types of cleft structures and the syntactic function of the clefted constituent.

Hypothesis 3 predicted that learners at less advanced levels may accept a standard-cleft-like structure without any element in C. In scenario (a), less advanced learners only accept a subject standard-cleft-like structure due to the L1 transfer of bare-*shì* subject clefts (Paul & Whitman, 2008) and the C domain of cleft structures remains inactive. Meanwhile, in scenario (b), these learners accept the object standard-cleft-like structure due to the L2 transfer of a possible null complementizer in the English object *it*-cleft. However, the test results did not meet either of the expectations.

Considering the low acceptance rates of the standard-cleft-like structure with an empty C by all groups of participants, the learners seem to know at an early stage that in an EP cleft structure initiated by a copula, the position of C cannot be left empty; in other words, there must be an overt element in the head of CP. Although the acceptance of an empty C shows a progressive decline from the learners of B1 to B2 and then to C1, the B1 learners' full mark rates were already below 30%. However, a slight subject-object asymmetry was attested in the B1 learners' responses, a case in which the subject standard-cleft-like structure was more strongly rejected than its object counterpart (percentage of marking 0: 51.5% vs. 37.9%), which might seem to be indicating the L2 transfer from English. In COPLE2, one occurrence of adjunct standard-cleft-like structure with an empty C was found produced by a B2 learner.

However, in the complementary English test, where the object *it*-cleft with a null complementizer should be better received than its counterpart where the subject is clefted, surprisingly, the B1 and B2 learners even accepted the subject *it*-cleft with a null complementizer more than that of the object.

Although the standard-cleft-like structure with an empty C was strongly rejected by all learner groups, such a structure was produced in the corrections of the *é-que* cleft, the standard cleft, and the semipseudocleft by some B1 learners and a few B2 learners. In such corrections, the total occurrences of the subject standard-cleft-like structure outnumbered the structure attempting to cleft an object (B1: 14 vs.7, B2: 9 vs. 3), instantiating a possible L1 transfer effect.

In sum, the majority of the learners had acquired that in an EP cleft structure, the [+focus] feature is associated to a projection in the C domain, and that in EP the complementizer cannot be null. The L2 transfer effect from English in this case does not substantially appear, while a possible L1 negative transfer effect resulting from the MC bare-*shì* subject cleft might remain in the L3 grammars of some

learners of lower levels of proficiency. Hypothesis 3 is thus contested.

Hypothesis 4, which expected difficulties in the acquisition of the semipseudocleft, is borne out by the test results.

The semipseudocleft was rejected by the non-native speakers irrespective of the clefted constituent. The absence of this cleft structure produced by the L1-MC learners in COPLE2 sustains the challenging nature of its acquisition. Since neither L1 MC nor L2 English have a copular verb lexicalizing the [+focus] feature and marking the left peripheral boundary of the vP as in L3 EP (Lobo, Santos & Soares-Jesel, 2012), this structure is naturally challenging due to the lack of a mapping resource. It is also worth noticing that according to the participants, such a structure had never been introduced in class.

Furthermore, although only unaccusative verbs occurred in the test items testing subject semipseudoclefts, the structure still received limited acceptance from the native speakers. The object semipseudocleft was, in turn, more welcomed by the native speakers. Although the acceptance rate of the object semipseudocleft was not considerably elevated, some of the native speakers in effect produced the structure while correcting other types of cleft structures. The fact that the semipseudocleft is an oral structure whereas the test was in written form might contribute to explain the native speakers' low acceptance.

Hypothesis 5, which foresaw a better performance of learners in terms of the pseudocleft, irrespective of the clefted constituent, is confirmed. Although the B1 learners showed relatively lower acceptance rates of the subject pseudocleft (62.7%) and the object pseudocleft (82.4%), mainly adapting the sentences into a *wh*-cleft or a standard cleft, the performances of the learners from groups B2 and C1 were at the ceiling level. A leap in the acquisition of such a strategy was seen from level B1 to B2. No obvious subject-object discrepancy was seen, while the object pseudocleft seemed marginally more accepted by the B1 and C1 learners than the subject pseudocleft. In the production of L1-MC learners, the pseudocleft is one of the two most recurrent structures in COPLE2 (7 out of 15 occurrences), on a par with the standard cleft.

In the complementary English test, the same leap was noticed between groups B1 and B2, although group B1 already demonstrated the full mark rates above 80%. In general, the English object

pseudocleft received a marginally better acceptance than the English subject pseudocleft, coherent with the results of the EP test.

Our position concerning the syntax of the pseudocleft holds that in MC, English and EP, the pseudoclefts all involve a small clause containing an embedded *wh*-clause. Hence, it is not surprising that the acquisition is facilitated due to the sufficiency of sources for mapping the structure onto the corresponding structure in the L3 EP. The sources for mapping can be solely from L1, L2 or one linguistic system typologically similar to EP, which is L2 English in this case, as predicted by the L1 transfer scenario (Hermas, 2014), the L2 Status Factor (Bardel and Folk, 2007, 2012) or the TPM (Rothman, 2010, 2011, 2013, 2015). Also, the facilitative effect can come from the cumulative or selected knowledge of both previously acquired languages, as the CEM (Flynn et al. 2004; Berkes and Flynn, 2012) or SM (Slabakova, 2017) expect. However, despite the considerably high acceptance, when compared with the standard cleft, the pseudocleft seems to have been somewhat more challenging for the B1 learners.

In spite of the low frequency of L1 spontaneous speech production (Lobo, Santos, Soares-Jesel, 2016), the native speakers clearly received the pseudocleft, regardless of whether the clefted constituent was the subject or the object, both with a full acceptance rate beyond 95%.

Hypothesis 6 predicted that the challenges of the agreement patterns of the standard cleft and the pseudocleft may be persistent even if such structures are well accepted in the earlier stages, which mirrors an incomplete acquisition of the grammar. Scenario (a) conjectured that learners either take both the standard cleft and the pseudocleft as a structure involving extraction (Lobo, 2006), if they consistently agree the clefted subject with the embedded verb; alternatively, they consider both structures an identificational structure involving a small clause (Lobo, 2006) if they consistently maintain the embedded verb 3rd person singular. Scenario (b) proposed another possibility: learners might make a semantic agreement, due to the anaphoric relation between the *wh*-word and the clefted constituent, if they systematically prefer to have agreement between the embedded verb and the clefted subject.

This hypothesis is partially corroborated by the data of the EP Test 2, as the agreement patterns did seem problematic even until the highest level of proficiency tested, whereas the assimilation of the

two structures did not constantly occur.

For the *é-que* cleft, only C1 learners showed a native-like performance in rejecting inflection on *é* and keeping the expression intact.

Concerning the plural subject standard cleft, although the grammatical subject-verb agreement was prevalently welcomed, when the agreement ceased to appear, the learners of groups B1 and B2 showed a slightly better acceptance of the ungrammatical configuration. However, only the B1 group showed a statistically significant difference when compared to the C1 group and the native speakers. One can reason that, from level B2, the learners started to form a grammar that was similar to that of the native speakers, that is, the embedded verb must agree with the clefted constituent in number. In the English test, the ungrammatical agreement between the verb *be* and the clefted plural subject was also well received by the learners of all levels.

The canonical agreement pattern of the pseudocleft received 100% full acceptance by the native speakers, whereas a few also accepted the agreement between the clefted constituent and the embedded verb, as in some nonstandard varieties. However, the case of the learners' performance in terms of the plural subject pseudocleft was not so clear as that of the previous two structures. The B1 and B2 learners seemed to accept both patterns, i.e. canonical agreement of the embedded verb and the *wh*-word and non-canonical agreement between the embedded verb and the clefted plural subject. The B2 learners showed higher acceptance, and their full mark rate of the non-canonical pattern even reached 80%. The C1 learners did not tend to fully accept both the canonical and the non-canonical agreement, showing low acceptance rates of 47.1% and 42.0%, respectively. In sum, the learners, especially those of level B, manifested a similar and relatively high acceptance of both agreement patterns. The non-standard agreement pattern of pseudoclefts, besides the results of the acceptability judgement task, was not only witnessed in some participants' corrections, but also attested twice in the spontaneous production in COPLE2.

This phenomenon suggests that due to the anaphoric relation established between the clefted constituent and the *wh*-morpheme, the learners may have been establishing the correct mental representation of the syntactic structure and, at the same time, effectively making the semantically motivated agreement, but only when encountering a pseudocleft. Also, it is observed that the learners

did not seem to be sensitive to the referential hierarchy of pronoun > DP > CP-Q (Tavares, 2005), which may condition the verbal agreement in nominal copulative sentences in EP.

6.2 Summary of the discussion

In general, the properties of certain EP cleft structures, namely, the *é-que* cleft, the standard cleft and the pseudocleft, can be acquired by L1-MC speakers by the time they reach level C, even though the latter two types can be acquired even earlier, at level B. In this study, the learners of all levels did not generally reject clefting an object with these structures. The standard-cleft-like structure with an empty C did not receive considerable acceptance, especially after level B1. Meanwhile, the semipseudocleft did not seem to have been acquired, even by learners of level C. Concerning the agreement pattern, even though progress was seen in terms of the *é-que* cleft and the standard cleft, the learners had not yet completely acquired the target grammar, especially regarding the agreement of the pseudocleft.

The results of the L3 EP tests coincide with those of the complementary L2 English tests, through which the learners of all levels manifested a high acceptance of the object cleft as well as a rejection of a cleft-like structure with an empty C, while the agreement pattern remained problematic for all levels. Overall, the results do not contradict a possible facilitation effect of L2 English on the acquisition of L3 EP.

Regarding L3 acquisition theories, our experimental data do not corroborate the (exclusive) L1 transfer scenario (Hermas, 2014) on account of the following results. First, no obvious subject-object asymmetry in learners' performance on the standard cleft was shown, even at the lowest level of proficiency, B1, in contrast to what the L1 transfer scenario expects, due to the impossibility of the association of the [+focus] feature to a functional projection in MC object clefts. Second, the standard-cleft-like structure without any element in C was not only rejected by all learners, but there was also no considerable preference for the items in which the subject was "clefted", although the B1 group showed a marginally better acceptance of the subject standard-cleft-like structure. If the L1 transfer scenario were borne out by our data, the acceptance rate of the subject standard-cleft-like structure would have overwhelmed that of its object counterpart, at least for B1 learners. Third, as the EP

pseudocleft shares an identical syntactic structure with the MC pseudocleft, the L1 transfer scenario would predict such a structure to be the most easily acquired; however, the B1 learners did not accept the EP pseudocleft perfectly as well as they accepted the standard cleft. Moreover, the agreement pattern of the pseudocleft remains problematic even until the C1 level.

As for the most challenging structure, the semipseudocleft, although in L1 MC the copula *shì* can function as a focus marker and emphasize any constituent to the right in a pattern of association with focus (Paul & Whitman, 2008), the learners did not seem to have associated such properties in L1 to the semipseudocleft in L3, since the similarities do not constitute the source of mapping due to the disparities of the syntactic configurations of the two structures in the two languages. The precocious acquisition of the standard cleft, together with the general rejection of the standard-cleft-like structure with an empty C, implies that the learners, from the level B1, are aware of the projection of C in the standard cleft and the association of the [+focus] feature to the C domain. This, in turn, more precisely entails that learners successfully untangled this feature from the low Foc^0 in MC and reassembled it onto the EP lexical items in a higher functional category, C. As the L2 English *it*-cleft was also considerably accepted, the process of the extrication from lower projections and the re-association of such a feature to the left periphery could be as early as during the learning of L2, and subsequently, the [+focus] feature in C could be mapped from the L2 to the L3 lexical items. Although the acquisition of the *é-que* cleft is hindered by the association of the [+focus] feature to the *é-que* expression, the occurrences of the mixed structure "*foi...é que*" suggest that, for some learners, even at a lower level of proficiency, the expression was already associated to the C head which it lexicalizes.

To sum up, the following hierarchy represents the level of acceptance of EP cleft structures: standard cleft >[51] pseudocleft > *é-que* cleft > semipseudocleft. In the light of the Feature Reassembly Hypothesis (Lardiere, 2008, 2009), the results of the experimental task, corroborated by the findings of the corpus search, suggest that for L1-MC learners, the main difficulty in the acquisition of the EP cleft structure involving the left periphery does not consist of the projection of C but rather the association of the [+focus] feature to the corresponding L3 lexical items in the C domain. As for the structure that involves the left periphery of the verb phrase, the semipseudocleft, the difficulties also

---

[51] The symbol > is intended to represent that the structure to the left of such symbol is more easily accepted earlier than the structure to the right.

lie in the association of the [+focus] feature to the L3 head lexicalized by the verb *ser* "be".

Chapter 7 Conclusion

The present study aimed at exploring the L3 acquisition of EP cleft structures by L1-MC learners in the light of the Feature Reassembly Hypothesis (Lardiere, 2008, 2009). In this regard, we investigated the syntax of such structures in the L1 MC, the L3 EP and the L2 English, and consequently, elaborated and applied an experimental task, which allowed us to answer the research questions presented in Chapter 1, resumed as below:

(i) what structural differences do EP clefts and MC clefts display?

(ii) are the syntactic differences a source of difficulties for the acquisition of EP clefts by MC speakers?

(iii) if so, can these learners overcome the potential obstacles?

(iv) what kinds of EP clefts present more difficulty for the learners?

(v) Will other languages previously acquired by learners influence the acquisition of EP?

For the first question, through a comparative analysis of the syntax of EP and MC clefts, we conclude that the cleft structures in EP and MC display the following main common properties and disparities: (1) the pseudoclefts of both languages share a similar base structure, a small clause selected by the copular verb and composed of a relative clause and the focused DP; (2) some EP cleft structures, namely, the *é-que* cleft and the standard cleft, involve the dislocation of the clefted constituent to the left periphery of the clause, which bears a [+focus] feature, whereas in the semipseudocleft the verb *ser* "be" functions as a focus marker on the boundary of the left periphery of the vP and focalizing all the elements within this domain. In MC cleft structures, the clefted constituent does not raise to the C domain, nor does the copula have such a function as in the EP semipseudocleft. In the bare-*shì* subject cleft, the [+focus] feature can only be associated to the prosodically prominent element immediately after *shì*, i.e. the subject. In the *shì…de* cleft, the [+focus] feature is associated to a projection below vP, FocP, of which the specifier position is the landing site for the clefted subject or adjunct, and the object bearing the focus stays in the original sentence-final position in light of the Nuclear Stress Rule in line with Cinque (1993). In terms of clefted constituents, in EP, the aforementioned cleft structures can focalize both subjects and objects, although in semipseudoclefts only an object or a post-verbal

subject can be clefted; in MC, the object cannot be clefted in bare-*shì* clefts or in *shì* V O *de* clefts, as opposed to the subject. In sum, the differences between cleft structures in EP and MC lie, to a certain extent, in the association of the [+focus] feature to different functional categories.

To answer to the questions (ii), (iii) and (iv), we designed an off-line experimental task, composed of two subtests in EP, evaluating the acquisition of different cleft structures and their agreement patterns, as well as two complementary English subtests to assess the learners' knowledge of English clefts, as most of these L1-MC learners had acquired English as L2 before acquiring EP. All the tests were applied to three groups of learners from levels B1, B2 and C1, and the EP tests were also applied to a control group of native EP speakers.

The results of the experimental task helped us answer the three questions. In what concerns question (ii), the results suggested a scale of difficulty concerning the L3 acquisition of EP cleft structures by L1-MC learners:

standard cleft > pseudocleft > *é-que* cleft > semipseudocleft

As shown above, the *é-que* cleft and the semipseudocleft are more challenging for the learners.

In what concerns questions (iii), (iv) and (v), according to the statistics, in terms of the acquisition of the standard cleft and the pseudocleft, learners from B1 to C1 did not show any substantial rejection of the object cleft, nor any evident preference for the subject cleft. The *é-que* cleft seems to be acquired only at the C level, and there seems to be a subject-object asymmetry, since the object *é-que* cleft was more rejected by the C1 learners. However, the object *é-que* cleft was also unwelcomed among the native speakers. The disparity between the two languages concerning the possibility of clefting an object in a structure of extraction thus does not seem to halt the acquisition of EP cleft structures as a whole. In terms of the difficulty of the association of features, the precocious acquisition of the standard cleft, among other evidence, implies that the association of the [+focus] feature to the C domain does not appear to be an obstacle since the level B1. The results of the complementary English tests, in terms of clefting an object, of the dislocation of the clefted constituent to the C domain and of the fulfillment of the C head, coincide with the results of the EP tests. Consequently, we reason that such challenges can be overcome in the acquisition of EP clefts in an early stage and, plausibly, were already surpassed when a learner was acquiring the L2 English. However, the problem of consistently

following the standard agreement pattern of the EP pseudocleft is persistent from the level B1 to the level C1, implicating that the learners may make semantic agreement.

Moreover, as presented above, the delayed development of the acquisition of the *é-que* cleft and the semipseudocleft showed that the reassembly of the [+focus] feature to the L3 lexical items is what essentially hinders the acquisition. One could remind that no equivalent is found in MC or English for either of such two structures, and that the results of the standard cleft and the pseudocleft suggest that the difficulty does not consist of the projection of the category C. Hence, we conclude that the main obstacle of the acquisition of the *é-que* cleft and the semipseudocleft lies in the association of the [+focus] feature to the corresponding L3 lexical items, while the challenges concerning the acquisition of the standard cleft and the pseudocleft could be overcome by resorting to *mapping* from L2 English, in which one can find the identical combination of the feature and the corresponding functional categories. Consequently, our study does not question a possible L2 transfer effect. A complementary corpus search also corroborated the findings of the experimental task.

Nonetheless, the present work is exploratory and far from exhaustive. The syntactic structures of various clefting strategies in EP, MC and English and the nature of such structures at the syntax-discourse interface require further analysis. The sensitivity to the different clefting patterns considering the different syntactic function of the clefted constituents shown by the native speakers also captured our attention and can lead to more profound investigation of these structures. Apart from testing the subject and the object, more testing could be conducted on other clefted constituents with diverse syntactic functions. The contrasts shown in the responses of the native EP speakers and the learners in the experiment have to be further investigated. For instance, in the field of linguistic processing, it would be noteworthy to examine whether the L1 speakers and the advanced L2 speakers process certain structures distinctly. In terms of SLA, although it was not the goal of this work to corroborate any SLA model, future studies could focus on discerning the potential effects of the L1 or/and the L2 on the L3 acquisition of syntactic structures of EP.

References

Aoun, J. & Li, Y.-H. A. (2003) *Essays on the representational and derivational nature of grammar*. Cambridge, MA: MIT Press.

Ambar, M. (1999). Aspects of the syntax of focus in Portuguese. In G. Rebuschi and L. Tuller, *The grammar of focus* (23-53). Amsterdam: John Benjamins.

Barbosa, M. (2013). As construções pseudoclivadas: perguntas e respostas. In: *Textos Selecionados, XXVIII Encontro Nacional da Associação Portuguesa de Linguística*, Coimbra, APL, 2013, pp. 131-148, ISBN: 978-989-97440-2-8

Bardel, C., & Falk, Y. (2007). The role of the second language in third language acquisition: the case of Germanic syntax. *Second Language Research*, SAGE Publications, 2007, 23 (4), pp.459-484. 10.1177/0267658307080557. hal-00570736

Bardel, C., & Falk, Y. (2012). Behind the L2 status factor: A neurolinguistic framework for L3 research. *Third language acquisition in adulthood*, 61-78.

Belletti, A. (2004). Aspects of the low IP area. The structure of CP and IP. *The cartography of syntactic structures*, 2, 16-51.

Berkes, É., & Flynn, S. (2012). Further evidence in support. *Third language acquisition in adulthood*, 46, 143.

Chao, Y.R. (1968). *A grammar of spoken Chinese*. Berkeley : University of California Press

Cheng, L.L.-S. (2008). Deconstructing the shi...de construction. *Linguistic Review, 25*(3/4)*, 235–266. DOI: 10.1515/TLIR.2008.007

Chomsky, N. (1977). On Wh-movement. In P. Culicover, T. Wasw, & A. Akmajian (Eds.), *Formal syntax.* San Francisco, London: Academic Press.

Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. Greenwood Publishing Group.

Chomsky N. (1995). *The Minimalist Program*. Cambridge, MA: The MIT Press.

Chomsky, N. (2004). Beyond explanatory adequacy. In *Structures and beyond*, Adriana Belletti (ed.), 104-131. Oxford: Oxford University Press.

Chomsky, N. (2005). Three factors in language design. *Linguistic inquiry*, 36(1), 1-22.

Chomsky, N. (2005). On Phases. Ms., MIT.

Cinque, G. (1993). A null theory of phrase and compound stress. *Linguistic Inquiry 24*(2), 239-297.

CORDIAL-SIN: Corpus Dialectal para o Estudo da Sintaxe/Syntax-oriented Corpus of Portuguese Dialects. Lisboa, Centro de Linguística da Universidade de Lisboa. URL: http://www.clul.ulisboa.pt/en/10-research/314-cordial-s

Costa, J. (1998). Word Order Variation. A Constraint-based Approach, Ph.D. Dissertation, University of Leiden. Published by Holland Academic Graphics, The Hague.

Costa, J. & Duarte, I. (2001). Minimizando a Estrutura: uma Análise Unificada das Construções de Clivagem em Português [Minimizing the structure: a unified analysis of the cleft constrcutions in Portuguese.]. *Actas do XVI Encontro Nacional da Associação Portuguesa de Linguística*, 627-638, Lisbon: APL/Colibri.

Costa, J. & Duarte, I. (2005). *Cleft Strategies in Portuguese: A Unified Approach.* Paper presented at the 11th Colloquium on Generative Grammar, Zaragoza, Spain.

Costa, J. & Lobo, M. (2009). Estruturas clivadas: evidência dos dados do português europeu não standard [Cleft structures: Evidence from non standard European Portuguese]. In Dermeval da Hora (ed.), *Anais—6o Congresso Internacional da Abralin, Vol. 2*, 3800–3806. João Pessoa, Brazil: Ideia Editora.

Dékány, É. (2010). The English cleft-construction: It-clefts. *The Odd Yearbook*, 8, 39-78.

Dikken, M. D., Meinunger, A., & Wilder, C. (2000). Pseudoclefts and ellipsis. S*tudia linguistica*, 54(1), 41-89.

English Grammar Today - Cambridge Dictionary. (n.d.) Retrieved December 9, 2019, from https://dictionary.cambridge.org/pt/gramatica/gramatica-britanica/cleft-sentences-it-was-in-

june-we-got-married

Falk, Y., Lindqvist, C., & Bardel, C. (2015). The role of L1 explicit metalinguistic knowledge in L3 oral production at the initial state. B*ilingualism: Language and cognition*, 18(2), 227-235.

Flynn, S., Foley C. & Vinnitskaya, I. (2004*).* The Cumulative-Enhancement Model for Language Acquisition: Comparing Adults' and Children's Patterns of Development. *First, Second and Third Language Acquisition of Relative Clauses, International Journal of Multilingualism, 1:1,* 3-16. DOI: 10.1080/14790710408668175

Fox, D., Pesetsky, D. (2004). Cyclic linearization of syntactic structure. *Theoretical Linguistics 31 (1/2)*, 1–46.

Friedmann, N., Belletti, A. & Rizzi, L. (2009). Relativized relatives: Types of intervention in the acquisition of A-bar dependencies. *Lingua 119*, 67–88.

Gao, L. (2013). College English Curriculum Requirements in China: Expectations and Responses. *Problems of Education in the 21st Century, 51*, 47-62.

Gu, Q. (2018). The listen-to-write approach proposed for EFL teachers of college English in China: Definition and its essentials. *Journal of Language Teaching and Research*, *9*(2), 398-407.

Hashimoto, A. (1966). *Embedding structures in Mandarin /.* (Electronic Thesis or Dissertation). Retrieved from https://etd.ohiolink.edu/

Hedberg, N. (2000). The referential status of clefts. *Language*, 891-920.

Heggie, L. (1993). The range of null operators: Evidence from clefting. *Natural Language & Linguistic Theory*, 11(1), 45-84.

Hermas, A. (2014). Multilingual transfer: L1 morphosyntax in L3 English. I*nternational Journal of Language Studies*, 8(2).

Heycock, C., & Kroch, A. (1999). Pseudocleft connectedness: Implications for the LF interface level. *Linguistic inquiry*, *30*(3), 365-397.

Heycock, C., & Kroch, A. (2002). Topic, focus, and syntactic representations. In *Proceedings of WCCFL* (Vol. 21, pp. 101-125).

Higgins, F.R. (1979). *The Pseudo-Cleft Construction in English.* New York: Garland.

Hole, D. & Zimmermann, M. (2013). Cleft partitionings in Japanese, Burmese and Chinese. 10.1075/la.208.11hol.

Hole, D. (2011). The deconstruction of Chinese "shì… de" clefts revisited. *Lingua 121, 11,* 1707-1733. DOI: 10.1016/j.lingua.2011.07.004

Hu, G. (2005). English language education in China: Policies, progress, and problems. *Language policy*, *4*(1), 5-24.

Huang, C.-T.J. (1994). Verb movement and some syntax-semantics mismatches in Chinese. *Chinese Languages and Linguistics* 2:587-613, 1994.

Huang, C.-T.J. (1988). Shuo shi he you [On shi and you]. *Bulletin of the Institute of History and Philology 59*, part 1: 43-64.

Huang, C.-T.J. (1998). Logical Relations in Chinese and the Theory of Grammar, Garland Publishing.

Huang, C.-T.J., Li, Y., & Li, Y. (2009). *The Syntax of Chinese (Cambridge Syntax Guides).* Cambridge: Cambridge University Press. doi:10.1017/CBO9781139166935

Kayne, R. S. (1994). *The antisymmetry of syntax (Vol. 25)*. MIT press.

Kiss, K. É. (1998). Identificational Focus versus Information Focus. *Language, 74,* 245– 273. doi: 10.2307/417867

Kou, K. (submitted). The L2 Acquisition of European Portuguese Sluicing by L1 Mandarin Chinese Speakers. (Unpublished master's thesis). University of Lisbon, Lisbon, Portugal.

Lardiere, D. (2008). Feature assembly in second language acquisition. *The role of formal features in second language acquisition,* 106-140. doi: 10.4324/9781315085340-5

Lardiere, D. (2009). Some thoughts on a contrastive analysis of features in second language acquisition. *Second Language Research*, 25(2), 173–227.

Li. C., & Thompson, S. (1981). *Mandarin Chinese: A Functional Reference Grammar.* Berkeley, CA: University of California Press.

Lin, J-W. (2006). Time in a language without tense: the case of Chinese. *Journal of Semantics* 23: 1–53.

Lin, J-W. (2010). A Tenseless Analysis of Mandarin Chinese Revisited: A Response to Sybesma 2007. *Linguistic Inquiry - LINGUIST INQ. 41*. 10.2307/40606841.

Lobo, M. (2006). Assimetrias em construções de clivagem do português: movimento vs. geração na base. [Asymmetries in Portuguese cleft constructions: Movement vs. base generation]. In Fátima Oliveira & Joaquim Barbosa (Eds.), *XXI Encontro Nacional da APL. Textos Selecionados (pp. 457-473)*. Lisboa: APL/Colibri.

Lobo, M., Santos, A. L. & Soares, C. (2012). Aquisição de estruturas clivadas no português europeu: produção espontânea e induzida. [Acquisition of cleft structures in European Portuguese: Spontaneous and elicited production]. In A. Costa, C. Flores & N. Alexandre (Eds.), *Textos Selecionados*. *XXVII Encontro Nacional da Associação Portuguesa de Linguística*. Lisboa: APL.

Lobo, M., Santos, A. L. e Soares-Jesel, C. (2016). Syntactic structure and information structure: the acquisition of portuguese clefts and be-fragments. *Language Acquisition, [online] 23(2),* 142-174. doi: 10.1080/10489223.2015.1067317

Lobo, M., Santos, A. L., Soares-Jesel, C., & Vaz, S. (2019). Effects of syntactic structure on the comprehension of clefts. *Glossa: A Journal of General Linguistics, 4(1)*, 74. DOI: http://doi.org/10.5334/gjgl.645

Long, H-P. & Xiao, X-P. (2011). 'Wo shi zuotian mai de piao' jushi jiqi xiangguan wenti. *Shijie hanyu jiaoxue*. [Sentences like 'It was yesterday that I bought the tickets' and related questions, *Chinese Teaching in the World*] 25: 305-318.

Mai, Z., & Yuan, B. (2016). Uneven reassembly of tense, telicity and discourse features in L2 acquisition of the Chinese shì… de cleft construction by adult English speakers. *Second Language Research*, 32(2), 247-276.

Meinunger, A. (1998). A monoclausal structure for (pseudo-) cleft sentences. In *Proceedings of NELS* (Vol. 28, pp. 283-298).

Mendes, A., Antunes S., Janssen, M. & Gonçalves, A. (2016). The COPLE2 Corpus: A Learner Corpus for Portuguese. *Proceedings of the Tenth Language Resources and Evaluation Conference – LREC'16,* 3207-321, 23-28 May 2016, Portoroz, Slovenia.

Mioto, C. (2012). Reduced pseudoclefts in Caribbean Spanish and in Brazilian Portuguese. In V. Bianchi, & C. Chesi (Eds.), *Enjoy Linguistics! Papers offered to Luigi Rizzi on the occasion of his 60th birthday*, 287-302. Siena: CISCL Press.

Modesto, M. (1995). As construções clivadas no PB: relações entre interpretação focal, movimento sintáctico e prosódia. *MA dissertation*. Universidade de São Paulo.

Moro, A. (1997). *The raising of predicates. Predicative noun phrases and the theory of clause structure.* Cambridge: Cambridge University Press.

Na Ranong, S., & Leung, Y-k. I. (2009). Null objects in L1 Thai-L2 English-L3 Chinese: An empiricist take on a theoretical problem. In Y-k. I. Leung (Ed.), *Third language acquisition and Universal Grammar* (pp. 162–191). Clevedon: Multilingual Matters.

Nash, L., & Rouveret, A. (2002). Cliticization as unselective attract. *Catalan Journal of Linguistics*, 1, 157-199.

Paris, M.-C. (1979). Nominalization in Mandarin Chinese. Paris: Département de Recherches linguistiques, Université Paris 7.

Paris, M.-C. (1998). Focus operators and types of predication in Mandarin. *Cahiers de Linguistique Asie Orientale 27(2)*: 129-159.

Paul, W. (2005). Low IP area and left periphery in Mandarin Chinese. R*echerches linguistiques de Vincennes*, (33), 111-134.

Paul, W. (2009). Consistent disharmony: sentence-final particles in Chinese. Paper presented at the Workshop on particles, Dept. of Linguistics, organized by the research project Structure and linearization in disharmonic word orders, University of Cambridge, 30-31 October 2008.

Paul, W. (2015). New perspectives on Chinese syntax. Berlin: De Gruyter Mouton.

Paul, W., Whitman, J. (2008). Shi...de focus clefts in Mandarin Chinese. *Linguistic Review 25 (3/4)*, 413–451. DOI: 10.1515/TLIR.2008.012

Reeve, M. (2011). The syntactic structure of English clefts. *Lingua*, 121(2), 142-171.

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.

Rothman, J. (2010). On the typological economy of syntactic transfer: Word order and relative clause high/low attachment preference in L3 Brazilian Portuguese. *International Review of Applied Linguistics in Language Teaching*, 48(2-3), 245-273.

Rothman, J. (2011). L3 syntactic transfer selectivity and typological determinacy: The typological primacy model. *Second Language Research*, 27, 107–127. doi: 10.1177/0267658310386439

Rothman, J. (2015). Linguistic and cognitive motivations for the Typological Primacy Model (TPM) of third language (L3) transfer: Timing of acquisition and proficiency considered. *Bilingualism: language and cognition*, 18(2), 179-190.

Santos, A. L. (2006). Minimal answers: Ellipsis, syntax and discourse in the acquisition of European Portuguese. Lisbon, Portugal: University of Lisbon dissertation.

Schwartz, B. D. & Sprouse, R. A. (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. A. (1996). L2 cognitive states and the full transfer/full access model. *Second language research*, 12(1), 40-72.

Shao, H. (2006). An empirical study of washback from CET-4 on college English teaching and learning, in: *CELEA Journal,* 29 (1) (2006), pp. 54-65

Simpson, A. & Wu, Z. (2002). From D to T - Determiner Incorporation and the Creation of Tense. *Journal of East Asian Linguistics* 11: 169 - 209.

Slabakova, R. (2017). The scalpel model of third language acquisition. *International Journal of Bilingualism*, 21(6), 651-665.

Soares, C. (2006). La syntaxe de la périphérie gauche en portugais européen et son acquisition [The syntax of left periphery in European Portuguese and its acquisition]. Paris: University of Paris 8 dissertation.

Sybesma, R. (2007). Whether We Tense-Agree Overtly or Not. *Linguistic Inquiry - LINGUIST INQ. 38*. 580-587. 10.1162/ling.2007.38.3.580.

Tang, Ting-chi. (1980). Cleft and pseudo-cleft constructions in Chinese: structure, function and constraint. *Bulletin of National Taiwan Normal University* 25.249-296.

Tavares, J. A. (2005). *Orações copulativas nominais com o verbo* ser *em Português Europeu.* [Nominal copulative clauses with the verb ser "be" in European Portuguese]. Mater's thesis. Nova University of Lisbon, Lisbon, Portugal.

Teng, S-H. (1979). Remarks on Cleft sentences in Chinese. *Journal of Chinese Linguistics* 7, 1: 101-114.

Vercauteren, A. M. W. (2010). Como é que é com o é que? Análise de estruturas com é que em variedades não standard do português europeu [How about "é que"? Analysis of "é que" structures in non-standard varieties of European Portuguese]. Master's thesis. Nova University of Lisbon, Lisbon, Portugal.

Vercauteren, A. M. W. (2016). A conspiracy theory for clefts: The syntax and interpretation of cleft constructions. Doctorial dissertation. Nova University of Lisbon, Lisbon, Portugal.

White, L. (2003). *Second Language Acquisition and Universal Grammar* (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511815065

Zhan, F. & Sun, C. (2013). A copula analysis of shì in the Chinese cleft construction. *Language and Linguistics.* 14: 755-788.

Zhu, D. (1978). 'De' zi jiegou he panduan ju <"De" construction and clefts>. *Zhongguo Yuwen* 144:23-27 and 104-109.

Appendix I: Acceptability Judgement Task in European Portuguese

Tarefa: Juízo de gramaticalidade

Nome: _____ Data de Nascimento: _____ Código: _____

Instruções:

Leia cada contexto, julgue se a frase <u>abaixo e a negrito</u> é <u>gramaticalmente</u> correta, e classifique-a, usando uma escala em que

  0 = frase incorreta, completamente agramatical

  3 = frase perfeita, completamente gramatical

Caso não considere a frase perfeita, corrija a frase riscando ou escrevendo por cima. Se não souber mesmo responder, escolha a opção "não sei".

Por favor, concentre-se na estrutura de cada frase e ignore qualquer problema com ortografia ou pontuação. Marque apenas UMA resposta para cada frase, e assegure-se de que julgou todas as frases.

提示：

请阅读下列每一个背景句，判断背景句下方加粗的句子是否符合语法，并运用以下梯度打分，圈出所给分数：

0 = 错误句，完全不符合语法

3 = 正确句，完全符合语法

如果不认为该句完全正确，请在原句上修改。如果实在无法判断，请选择"não sei"。

请专注于句子结构，并忽略任何拼写或标点问题。请为每一题圈出一个答案，并最后确认已经完成了所有题目。

Exemplo 例子:

CONTEXTO: *Eu não comi o bolo.*

      comeu

                                      Ⓞ 1   2   3   não sei

O   João ~~comi~~ o bolo.

1. CONTEXTO: *O que disse a Amália?*

    A Amália disse que se levantou cedo.   0   1   2   3   não sei

2. CONTEXTO: *A Ana comeu um bife.*

    O Pedro não sabe a Ana comeu o bife onde.   0   1   2   3   não sei

3. CONTEXTO: *O Pablo não destruiu a torneira.*

    Quem destruiu a torneira foi a Letícia.   0   1   2   3   não sei

4. CONTEXTO: *A Tatiana ainda não visitou o Porto.*

    A Eduarda já visitou-o.   0   1   2   3   não sei

5. CONTEXTO: *O Simão não atendeu o telefone.*

    A Rute é que atendeu o telefone.   0   1   2   3   não sei

6. CONTEXTO: *O Afonso acabou a aula.*

    A Diana não sabe quando é que o Afonso acabou a aula.   0   1   2   3   não sei

7. CONTEXTO: *A Filipa leu a revista.*

    O Álvaro também leu-a.   0   1   2   3   não sei

8. CONTEXTO: *O Carlos bebeu leite.*

    Quando é que o Carlos bebeu o leite?   0   1   2   3   não sei

9. CONTEXTO: *O rapaz não corrigiu o texto.*

    As raparigas foram que corrigiram o texto.   0   1   2   3   não sei

10. CONTEXTO: *A Isabel leu um livro.*

    O João não sabe a Isabel leu o livro quando.   0   1   2   3   não sei

11. CONTEXTO: *A Cecília não ganhou a bolsa.*

    Foi a Carlota ganhou a bolsa.   0   1   2   3   não sei

12. CONTEXTO: *A Ana informou alguém?*

    A Ana informou alguém, mas a Sofia não sabe quem.   0   1   2   3   não sei

13. CONTEXTO: *A Diana convidou alguém.*

    Quem é que a Diana convidou?   0   1   2   3   não sei

14. CONTEXTO: *O Francisco não lavou a cozinha.*

Os pais é que lavaram a cozinha.   0   1   2   3   não sei

15. CONTEXTO: *A Diana encontrou alguém?*

A Diana encontrou alguém, mas a Ana não sabe é quem.   0   1   2   3   não sei

16. CONTEXTO: *A Frederica não vendeu os óculos.*

O que a Frederica vendeu foi o casaco.   0   1   2   3   não sei

17. CONTEXTO: *O Diogo fez alguma coisa?*

O Diogo fez alguma coisa, mas a Maria não sabe o quê.   0   1   2   3   não sei

18. CONTEXTO: *A Joana estudou alguma coisa.*

A Maria não sabe a Joana estudou o quê.   0   1   2   3   não sei

19. CONTEXTO: CONTEXTO: *O Prof. João não chumbou a turma toda.*

A Profª. Lúcia é que chumbou a turma toda.   0   1   2   3   não sei

20. CONTEXTO: *A Ana lavou o carro?*

A Ana lavou o carro em algum lugar, mas o Carlos não sabe é onde.   0   1   2   3   não sei

21. CONTEXTO: *Os rapazes não abriram a porta.*

Quem abriram a porta foram os avós.   0   1   2   3   não sei

22. CONTEXTO: *A Susana comprou alguma coisa.*

A Maria não sabe o que é que a Susana comprou.   0   1   2   3   não sei

23. CONTEXTO: *O Fernando não mostrou um filme.*

Foi uma série o Fernando mostrou.   0   1   2   3   não sei

24. CONTEXTO: *A Maria passeou o cão?*

A Maria passeou o cão em algum lugar, mas o Bruno não sabe onde.   0   1   2   3   não sei

25. CONTEXTO: *O Cesário não perdeu a carteira.*

A mala é que o Cesário perdeu.   0   1   2   3   não sei

26. CONTEXTO: *O Vítor não lavou a roupa.*

A Luciana já a lavou.   0   1   2   3   não sei

27. CONTEXTO: *A Dra. Ferreira não interrompeu o diálogo.*

Foi o Dr. Machado que interrompeu o diálogo. 0 1 2 3 não sei

28. CONTEXTO: *A Madalena entregou o trabalho de casa?*

Acho que a Madalena entregou-o. 0 1 2 3 não sei

29. CONTEXTO: *O Pedro leu um livro.*

Quando é que o Pedro leu o livro? 0 1 2 3 não sei

30. CONTEXTO: *A Vera não sujou a roupa.*

O Vítor é que sujou a roupa. 0 1 2 3 não sei

31. CONTEXTO: *O Carlos arrumou o quarto.*

A Maria não sabe o Carlos arrumou o quarto quando. 0 1 2 3 não sei

32. CONTEXTO: *A Emília não entrou na Sala de Atos.*

Entrou foi a Mafalda. 0 1 2 3 não sei

33. CONTEXTO: *Quem ofereceu as flores aos professores?*

Todos ofereceram-lhes as flores. 0 1 2 3 não sei

34. CONTEXTO: *O Afonso arrumou o quarto?*

O Afonso arrumou o quarto num desses dias, mas o Carlos não sabe quando. 0 1 2 3 não sei

35. CONTEXTO: *A Madalena não publicou o livro.*

Foi o Eduardo publicou o livro. 0 1 2 3 não sei

36. CONTEXTO: *A Maria comprou um livro.*

Onde é que a Maria comprou o livro? 0 1 2 3 não sei

37. CONTEXTO: *O Rúben não escreveu uma carta.*

O que o Rúben escreveu foi um romance. 0 1 2 3 não sei

38. CONTEXTO: *A Sofia viu alguém?*

A Sofia viu alguém, mas o Paulo não sabe quem. 0 1 2 3 não sei

39. CONTEXTO: *Os meninos não resolveram a equação.*

Quem resolveu a equação foram as meninas. 0 1 2 3 não sei

40. CONTEXTO: *Tenho uma nova notificação.*

Alguém enviou-me um e-mail. 0 1 2 3 não sei

41. CONTEXTO: *A Glória leu alguma coisa.*

A Glória leu o quê? 0 1 2 3 não sei

42. CONTEXTO: *O João não recebeu o prémio.*

Foram as meninas que recebeu o prémio. 0 1 2 3 não sei

43. CONTEXTO: *A Sofia tirou uma fotografia.*

A Joana não sabe onde é que a Sofia tirou a fotografia. 0 1 2 3 não sei

44. CONTEXTO: *O dono não fechou a loja.*

Quem fechou a loja foram os filhos. 0 1 2 3 não sei

45. CONTEXTO: *A Júlia ouviu alguma coisa.*

A Júlia ouviu o quê? 0 1 2 3 não sei

46. CONTEXTO: *Estes atores não deram entrevistas.*

Foram as cantoras que deram entrevistas. 0 1 2 3 não sei

47. CONTEXTO: *O Nelson usou o computador.*

A Matilde também o usou. 0 1 2 3 não sei

48. CONTEXTO: *O gerente não estragou o piano.*

Os músicos é que estragaram o piano. 0 1 2 3 não sei

49. CONTEXTO: *O Luís visitou alguém.*

O Carlos não sabe quem é que o Luís visitou. 0 1 2 3 não sei

50. CONTEXTO: *A Rafaela não cancelou a reunião.*

Foi o Santiago que cancelou a reunião. 0 1 2 3 não sei

51. CONTEXTO: *A Margarida vendeu o apartamento.*

O João não sabe quando é que a Margarida vendeu o apartamento. 0 1 2 3 não sei

52. CONTEXTO: *O Guilherme não aqueceu a sopa.*

    Quem aqueceu a sopa foi a Teresa.   0   1   2   3   não sei

53. CONTEXTO: *A Susana quebrou o vidro.*

    Não, a Susana não quebrou-o.   0   1   2   3   não sei

54. CONTEXTO: *A Joana não tomou chá.*

    Foi café a Joana tomou.   0   1   2   3   não sei

55. CONTEXTO: *O Roberto cumprimentou alguém.*

    O Roberto cumprimentou quem?   0   1   2   3   não sei

56. CONTEXTO: *A Diana limpou a mesa?*

    A Diana limpou a mesa num desses dias, mas a Joana não sabe quando.   0   1   2   3   não sei

57. CONTEXTO: *A Anita não preparou sobremesas.*

    Os irmãos é que prepararam sobremesas.   0   1   2   3   não sei

58. CONTEXTO: *A Inês comprou alguma coisa.*

    O que é que a Inês comprou?   0   1   2   3   não sei

59. CONTEXTO: *A Sónia não estudou um romance realista.*

    Foi um poema modernista que a Sónia estudou.   0   1   2   3   não sei

60. CONTEXTO: *A Isabel levantou o dinheiro.*

    A Isabel levantou o dinheiro onde?   0   1   2   3   não sei

61. CONTEXTO: *O Diogo não fez o teste.*

    Foram os amigos que fizeram o teste.   0   1   2   3   não sei

62. CONTEXTO: *A Íris faz anos hoje.*

    Todos deram-lhe os parabéns.   0   1   2   3   não sei

63. CONTEXTO: *A Antónia não visitou o museu.*

    O castelo é que a Antónia visitou.   0   1   2   3   não sei

64. CONTEXTO: *O Bruno não apareceu na cerimónia.*

    Apareceu foi a Margarida.   0   1   2   3   não sei

65. CONTEXTO: *A Luísa ficou assustada.*

    Alguém empurrou-a.   0   1   2   3   não sei

66. CONTEXTO: *O Bruno ouviu alguma coisa.*

    O Diogo não sabe o que é que o Bruno ouviu.   0   1   2   3   não sei

67. CONTEXTO: *A Fabiana não aprendeu francês.*

    Quem aprendeu francês foi o Alberto.   0   1   2   3   não sei

68. CONTEXTO: *O Adriano tomou o medicamento.*

    O Adriano tomou o medicamento quando?   0   1   2   3   não sei

69. CONTEXTO: *A Laura encontrou a Sandra.*

    O Miguel também encontrou-a.   0   1   2   3   não sei

70. CONTEXTO: *A Ana perdeu a carteira.*

    Onde é que a Ana perdeu a carteira?   0   1   2   3   não sei

71. CONTEXTO: *A Lorena não comprou um quadro.*

    A Lorena comprou foi um CD.   0   1   2   3   não sei

72. CONTEXTO: *O Henrique não cometeu esse erro.*

    Foi o Dinis que cometeu o erro.   0   1   2   3   não sei

73. CONTEXTO: *A Elisa comeu alguma coisa?*

    A Elisa comeu alguma coisa, mas o Afonso não sabe é o quê.   0   1   2   3   não sei

74. CONTEXTO: *A Rosa não elaborou esta teoria.*

    Quem elaboraram esta teoria foram as amigas.   0   1   2   3   não sei

75. CONTEXTO: *A Maria encontrou um gato.*

    O Pedro não sabe onde é que a Maria encontrou o gato.   0   1   2   3   não sei

76. CONTEXTO: *A Noa não desligou a máquina de lavar roupa.*

    A televisão é que a Noa desligou.   0   1   2   3   não sei

77. CONTEXTO: *A Margarida fez uma festa.*

    O Pablo não sabe a Margarida fez a festa onde.   0   1   2   3   não sei

78. CONTEXTO: *Os enfermeiros não divulgaram a notícia.*

    Os médicos foram que divulgaram a notícia.   0   1   2   3   não sei

79. CONTEXTO: *A Beatriz matou alguém.*

    Quem é que a Beatriz matou?   0   1   2   3   não sei

80. CONTEXTO: *A empregada não limpou as mesas.*

    Foi o chão a empregada limpou.   0   1   2   3   não sei

81. CONTEXTO: *A Joana viu alguém.*

    A Joana viu quem?   0   1   2   3   não sei

82. CONTEXTO: *A Nina não leu o poema.*

    Foram os colegas que leram o poema.   0   1   2   3   não sei

83. CONTEXTO: *A Beatriz abriu a janela?*

    A Beatriz abriu a janela num desses dias, mas a Isabel não sabe é quando.   0   1   2   3   não sei

84. CONTEXTO: *O Samuel não congelou a carne.*

    Foi o peixe que o Samuel congelou.   0   1   2   3   não sei

85. CONTEXTO: *O Bernardo não guardou os documentos.*

    Quem guardaram os documentos foram os funcionários.   0   1   2   3   não sei

86. CONTEXTO: *O Leonardo resolveu o problema.*

    Alguém o ajudou.   0   1   2   3   não sei

87. CONTEXTO: *A Liliana não chegou antes da aula.*

    Chegou foi a Teresa.   0   1   2   3   não sei

88. CONTEXTO: *A Maria bebeu alguma coisa?*

    A Maria bebeu alguma coisa, mas o João não sabe o quê.   0   1   2   3   não sei

89. CONTEXTO: *O Sérgio vendeu o livro.*

    Não, ele não o vendeu.   0   1   2   3   não sei

90. CONTEXTO: *Os caloiros não acabaram o trabalho.*

    Os finalistas foram que acabaram o trabalho.   0   1   2   3   não sei

91. CONTEXTO: *O Danilo comeu uma sopa.*

    O Danilo comeu a sopa quando?   0   1   2   3   não sei

92. CONTEXTO: *Os engenheiros não sugeriram este plano.*

    Foram os arquitetos que sugeriu este plano.   0   1   2   3   não sei

93. CONTEXTO: *A Joana vendeu o carro?*

    A Joana vendeu o carro num desses dias, mas o David não sabe é quando.   0   1   2   3   não sei

94. CONTEXTO: *Onde está o chocolate?*

    O Ricardo disse que comeu-o.   0   1   2   3   não sei

95. CONTEXTO: *O Pedro beijou alguém.*

    A Ana sabe o Pedro beijou quem.   0   1   2   3   não sei

96. CONTEXTO: *O Bruno tomou o café?*

    O Bruno tomou o café em algum lugar, mas a Beatriz não sabe é onde.   0   1   2   3   não sei

97. CONTEXTO: *A Cristina não vestiu os calções.*

    A Cristina vestiu foi a saia.   0   1   2   3   não sei

98. CONTEXTO: *A Nádia viu um filme.*

    A Nádia viu o filme onde?   0   1   2   3   não sei

99. CONTEXTO: *O Romeu beijou a Julieta.*

    Não, o Romeu não beijou-a.   0   1   2   3   não sei

100. CONTEXTO: *O Pedro bebeu alguma coisa.*

    A Susana não sabe o Pedro bebeu o quê.   0   1   2   3   não sei

101. CONTEXTO: *O Gilberto não partiu a janela.*

    Foi a Margarida partiu a janela.   0   1   2   3   não sei

102. CONTEXTO: *A Joana partiu o prato?*

    A Joana partiu o prato em algum lugar, mas o Afonso não sabe onde.   0   1   2   3   não sei

103. CONTEXTO: *O Enzo não ouviu esta música.*

    O Gustavo já ouviu-a.   0   1   2   3   não sei

104. CONTEXTO: *Os residentes não construíram o edifício.*

    Quem construiu o edifício foram os trabalhadores.   0   1   2   3   não sei

105. CONTEXTO: *O Carlos visitou alguém?*

    O Carlos visitou alguém, mas a Margarida não sabe é quem.   0   1   2   3   não sei

106. CONTEXTO: *A Inês não encomendou a tarte.*

    Foram os colegas que encomendou a tarte.   0   1   2   3   não sei

107. CONTEXTO: *O Miguel procurou alguém.*

    A Isabel sabe quem é que o Miguel procurou.   0   1   2   3   não sei

108. CONTEXTO: *O Duarte não alugou uma casa.*

    O que o Duarte alugou foi um quarto.   0   1   2   3   não sei

109. CONTEXTO: *Hoje é o Dia de Natal.*

    Todos o festejam em casa.   0   1   2   3   não sei

110. CONTEXTO: *A Sofia escreveu alguma coisa.*

    O que é que a Sofia escreveu?   0   1   2   3   não sei

111. CONTEXTO: *O André não comprou um computador.*

    Foi um tablet que o André comprou.   0   1   2   3   não sei

112. CONTEXTO: *A Beatriz escreveu alguma coisa?*

    A Beatriz escreveu alguma coisa, mas o Paulo não sabe é o quê.   0   1   2   3   não sei

113. CONTEXTO: *O Xavier não cantou uma música de pop.*

    O Xavier cantou foi um fado.   0   1   2   3   não sei

114. CONTEXTO: *O Diogo convidou alguém.*

O Paulo não sabe o Diogo convidou quem.   0   1   2   3   não sei

Appendix II: Acceptability Judgement Task in English

Nome: _____ Data de Nascimento: _____ Código: _____

提示：

请阅读下列每一个背景句，判断背景句下方加粗的句子是否符合语法，并运用以下梯度打分，圈出所给分数：

0 = 错误句，完全不符合语法

3 = 正确句，完全符合语法

如果不认为该句完全正确，请在原句上修改。如果实在无法判断，请选择"I do not know"。

请专注于句子结构，并忽略任何拼写或标点问题。请为每一题圈出一个答案，并最后确认已经完成了所有题目。

1. CONTEXT: *Who has solved the problem?*

   The problem had been solved by Michelle.　　　　0　1　2　3　I do not know

2. CONTEXT: *Sally lost her wallet.*

   Sam knows Sally lost her wallet where.　　　　0　1　2　3　I do not know

3. CONTEXT: *Ivonne didn't answer the door.*

   It was Joanne that answered the door.　　　　0　1　2　3　I do not know

4. CONTEXT: *Who cleans the table every day?*

   The table was cleaned by John every day.　　　　0　1　2　3　I do not know

5. CONTEXT: *Who had taken a photo?*

   A photo had been taken by Fiona.　　　　0　1　2　3　I do not know

6. CONTEXT: *Did Sam lose something?*

   Sam lost something, but Peter doesn't know what.　　　　0　1　2　3　I do not know

7. CONTEXT: *Laura didn't read the novel.*

   What Laura read was a film script.　　　　0　1　2　3　I do not know

8.   CONTEXT: *Who kicked the ball?*

    The ball is kicked by the children.   0 1 2 3   I do not know

9.   CONTEXT: *Daniel didn't accept the offer.*

    It was Ronan accepted the offer.   0 1 2 3   I do not know

10. CONTEXT: *Cindy ate an apple.*

    Jenny knows when Cindy ate an apple.   0 1 2 3   I do not know

11. CONTEXT: *Who eats a banana every day?*

    A banana was eaten by Doris every day.   0 1 2 3   I do not know

12. CONTEXT: *Did Sam have an accident?*

    Sam had an accident somewhere, but
    Peter doesn't know where.   0 1 2 3   I do not know

13. CONTEXT: *Raymond didn't learn German.*

    It was Dutch Raymond learnt.   0 1 2 3   I do not know

14. CONTEXT: *Who was writing an e-mail?*

    An e-mail is being written by Robert.   0 1 2 3   I do not know

15. CONTEXT: *Did Susan hit someone?*

    Susan hit someone, but Peter doesn't
    know is who.   0 1 2 3   I do not know

16. CONTEXT: *The teachers didn't organize the parade.*

    It was the students that organized the parade.   0 1 2 3   I do not know

17. CONTEXT: *Who catches a fish every day?*

    A fish is caught by Paul every day.   0 1 2 3   I do not know

18. CONTEXT: *Did Sam kiss someone?*

    Sam kissed someone, but Peter doesn't
    know who.   0 1 2 3   I do not know

    CONTEXT: *Who beat Susana?*

    Susana is beaten by Tom.   0 1 2 3   I do not know

19. CONTEXT: *David didn't drink the milk.*

    Who drank the milk was Paul.   0   1   2   3   I do not know

20. CONTEXT: *Jenny bought something.*

    Sandy knows Jenny bought what.   0   1   2   3   I do not know

21. CONTEXT: *Who has received a letter?*

    A letter has been received by Sophia.   0   1   2   3   I do not know

22. CONTEXT: *Levin didn't drop his key.*

    It was Peter that dropped his key.   0   1   2   3   I do not know

23. CONTEXT: *Who is selling Tom's car?*

    Tom's car was being sold by Clare.   0   1   2   3   I do not know

24. CONTEXT: *Did Joanne buy something?*

    **Joanne bought something, but Becky doesn't
    know is what.**   0   1   2   3   I do not know

25. CONTEXT: *Cristine didn't write a post card.*

    It was a letter that Cristine wrote.   0   1   2   3   I do not know

26. CONTEXT: *David met his friend.*

    Susana knows where David met his friend.   0   1   2   3   I do not know

27. CONTEXT: *Who was building a house?*

    The house is being built by Harry.   0   1   2   3   I do not know

28. CONTEXT: *John didn't borrow a dictionary.*

    It was a book John borrowed.   0   1   2   3   I do not know

29. CONTEXT: *Did Peter hold a party?*

    **Peter held a party somewhere, but Cindy
    doesn't know is where.**   0   1   2   3   I do not know

30. CONTEXT: *Caroline didn't draw the picture.*

    Who drew the picture was Anita.   0   1   2   3   I do not know

31. CONTEXT: *Who bought a t-shirt?*

   A t-shirt was bought by Mary.   0 1 2 3   I do not know

32. CONTEXT: *Tom sold his car.*

   Peter knows Tom sold his car when.   0 1 2 3   I do not know

33. CONTEXT: *The singers didn't practice the musical.*

   It was the dancers that practiced the musical.   0 1 2 3   I do not know

34. CONTEXT: *Who had fed the cat?*
   The cat has been fed by Jenny.   0 1 2 3   I do not know

35. CONTEXT: *Adam didn't pick an apple.*

   What Adam picked was a pear.   0 1 2 3   I do not know

36. CONTEXT: *Peter saw someone.*

   Sally knows Peter saw who.   0 1 2 3   I do not know

37. CONTEXT: *Who has caught the mouse?*

   The mouse had been caught by Sarah.   0 1 2 3   I do not know

38. CONTEXT: *The company didn't sign the contract.*

   It were the lawyers that signed the contract.   0 1 2 3   I do not know

39. CONTEXT: *Sam ate something.*

   Peter knows what Sam ate.   0 1 2 3   I do not know

40. CONTEXT: *Who was washing the dishes?*

   The dishes were being washed by Sandy.   0 1 2 3   I do not know

41. CONTEXT: *Teddy didn't break the window.*

   It was Fredric broke the window.   0 1 2 3   I do not know

42. CONTEXT: *Ivan visited someone.*

   Susan knows who Ivan visited.   0 1 2 3   I do not know

43. CONTEXT: *Philip didn't buy the beer.*

   It was the juice that Philip bought.   0 1 2 3   I do not know

158

44. CONTEXT: *Did Robert call Jenny?*

   **Robert called Jenny one day last week,
   but Jenny doesn't know is when.**   0 1 2 3   I do not know

45. CONTEXT: *Who is writing a story?*

   The story is being written by Susana.   0 1 2 3   I do not know

46. CONTEXT: *Who had made a cake?*

   A cake has been made by Kylie.   0 1 2 3   I do not know

47. *The police didn't search the whole town.*

   **It were the volunteers that searched
   the whole town.**   0 1 2 3   I do not know

48. CONTEXT: *Did Sam visit the museum?*

   **Sam visited the museum one day last week,
   but Peter doesn't know when.**   0 1 2 3   I do not know

49. CONTEXT: *Who is reading a book?*
   The book was being read by Peter.   0 1 2 3   I do not know

Appendix III: Consent

CONSENTIMENTO INFORMADO

"Aquisição do português europeu como língua não materna por falantes de mandarim e aquisição do mandarim como língua não materna por falantes de português"

Objectivo do estudo: Este estudo pretende identificar possíveis problemas na aquisição do português europeu como língua não materna por falantes de mandarim e possíveis problemas na aquisição do mandarim por falantes de português.

Descrição e métodos: Para alcançar os objetivos definidos, será necessário comparar o desempenho em português de alunos chineses que têm mandarim como língua materna com o desempenho de falantes monolingues do português; será ainda necessário comparar o desempenho em mandarim de falantes que têm o português como língua materna com o desempenho de falantes monolingues do mandarim. O estudo foca-se apenas na aquisição em idade adulta. A participação neste estudo implica que o participante responda a diferentes questionários, que envolvem o julgamento da gramaticalidade de frases, a avaliação do seu valor de verdade e, ainda, a produção escrita (produção provocada). A recolha de dados será feita em suporte escrito. A participação de cada participante no estudo demorará em conjunto cerca de 60 minutos. Os resultados expectáveis poderão contribuir para uma compreensão mais completa da aquisição de uma língua não materna, podendo ter, especificamente, consequências para a reflexão sobre o ensino de português a falantes de mandarim e para o ensino de mandarim a falantes de português.

Riscos previsíveis: Não aplicável.

Possíveis benefícios para os participantes: Não se garante que este estudo proporcione benefícios diretos para o participante. No entanto, a informação obtida vai contribuir para aumentar o conhecimento científico sobre a aquisição da sintaxe da língua não materna e poderá vir a beneficiar terceiros. O participante não terá quaisquer benefícios financeiros decorrentes deste estudo.

Participação voluntária: O participante terá toda a liberdade para recusar a participação no estudo ou retirar o seu consentimento, suspendendo a participação em qualquer momento. A participação é voluntária e a recusa em participar não acarreta qualquer penalização ou perda de benefícios.

Confidencialidade: Os dados obtidos serão utilizados exclusivamente para investigação. A informação recolhida de cada participante será combinada e analisada em conjunto com informação de outros participantes. Todos os dados de identificação de cada participante serão mantidos em confidencialidade. Para o estudo, a cada participante será atribuído um número codificado. A identidade dos participantes nunca será revelada em qualquer relatório ou publicação decorrente do estudo.

A quem devo colocar questões relacionadas com este estudo: Chang Gao (mestranda, FLUL), 18811760223@163.com; JinwenYu (mestrando, FLUL), isyujw@163.com; Ka Man Kou (mestranda, FLUL), joanne1995kou@gmail.com; Xinyi Li (mestranda, FLUL), lixinyi@campus.ul.pt; Ana Lúcia Santos (Professora Associada; Faculdade de Letras, Universidade de Lisboa), als@letras.ulisboa.pt (Investigadora Responsável).

Appendix IV: Corrections made in Acceptability Judgement Task in European Portuguese

grupo control	sujeito								
É QUE	canónica	pseudo	invertida	Q	frase simples	semipseudo	é que	sem C	outros
1									
2			1		1				
3			1						
total			2		1				
CANÓNICA				Q					
4				4					
5				5					
6				3					
total				12					
sem QUE	canónica			Q					
7	13			8	1				
8	12			7					
9	13			6					
total	38			21	1				
SEMIPSEUDO		pseudo							
10	2	11					1		
11		12	1				1		
12	2	12	1						
total	4	35	2				2		
PSEUDO									
13									
14									
15									
total									

objecto										
grupo	É QUE	canónica	pseudo	invertida	Q	frase simples	semipseudo	é que	sem C	outros
control	16	1	4	1		9	1			2
	17	1	5	3		5		1 (subj.)		2
	18	2	2	1		10	2			1
		4	11	5		24	3	1		5
	CANÓNICA		pseudo	invertida	Q	frase simples	semipseudo			
	19		2		1	1				
	20		3		4	1	1			
	21		1	1	4		1			
			6	1	9	2	2			
	sem QUE	canónica	pseudo	invertida	Q	frase simples	semipseudo			
	22	12		1	4	1	1			1
	23	10	1	2	6					1
	24	7	2	1		6	2			1
		29	3	3	11	7	3			3
	SEMIPSEUDO		pseudo			frase simples				
	25		4			3				1
	26		5			4				1
	27		5			3				
			14			10				2
	PSEUDO									
	28									1(aquilo que)
	29									
	30									
										1

B1		canónica	pseudo	invertida	Q	frase simples	semipseudo	é que	sem C	Foi que
					sujeito					
B1	É QUE	canónica	pseudo	invertida	Q	frase simples	semipseudo	é que	sem C	Foi que
	1	4		3	1	6				1
	2	2		2		8				
	3	2		3		8			1	1
	total	8		8	1	22			1	2
	CANÓNICA				Q				sem C	
	4				1				1	
	5				2				2	
	6				1					
	total				4				3	
	sem QUE	canónica			Q	frase simples				Foi...fazER...
	7	12			1	1				1
	8	11			6					1
	9	9			3					
	total	32			10	1				2
	SEMIPSEUDO	canónica	pseudo				chegou SER...		sem C	chegou, foi...
	10	2	9		1	1	1		3	1
	11	4	8		1	2	2		3	1
	12	6	7		1	1	1		4	1
	total	12	24		3	4	4		10	3
	PSEUDO	canónica			Q		quem...SER			WH..., foi...
	13	4			2	1	1		1	
	14	1			3					2
	15	1			2					2
	total	6			7	1	1		1	4

		canónica	pseudo	invertida	Q	frase simples	semipseudo	outros	sem C	Foi que
						objecto				
B1	É QUE	canónica	pseudo	invertida	Q	frase simples	semipseudo	outros	sem C	Foi que
	16	2		2		1			3	1
	17	5				4		1	1	1
	18	2		1		7		1	1	1
		9		3		12		2	5	3
	CANÓNICA								sem C	
	19									
	20								1	
	21								1	
									2	
	sem QUE	canónica			Q		sem QUE (+,/ invertida)			Foi…é que
	22	10	1		1		2			1
	23	13			1		1			
	24	13			1		1			
		36	1		3		4			1
	SEMIPSEUDO	canónica	(incluindo 3: o que é que…)			frase simples	outros		sem QUE	
	25	3	12			6	1			
	26	2	12			5	1		1	
	27	3	12			5	1		1	
		8	36			16	3		2	
	PSEUDO						outros(+,)			
	28	1			3		1			
	29				1		1			
	30	1			1	…que foi				
		2			5		3			

objecto									
B2 É QUE	canónica	pseudo	invertida	Q	frase simples	semipseudo	é que	m C (-qu	Foi que
1	7	1	2		2				
2	3	1	1		5			1	
3	6		1		3			1	1
Total	16	2	4		10			2	1
CANÓNICA				Q					
4				1					
5				1					
6				1					
Total				3					
SEM QUE	canónica			Q					
7	12			2					
8	15			1					
9	17			1					
Total	44			4					

SEMIPSEUDO	canónica	Pseudo	invertida	frase simples	sem C
10	4	13		1	2
11	5	13			
12	3	10	2	1	3
Total	12	36	2	2	5

PSEUDO	outros
13	
14	
15	quem…, foi
Total	1

						objecto				
B2	É QUE	canónica	pseudo	invertida	Q	frase simples	semipseudo	foi…é que	sem C	Foi que
	16	7	2	3					2	
	17	8	1	2						
	18	6	1	1		2		1	1	1
total		21	4	6		2		1	3	1
	CANÓNICA									
	19									
	20									
	21									
total										
	SEM QUE	canónica								
	22	18	1							
	23	18								
	24	16								
total		52	1							
	SEMIPSEUDO	canónica	pseudo			frase simples				
	25	1	16			3				
	26	2	17			1				
	27	4	12			2				
total		7	45			6				
	PSEUDO									
	28									
	29						o que é que foi…			
	30									
total								1		

		sujeito							
C1	É QUE	canónica	pseudo	invertida	Q	frase simples	semipseudo	é que	sem C
	1	3				1			
	2	3				1			
	3	3				1			
	Total	9				3			
	CANÓNICA								
	4								
	5								
	6								
	Toal								
	sem QUE	canónica			Q				Foi é que
	7	14			1				1
	8	15							
	9	13							1
	Total	42			1				2
	SEMIPSEUDO	canónica	pseudo		Q				sem C
	10	6	8						
	11	4	9		1				
	12	3	9		1				1
	Total	13	26		2				1
	PSEUDO				Q				
	13								
	14				1				
	15				1				
	Total				2				

							sujeito				
C1	É QUE	canónica	pseudo	invertida	Q	frase simples	semipseudo	é que	sem C	Foi... é que	
	16	6		1						2	
	17	6		1						1	
	18	6		1						2	
	TOTAL	18		3						5	
	CANÓNICA										
	19										
	20										
	21										
	TOTAL										
	sem QUE	canónica								Foi é que	
	22	14								1	
	23	12								1	
	24	15								1	
	TOTAL	41								3	
	SEMIPSEUDO	canónica	pseudo							que..vestiu foi	
	25	3	13			1					
	26	2	13			1					
	27	2	12			1				1	
	TOTAL	7	38			3				1	
	PSEUDO										
	28										
	29										
	30										
	TOTAL										

grupo control

É QUE	canónica	frase simples	pseudo	invertida	invert. Concord.	Outros	Q. Concord.
31						5	
32						1	
33						1	
total						7	

É QUE flex.	canónica	frase simples	é que	invertida	invert. Concord.	Outros	Q. Concord.
34	9	1	5	1	2	3	
35	5		7	3	2	2	2
36	6		6	4	2	3	
total	20	1	18	8	6	8	2

CANÓNICA	canónica sem concord	Outros	Q. Concord.
37		1	1
38			2
39		1	4
total		2	7

CANÓNICA	canónica com concord	Q.	Q. Concord.	Outros	
40	15	1	1	3	1
41	16	2		1	
42	17			2	
total	48	3	1	6	1

PSEUDO	pseudo com concord
43	
44	
45	
total	

PSEUDO COM CONCORD	pseudo sem concord	Q.
46	20	1
47	16	1
48	21	
Total	57	2

grupo B1

É QUE	canónica	frase simples	pseudo	sem C	invert. Concord	Q
31	6	9				1
32	5	8		1		1
33	5	10		1		1
total	16	27		2		3

É QUE flex.	canónica	frase simples	Q. Concord	sem C	invert. Concord	Outros
34	5	7		1	2	3
35	7	2	1	1	2	
36	7	1	1	1	2	1
total	19	10	2	3	6	4

CANÓNICA	canónica sem concord	Q. Concord	sem C	invert. Concord
37		1	1	1
38		1		
39		2		
total		4	1	1

CANÓNICA SI	canónica com concord	Q. Concord	sem C		Q
40	6	1	1		2
41	11				2
42	7	1	2		3
total	24	2	3		7

PSEUDO	canónica	cano -concord	pseudo +con	sem C	Outros	Q
43		1		1	quem..SG foi 1	1
44	1	1	2		1	1
45	1	1	1		2	2
total	2	3	3	1	4	4

PSEUDO+Con	canónica	Q. concord.	pseudo - con	sem C	outros
46	1	2	5		2
47	1	3	4		1
48		3	1(foi)	1	3
Total	2	8	10	1	6

grupo B2	É QUE	canónica	frase simples	pseudo	é que flex.	Invertida+con	Q
	31	7	5		1	1	
	32	6	4		1	1	
	33	6	6		1		
	total	19	15		3	2	

	É QUE flex.	canónica	frase simples	foram QUEM que	é que flex.	Invertida+con	é que
	34	11	2	1	1	1	1
	35	14			1		
	36	13					
	total	38	2	1	2	1	1

	CANÓNICA	canó. - con.	Q.+ Con.
	37		
	38		1
	39		1
	total		2

	CANÓNICA - CON	canó + con.
	40	13
	41	15
	42	11
	total	39

	PSEUDO	pseudo+ concord	quem...FOI
	43	4	2
	44	2	1
	45	3	1
	total	9	4

PSEUDO COM CONCORD		pseudo- concord
	46	3
	47	2
	48	1
Total		6

É QUE	canónica	frase simples	pseudo	invertida	invert.con	Q
31	3	1				
32	3	1				
33	3	1				
total	9	3				

É QUE flexiona	canónica	FOI ...que (canó.)	é que	foram aqueles que	invert.con	Q
34	12	1	2			
35	9	1	3	1	1	
36	10	1	2	1		
total	31	3	7	2	1	

CANÓNICA	canónica - con.	FOI ...que (canó.)
37		1
38		1
39		1
total		3

CANÓNICA - CO	canónica + con.	FOI ...que (canó.)		Q
40	12			1
41	15			
42	10	1		1
total	37	1		2

PSEUDO		pseudo com concord	quem...FOI	Q
43		2	5	1
44		2	6	1
45		2	6	1
total		6	17	3

PSEUDO COM CONCORD		pseudo sem concord	quem...FOI	Q
46		4	3	
47		4	4	
48		4	3	
Total		12	10	