



A contrastive study of questions in Mandarin and French: from theory to corpus analysis

Hung-Hsin Hsu

Graduate Institute of Linguistics, National Chengchi University, Taiwan
Centre for English Corpus Linguistics, Institut Language et Communication, Université
catholique de Louvain, Belgique

Committee members

One-Soon Her (Tunghai University, National Chengchi University), supervisor
Marie-Aude Lefer (Université catholique de Louvain), supervisor
Kristel Van Goethem (Université catholique de Louvain), chair and examiner
Kevin Henry (Université de Mons), external examiner
Marc Allasonnière-Tang (Muséum National d'Histoire Naturelle), external examiner
Elizabeth Zeitoun (Academia Sinica), external examiner

Taipei (Taiwan) & Louvain-la-Neuve (Belgium)
2026.04.24

Acknowledgement Remerciement 致謝

While this dissertation appears to be the work of a single author, it is, on the contrary, the result of the support of many people around me: my two supervisors, my committee members, my colleagues, my friends, and my family.

I would first like to express my deepest gratitude to my two supervisors, One-Soon Her and Marie-Aude Lefer. Throughout my PhD training, I have learned a lot from both of them. Our discussions have always provided me with valuable insights and guidance. One-Soon has guided me in developing theoretical reasoning and in shaping the framework of this research. It is also thanks to him that I had the opportunity to explore this topic. Marie-Aude has always shown great enthusiasm for my cross-linguistic work and has generously devoted her time to guide me through every step of corpus design and analysis. Thanks to the confidence and trust they have placed in me; I have been able to conduct my doctoral project with a great deal of independence.

The quality of this dissertation has also greatly benefited from the careful evaluation of my committee members. I would like to thank Kevin Henry and Marc Allasonnière-Tang for their valuable comments on corpus design and statistical analysis during our regular annual meetings. I am also grateful to Kristel Van Goethem for serving as the president of my defense and for sharing her expertise in contrastive linguistics. Moreover, I am deeply thankful to Elizabeth Zeitoun for accompanying me throughout my academic journey and witnessing my growth from the master's level to the PhD stage.

Along this journey toward my goal, my PhD path has been made all the more joyful by the companionship of my colleagues and friends. During the first three years of my PhD at NCCU, I had the pleasure of getting to know 宋佳興學長, 林志安, 曹子允, 梁永平, 吳欣蓉, 范姜沛柔, 傅怡萱, 陳宇倫. Thanks to their companionship, I never felt along. During the second part of my journey at UCLouvain, I was fortunate to be surrounded by a large community. I am especially grateful to the Alpha team under the supervision of Marie-Aude, including Thomas Prinzie, Alice Heylens, Romane Bodart, and Laura Aguiar De Souza Penha Marion. I would also like to thank my colleagues at the Centre for English Corpus Linguistics (CECL): 陶亞婷, Laurie Radar, Tanguy Dubois, Pauline Jadouille, 黃靈敏, 沈家成, Gabriela Vaughan, Deborah Fifelola, Thomas De Wispelaere, Florence Detienne, Peter Thwaites, Elen Le Foll, and Elaine Millar. In

addition, I am thankful to colleagues from other research groups and institutes: Laetitia Louis, Laetitia Aulit, Pauline Degrave, Stéphanie Pécher, Coline Rondiat, Remy Decorte, 何百羚, 李宥澤, 顏英玉, 胡駿飛, Sarah Schroyens, Jorina Brysbaert, Natasha Devroede, Antoine Ghislain, and Letizia Sassi. Outside academia, I am deeply grateful to my friends who have supported me consistently over the years: 陳仲豪, 任永慈, 李書帆, 林孟潔, Corrinne Quemener, Jean-Michel Quemener, Florent Courally, 施以潔學姊. I am aware that this list is far from exhaustive and could easily extend for several more pages. I sincerely appreciate everyone who has accompanied me on this journey.

In addition to the emotional support, I am also grateful for the financial support that made my PhD journey possible. I was fortunate to receive a three-year PhD scholarship from National Chengchi University (2019-2022). During my time abroad, I was supported by a scholarship from the Ministry of Education (Scholarships for Excellent Students to Study Abroad, 2022-2023), as well as two scholarships from the National Science and Technology Council (Graduate Students Study Abroad Program, 2023-2024, and the Grant for Doctoral Candidates in the Humanities and Social Sciences, 2024-2025). I am equally thankful for additional financial support from the College of Foreign Languages & Literature (NCCU), and Institut de Langage et Communication (UCLouvain). Altogether, these resources enabled me to attend conferences and share the results of my research. Most importantly, this support has been indispensable in helping me achieve my goal.

最後，這本論文的完成也有賴於家人們給予我的愛與支持，從碩士班到博士班，我的父母、妹妹以及弟弟始終給我最溫暖而堅實的力量。無論身處高山或是低谷，他們都一直與我同行。感謝我的父母信任我，讓我能夠選擇屬於自己的人生道路，也感謝他們身體健康，使我得以無後顧之憂地追求並完成自己的目標。感謝我的妹妹和弟弟，與我分享生活中的點滴，也感謝他們在我遠行之際，替我陪伴與照顧在台灣的家人，讓我能夠安心前行。

徐宏欣 Hung-Hsin Romain Hsu

Table of contents

Abstract	ix
Abstract (Mandarin)	x
Abstract (French)	xi
Notational Conventions	xiii
Chapter 1 Introduction	1
1.1. Background and motivation	1
1.2. Objectives of the study and research questions	3
1.2.1 Objectives.....	3
1.2.2 Research questions	3
1.3. Theoretical Framework	6
1.4. Defining the key constructs through <i>tertium comparationis</i>	11
1.5. Methodology: Corpus linguistics.....	15
1.6. Outline of the dissertation.....	21
Chapter 2 Formal syntax and semantics of questions.....	23
2.1. Overview of the construct of questions.....	23
2.1.1 Question vs. Interrogative	23
2.1.2 Typology of questions and question types.....	25
2.1.3 Typology of questions and marking strategies for questions.....	29
2.1.3.1 Marking strategies for polar questions	30
2.1.3.2 Marking strategies for constituent questions	35
2.1.3.3 Marking strategies for disjunctive questions	38
2.1.4 Summary of typology of question types and marking strategies	40
2.2 Questions in Mandarin and French.....	42

2.2.1.	Question types and marking strategies in Mandarin.....	43
2.2.2.	Question types and marking strategies in French	48
2.2.3.	Summary	54
2.3	The dichotomic view of questions in Mandarin and French	55
2.3.1	The delineation of a disjunctive question	55
2.3.2	Empirical evidence of taxonomy of question in Mandarin and French	61
2.3.3	Summary	69
2.4	Question words	70
2.4.1	Question words and their syntactic positions.....	70
2.4.2	Question words in Mandarin and French.....	73
2.4.3	<i>Wh</i> -in-situ in Mandarin	85
2.4.4	<i>Wh</i> -in-situ in French.....	95
2.4.5	Summary	102
2.5	Question particles	105
2.5.1	Particles in Mandarin	107
2.5.1.1	Ma particle.....	110
2.5.1.2	Ba particle.....	114
2.5.1.3	A particle	119
2.5.1.4	Ne particle.....	121
2.5.2	Summary of four particles in Mandarin.....	127
2.5.3	Question particles in French	131
2.5.3.1	Question particle ‘Est-ce que’	132
2.5.3.2	Question particle ‘ça’ and ‘done’	137
2.5.4	Summary of particles in French.....	144

2.5.5	A tentative conclusion on question particles from a cross-linguistic perspective	146
2.6	Summary of Chapter 2	148
Chapter 3 Data and Methodology: The design, compilation and analysis of a bidirectional parallel corpus		150
3.1	Corpus-based contrastive studies: tools, languages and topics	150
3.2	Corpus design and compilation	159
3.2.1	Selection of Taiwan Mandarin novels	161
3.2.2	Selection of French novels	167
3.2.3	Summary on the selection of novels for the bilingual corpus	172
3.3	Data preparation: digitalization, alignment, extraction	173
3.4	Data coding: question type, morphosyntactic structure, register	174
3.4.1	Question type	175
3.4.2	Morphosyntactic structure	176
3.4.3	Register	182
3.5	Data coding: interrater reliability	184
3.5.1	Question type	185
3.5.2	Morphosyntactic structure	186
3.5.3	Register	189
3.5.4	Overall results of interrater reliability	190
3.6	Data analysis: statistical tests	191
3.7	Summary of Chapter 3	191
Chapter 4 Comparable analysis of questions in Taiwan Mandarin and French		193
4.1	Overall frequency of questions in comparable corpus data	193

4.2	The distribution of questions in registers, question types, and morphosyntactic structures	199
4.2.1	Register	199
4.2.2	Question type	201
4.2.3	Morphosyntactic structures of questions in Taiwan Mandarin and in French.....	206
4.2.3.1	Morphosyntactic structures of questions in Taiwan Mandarin	207
4.2.3.2	Morphosyntactic structures of questions in French.....	214
4.2.3.3	Summary of morphosyntactic structures of questions in Taiwan Mandarin and French	221
4.2.4	Contrastive insights of questions in Mandarin and French.....	222
4.3	Contrastive analysis: taxonomy of questions	223
4.4	The analysis of <i>wh</i>-in-situ in French	234
4.5	Contrastive analysis: question particles	244
4.5.1	Particles in Mandarin	245
4.5.2	Particles in French.....	251
4.6	Contrastive analysis: summary and insights to parallel analyses.....	258
Chapter 5 Bidirectional parallel and monolingual comparable analyses of Mandarin and French		265
5.1	Bidirectional parallel analysis	265
5.2	Bidirectional parallel analysis: Taxonomy of questions	267
5.2.1	Translation of questions types from Taiwan Mandarin to French	268
5.2.2	Translation of questions types from French to Taiwan Mandarin	275
5.2.3	Discussion on bidirectional parallel analysis of question types.....	284
5.3	Bidirectional parallel analysis: <i>Wh</i>-in-situ	288
5.3.1	Translation of <i>wh</i> -questions from Taiwan Mandarin to French <i>wh</i> -in-situ ..	289

5.3.2	Translation of <i>wh</i> -in-situ from French to Taiwan Mandarin <i>wh</i> -questions ..	295
5.3.3	Discussion on bidirectional parallel analysis of <i>wh</i> -in-situ	297
5.4	Bidirectional parallel analysis: Question particles	299
5.4.1	Translation of question particles from Taiwan Mandarin to French	301
5.4.2	Translation of question particles from French to Taiwan Mandarin	306
5.4.3	Discussion on bidirectional parallel analysis of question particles.....	310
5.5	Summary of bidirectional parallel analysis of questions in Taiwan Mandarin and French	314
5.6	Monolingual comparable analysis.....	317
5.6.1	Translation features: Interference and normalisation in translated language	318
5.6.2	Overall frequency of questions in each translated language.....	322
5.7	Monolingual comparable analysis: Taxonomy of question types	329
5.7.1	Comparison of question types in non-translated and translated Taiwan Mandarin.....	330
5.7.2	Comparison of question types in non-translated and translated French	332
5.7.3	Discussion of question types in non-translated and translated languages	333
5.8	Monolingual comparable analysis: <i>Wh</i>-in-situ	334
5.8.1	Comparison of <i>wh</i> -in-situ in non-translated and translated French	334
5.8.2	Discussion of <i>wh</i> -in-situ in non-translated and translated French	338
5.9	Monolingual comparable analysis: Question particles	339
5.9.1	Comparison of question particles in non-translated and translated Taiwan Mandarin.....	340
5.9.2	Comparison of question particles in non-translated and translated French ..	347
5.9.3	Discussion of question particles in non-translated and translated languages.....	353

5.10	Summary of monolingual comparable analysis of questions in Taiwan Mandarin and French	354
5.11	Exploring the relationship between questions, genre, authorship, and translation status.....	356
5.11.1	Question types in Taiwan Mandarin and French	357
5.11.2	Morphosyntactic structures in Taiwan Mandarin and French	360
5.11.3	Cross-linguistic summary of question types and morphosyntactic structures with respect to register, translation status, and genre	369
Chapter 6	General conclusion	373
6.1	Summary of the main findings	373
6.1.1	Question types.....	374
6.1.2	<i>Wh</i> -in-situ.....	377
6.1.3	Question particles.....	378
6.1.4	Integrated analysis based on Bayesian regression models.....	380
6.1.5	Summary of the cross-linguistic findings in Taiwan Mandarin and French.....	382
6.2	Main contributions of the study	383
6.2.1	Contribution to contrastive linguistics	383
6.2.2	Contribution to the theory of questions.....	387
6.2.3	Contribution to corpus linguistics	388
6.3	Limitations of the study	390
6.4	Promising avenues for future research.....	392
6.5	Implications for related fields.....	394
	Appendix.....	396
	References.....	409
	References in Mandarin.....	443
	References of novels.....	446

Abstract

The aim of this dissertation is to undertake a cross-linguistic examination of questions in Mandarin and French. By combining theoretical discussions with empirical corpus analyses, I endeavour to shed light on three key topics: question types, *wh*-in-situ, and question particles.

The theoretical part initiates with a comprehensive overview of questions based on typology studies. Two primary issues are addressed: the classification of question types (i.e., polar questions, constituent questions, and disjunctive questions) and the marking strategies of questions (e.g., intonation, word order, tags, particles, question words). Following the overview in typology, the focus narrows to the two languages under investigation, Mandarin and French. In addition to presenting their question types and marking strategies, considerable attention is given to the **distinction of question types**, particularly the debated status of disjunctive questions. Utilizing empirical data gathered from the grammaticality judgement tasks in both languages, I explore the plausibility of different theoretical question taxonomies. Subsequently, attention shifts to two linguistic phenomena attested in both Mandarin and French. ***Wh-in-situ***, a default position of question words in Mandarin, is optional in French. Despite this difference, two syntactic accounts for *wh*-in-situ (e.g., LF movements and unselective binding) is applicable in both languages. However, when applying these mechanisms, the effect of syntactic constraints might be imposed differently between the two languages. Additionally, **question particles** are shared (e.g., Mandarin: 嗎 *ma* and 呢 *ne*; French: *est-ce que* and *ça*). Alongside an overview of these particles across different linguistic levels, with pertinent discussions in formal syntax, primary focus is placed on comparing the 呢 *ne* and the *ça* particle. Both particles allow occurrence with constituent questions and, pragmatically, are used for co-referencing to the preceding context.

On the basis of theoretical overview and discussion, this study examines authentic language use through a self-compiled bidirectional parallel corpus, comprising six original novels in Mandarin and French, along with their respective translations. The texts are processed using Sketch Engine for automatic extraction, and subsequent coding of extracted data involves identifying question types and marking strategies. The coding facilitates following analyses. Analyses are conducted from three perspectives. First, the bilingual comparable approach investigates the distribution of question types and marking strategies across the two source languages. Second, the bidirectional parallel approach examines how cross-linguistic differences are reflected or accommodated in translations. Third, the monolingual comparable approach assesses the translation effect by comparing non-translated and translated languages within the same language. To further examine the data, a Bayesian regression model is applied to evaluate the influence of multiple variables, including register, language, translation status, and literary genre, on the distribution of questions and related syntactic phenomena.

The concluding chapter integrates findings from both theoretical and empirical analyses, highlighting the contributions to contrastive linguistics, corpus linguistics and the theory of questions. It also discusses limitations, proposes directions for future research, and suggests practical applications for pertinent fields, such as language pedagogy, language learning, and translation studies.

Abstract (Mandarin)

本論文的研究目的在於探討中文以及法文跨語言的問句對比，藉由理論以及實際語料兩者所提供的視角，特別針對以下三個主題進行探討：問句類型的分類，疑問詞原位現象（*wh-in-situ*），以及疑問助詞。

首先，理論方面將從語言類型學對於問句的描述開始，並主要針對兩個議題進行闡述：問句類型（是非問句、疑問詞問句、選擇問句）以及問句標記策略，例如：語調、語序、附加問句、疑問助詞、疑問詞等。在完整介紹語言類型學中針對問句的討論後，我將焦點限縮至中文和法文並針對這兩個語言的問句類型以及問句標記策略提供細部的介紹。但由於長久以來問句分類一直都是相關研究的討論重點之一，因此我將會針對這個議題有較深入的探討，並提出合理性相對較高且能夠涵蓋較廣的問句分類法。並另外針對中文及法文，使用語言實證的實驗（文法判斷任務）來觀察此分類法的有效性。接著，研究的焦點將轉移到中文及法文共有的語言現象。**疑問詞原位**在中文是一種語法常規，在法文卻是一種選擇性的語法表現，雖然這個現象在兩個語言間有著本質上的差異，但這樣的句法現象卻能夠使用相同的句法模式來解釋，例如：邏輯形式移位（LF movement）以及無擇約束（unselective binding）；但實際上，在推導這些句法操作的過程時，兩個語言間仍舊存在著對於相關句法限制的差異。另一個兩者共有的語言現象則是**疑問助詞**，例如：中文的“嗎”和“呢”以及法文的“*est-ce que*”和“*ça*”。除了從不同語言學角度的描述來分析這幾個疑問助詞外，亦會針對形式語法對於疑問助詞相關的研究進行討論。另外，亦會特別針對中文的“呢”以及法文的“*ça*”進行跨語言的對比討論，因為兩者除了在語法的分布上接近外，同時在語用上亦具備指涉並與前文連結的效果。

在理論相關的闡述與討論基礎下，我將透過語料庫進一步觀察語言實際使用的情形。該語料庫是由研究者自行建構的雙向平行語料庫（bidirectional parallel corpus），語料來源為中文以及法文各六本原文小說以及他們相對應的翻譯本。這些語料透過 Sketch Engine 軟體進行自動抽樣，接著，根據問句類型以及標記策略來註記所抽出的目標語料。語料的分析將從三個層面展開，首先，透過雙語可比（bilingual comparable）的角度比較中文與法文在問句使用上的跨語言差異，其次，從雙向平行（bidirectional parallel）角度檢視譯文如何反映或處理跨語言的差異，最後，透過單語可比（monolingual comparable）的方式，比較同一語言中原文與譯文的差異是否受到翻譯因素的影響。透過這三種分析角度，本研究得以細緻呈現兩種語言問句使用的異同。進一步地，本研究亦運用 Bayesian regression model 來評估多項變數（語域、語言、翻譯狀態、文學類型等）對於問句分佈與句法現象的影響程度。

在完成理論以及實證兩部分的分析後，結論章節整合跨不同角度的研究結果，並說明本論文在對比語言學、語料庫語言學以及問句研究領域的主要貢獻。同時，本章將基於研究限制提出未來可能的改善方向與進一步的研究議題。此外，本論文將從應用層面提出建議，例如：對二語教學與習得、翻譯研究等領域可能的幫助。

Abstract (French)

Cette thèse vise à analyser les questions en chinois et en français dans une perspective contrastive. En se fondant sur la discussion théorique et une analyse empirique de corpus, l'objectif principal de ce travail est de mettre en lumière trois aspects: les types de questions, les interrogatifs in-situ et les particules interrogatives.

Le volet théorique commence par une description des questions, fondée d'un panorama approfondi des travaux antérieurs en études typologiques. Deux thèmes sont traités en détail: les types de questions (c'est-à-dire les questions fermées, ouvertes et alternatives) et les marqueurs de questions (par exemple, l'intonation, l'ordre des mots, les particules et les mots interrogatifs). Après un état de la question générale, nous nous concentrons sur les deux langues étudiées dans le cadre de cette thèse, le mandarin et le français. Outre la présentation de leurs types de questions et marqueurs de questions, nous prêtons une attention particulière à la **distinction des types de questions**, étant donné les débats théoriques actuels, particulièrement en ce qui concerne la classification des questions alternatives. Au moyen de données empiriques issues de tests de grammaticalité dans les deux langues étudiées, nous évaluons la validité de différentes taxonomies de questions proposées dans les travaux antérieurs. S'ensuivent des discussions des deux phénomènes linguistiques présents dans les langues chinoise et française. **Les interrogatifs in-situ** émergent comme une composante fondamentale dans les questions en mandarin, tandis qu'ils sont optionnels dans les questions en français. Malgré cette différence interlinguistique, des analyses syntaxiques de cette structure peuvent être opérées dans les deux langues, comme en témoignent le mouvement de la Forme logique (FL) et le liage indifférencié. Cependant, lors de l'application de ces mécanismes, les effets syntaxiques s'imposent différemment dans les deux langues. Outre les interrogatifs in-situ, **les particules interrogatives** sont observées dans ces deux langues (par exemple, en chinois 嗎 *ma* et 呢 *ne*; en français *est-ce que* et *ça*). À travers une description détaillée de ces particules à divers niveaux linguistiques, ainsi que des apports théoriques pertinents en lien avec la syntaxe formelle, nous comparons la particule 呢 *ne* et *ça*. En effet, ces deux éléments sont utilisés dans les questions ouvertes. De plus, d'un point de vue pragmatique, leur utilisation vise à établir un lien avec le contexte précédent.

Dans un second temps, nous étudions l'usage authentique à travers des données électroniques réunies afin de constituer un corpus parallèle bidirectionnel. Ce corpus comprend six romans originaux en chinois et en français, ainsi que leurs traductions respectives dans l'autre langue. Les textes ont été intégrés à la plateforme Sketch Engine afin de procéder à l'extraction automatique des données pertinentes (les questions dans les romans originaux et leurs équivalents dans les romans traduits). Les données extraites ont ensuite été codées en fonction des types de questions et de leurs marqueurs spécifiques. Nos analyses s'appuient sur trois approches complémentaires. Premièrement, l'approche bilingue comparable permet d'examiner l'usage des questions à travers les deux langues. Deuxièmement, l'approche parallèle bidirectionnelle sert à vérifier dans quelle mesure les traductions reflètent les différences interlinguistiques observées dans les données comparables. Enfin, l'approche monolingue comparable compare les données originales et traduites au sein d'une même langue afin d'évaluer les effets de la traduction. Les analyses sont ensuite approfondies à l'aide d'un modèle statistique, le *Bayesian regression*, afin d'examiner l'effet de plusieurs variables (e.g., registre, langue, statue de traduction, et genre littéraire) sur la distribution des questions.

Enfin, la conclusion présente une discussion intégrant les analyses et les résultats, combinant perspectives théoriques et observations empiriques. Elle met en lumière les principales contributions de cette étude à la linguistique contrastive, à la linguistique de corpus et à la théorie des questions. En outre, sur la base des limites identifiées dans cette recherche, des pistes prometteuses pour des études futures sont proposées. Enfin des applications pratiques sont suggérées pour des domaines pertinents tels que l'enseignement et l'apprentissage des langues étrangères, ainsi que la didactique de la traduction.

Notational Conventions

The glossary abbreviations used in this dissertation follows *The Leipzig Glossing Rules* (Comrie et al. 2008) and Creissels (2006) for French-specific glosses. They occur in the interlinear glosses of examples to represent the grammatical functions of words. Note that some glosses may remain identical if they are examples cited directly from literature.

1	first person
2	second person
3	third person
ACC	accusative
AGT	agent
APRT	present participle
ART	article
ASP	aspect
CL	classifier
CLIT	clitic
COMP	complementiser
COND	conditional
COP	copula
DEM	demonstrative
FMN	feminine
FOC	focus
FUT	future
ILL	illative
IMPARF	imperfect
IND	indicative
INF	infinitive
INS	instrumental
INT	interrogative
NEG	negation, negative

NOM	nominative
OBJ	object
PAR	partitive
PL	plural
POL	polite
POSS	possessive
PREP	preposition
PRF	perfective
PROG	progressive
PRS	present tense
PST	past tense
PTCL	particle
PTCP	past participle
Q	question
REFL	reflexive
REL	relative pronoun
SBJ	subject
SG	singular

Chapter 1 Introduction

1.1. Background and motivation

Questions are a common component of daily life and they are a fundamental property in all human languages (Siemund 2001: 1010; Moyer & Syrett 2021: 1). Not only do questions appear in colloquial conversations to seek confirmation or substantial information from addressees, but questions can also be found in formal situations, such as interrogation by law enforcement officials during investigations. Moreover, questions help inspire people and further the development of science and society (Seyferth, Ratna & Chung 2022: 1031). For instance, when Isaac Newton once asked: *If the apple falls, does the moon also fall?*, the curiosity raised by a question led him to discover gravity. Since questions play such an important role in humans' daily lives, understanding the nature of questions is a continuing concern within various linguistic disciplines.

In the field of phonology, studies have tried to capture the distinctiveness of questions by examining intonation patterns (Geluykens 1988; Hedberg, Sosa & Görgülü 2017; Yang, Gryllia & Cheng 2020); sentence stress patterns (Ladd 2001); the interaction between intonations and lexical tones (Chao 1968: 26–39; Kung, Chwilla & Schriefers 2014; Torreira, Roberts & Hammarström 2014); and acoustic perception of questions (Gussenhoven & Chen 2000; Yuan 2006, 2011). Regarding the morphosyntactic level, word order and particles often come into play. Some languages have a rearrangement of word order (i.e., inversion); however, it has been suggested that such constructions are limited to the European languages (Ultan 1978; Siemund 2001; Hölzl 2018: 63). Consequently, studies related to inversions in questions are relatively common in languages, such as French (Lahousse 2003a; Stark & Binder 2021), Spanish (Ordóñez 1998), Italian (Bianchi, Bocci & Cruschina 2017). In contrast, languages like Mandarin use particles to indicate questions without any shift in word order (Lin 2015; Luo & Wu 2017; Bhatt & Dayal 2020). Concerning semantics, researchers are concerned with the sets of propositions contained in different question types (Hamblin 1973; Karttunen 1977; Groenendijk & Stokhof 1984; Ciardelli, Groenendijk & Roelofsen 2018: chap. 5), despite that studies have differed in the conceptualization of propositions in questions (Roelofsen 2019: 2). Pragmatics pays attention to the pragmatic functions that questions may carry,

such as rhetorical functions (Caponigro & Sprouse 2019) and other indirect uses (e.g., suggestion, permission, discourse initiating) (Stivers & Enfield 2010: 2623; Siemund 2017). As questions are pervasive in human communication, they have also caught the attention of sociolinguists who aim to explore how questions might be used differently based on gender, age, region, social class, and social status. In one study, the authors analyse how children from different social classes (i.e., working class vs. middle class) adopt the question-answer system (Robinson & Rackstraw 1977). Another study shows that word order rearrangement for questions is quite infrequent in the SMS of the younger French-speaking population between the ages of 12 to 19 in Switzerland, compared to the group between the ages of 30 to 68 (Guryev 2017: 221). In addition, the sociolinguistic aspect has identified a variety of question forms in different social contexts. Such a variety can only be uncovered if the studies of questions go beyond theoretical discussions and take language use into account.

Clearly, questions as a linguistic topic have been widely treated at different theoretical levels along with empirical studies, yet most studies are limited to monolingual analyses. Cross-linguistic work can help better shape and understand the question system of languages. Despite the advantages brought by cross-linguistic works, there is concern that the latter often lacks diversity in terms of the languages that are taken to compare. Hence, only a few contrastive studies to date have been attributed to language pairs outside of European languages. For instance, according to the statistical evidence provided by Gast (2012), the majority of languages studied in the journal *Languages in Contrast* are from Germanic and Romance language families. Based on the contents of *Languages in Contrast* from the past three decades, Hasselgård (2020) also finds that English is still the dominant language involved in such studies. However, this situation might have gradually shifted in the past ten years.

Thus, in reviewing the literature, there are two motivations for this study. First, I think a contrastive study of questions is necessary in that such cross-linguistic findings assist in our understanding of how questions in languages are used differently or similarly. Secondly, to grasp the full picture of questions in language systems, it is best to look beyond the English-centred vision and to compare languages that are distant in language family, such as the two languages examined in this dissertation, Mandarin and French.

1.2. Objectives of the study and research questions

1.2.1 Objectives

As questions have been addressed in various linguistic studies, they have not yet been extensively studied, especially in corpus-based contrastive studies (Curry & Chambers 2017: 327; Axelsson 2020: 238). Thus, the first objective of this dissertation is to use electronic corpora and corpus-linguistic methods to compare the actual use of questions in the two languages.

Secondly, language pairs taken for comparison in contrastive linguistics too often involve English or European languages; other languages seem to have received less attention. Not to mention that the language pair of Mandarin and French is far less frequent than the pair of Mandarin-English (Poiret 2020: 6). Hence, the second objective of this research is to break away from the long-standing practice of using Mandarin-English contrastive language combination and to use Mandarin-French in this study.

A third objective is to compensate for the lack of bidirectional parallel Taiwan Mandarin-French corpora. I have therefore compiled my unique corpus, comprising original texts in Mandarin and French and their translations in both language directions.

Finally, this dissertation is concerned on the one hand with theoretical considerations and on the other hand with corpus analysis. Therefore, the last objective of this research is to demonstrate how to deal with questions from two different perspectives. The ultimate purpose of this study is to examine and compare questions in Mandarin and French based on the theory of comparative syntax and a self-built bidirectional parallel corpus. Combining the two different approaches can not only contribute to a more thorough understanding of the topic with which I work, but also highlight the extent to which theory and language use can match or mismatch.

1.2.2 Research questions

As shown from the title of this dissertation, the main topic being addressed is the construct of ‘question’ in Mandarin and French through theoretical discussions and corpus analyses. More precisely, the topic of ‘question’ is approached from three angles: question types, the position of question words, and the particles that appear in questions. I will start with the delineation of question types based on a currently proposed taxonomy

of questions before moving on to the other topics which concern features that are shared in the two languages studied here.

First, regarding question types, Her, Che & Bodoño (2022) recently proposed a dichotomy of questions in Mandarin along with a tentative view on English question types. They argue that a dichotomy of questions in Mandarin (i.e., confirmation-seeking vs. information-seeking questions) can be justified by a series of syntactic tests and inferences from semantics. Based on the results, they boldly suggest that theoretically, English has only one question type, namely information-seeking questions. Building upon this perspective, I extend this taxonomy to French. Accordingly, the first two research questions are as follows:

1. Theoretically, what is the taxonomy of question types in French in comparison to a dichotomy of questions in Mandarin?
2. In the corpus study, what is the distribution of question types? Which factors can explain such distribution? Are the cross-linguistic similarities and differences found in comparable corpus reflected in the parallel corpus?

Secondly, the topic of question word position has received considerable critical attention in monolingual studies (see Huang 1982, Cheng 1991 and Yuan 2020 for Mandarin and Obenauer 1976, Adli 2006 and Shlonsky 2012 for French). However, when closely examined, Mandarin and French have a common property in terms of question word position, which is termed *wh*-in-situ. This occurs when a question word resides in its original syntactic position in a non-interrogative sentence without fronting, as shown by *shénme* ‘what’ in Mandarin in example (1.1) and *quoi* ‘what’ in French in (1.2). Note that Mandarin is considered a *wh*-in-situ language (Huang, Y.-H. Audrey Li & Li 2009: 260; Yip & Rimmington 2021: 161), while question words in French normally appear at the beginning of questions, as in (1.3).

(1.1) 你 吃 什麼 (呢)?
 nǐ chī shénme (ne)?
 2SG eat what PTCL
 ‘What do you eat?’

(1.2) *Tu manges quoi ?*
 2SG eat what
 ‘What do you eat?’

- (1.3) *Que* *manges-tu ?*
 what eat.2SG-2SG
 ‘What do you eat?’

Despite this obvious syntactic similarity, only a few studies have paid attention to this phenomenon contrastively (see Cheng & Rooryck 2000 and Wu 2019). Both studies proceed with the topic under the generative framework. Cheng & Rooryck (2000) have primarily demonstrated the licensing conditions of *wh*-in-situ in French, given that this syntactic feature is optional. Only a few contrastive aspects with Mandarin *wh*-in-situ are mentioned. On the other hand, Wu (2019) analyses cross-linguistically *wh*-in-situ in both languages, but only within ten pages. He has pointed out that there are more differences than similarities in terms of the syntactic mechanisms behind such a phenomenon. Accordingly, in addition to the fact that no previous study has comprehensively investigated this topic, the contrastive analyses to date have extensively tended to account for such a linguistic phenomenon from a formal perspective. Therefore, it is worth studying this topic with not only a thorough theoretical discussion but also with an examination of corpus data. The synergy of two different perspectives can shed new light on the topic of *wh*-in-situ. Accordingly, the following two research questions are as follows:

3. Theoretically, what are the differences and similarities of *wh*-in-situ questions between French and Mandarin?
4. In the corpus study, what is the proportion of *wh*-in-situ in French? How French *wh*-in-situ questions are translated into Mandarin? What types of structure in Mandarin *wh*-questions often translated with *wh*-in-situ in French?

Lastly, in Mandarin, the final particles used in questions (i.e., *ma* and *ne*) have been frequently addressed in the literature, such as in Chu (2009) and Fang & Hengeveld (2022). As for French, *est-ce que* has been recognised as a question particle in the literature (Harris 1998: 237; Dryer 2013a). Recently, works have also been dedicated to studying another question particle (i.e., *ça*), such as in Lefeuvre (2018) and Smirnova & Abeillé (2021). However, to the best of my knowledge, there have not been any contrastive studies of these particles to date. I thus aim for an in-depth cross-linguistic analysis, paying particular attention to the morphosyntactic representations of these

particles and the distribution and the use of question particles in corpus data. The research questions are as follows:

5. Theoretically, what are the differences and similarities of question particles between the cases recently discussed in French and those in Mandarin?
6. In the corpus study, what are the frequencies of these question particles in the comparable corpus? How are question particles translated in both language directions?

1.3. Theoretical Framework

Studies concerning the comparison of languages are often related to the following two disciplines: typology and contrastive linguistics. However, they serve different objectives. Typology is the study of linguistic systems from both synchronic and diachronic perspectives and it aims to uncover the differences and similarities between a larger number of languages (Velupillai 2012: 15,18-19). In addition, by observing recurring patterns across languages, the ultimate goal of typology studies is to answer the question of *what's where why* (Bickel 2007: 239), namely, typologists need to account for the distribution of such recurring patterns from geographical, genealogical, and historical perspectives. Likewise, contrastive linguistics does compare languages as do the studies of typology. Accordingly, contrastive linguistics can also be defined as the systematic study of discovering the differences and similarities in the structure and use of two or more languages (Bugarski 1991: 77). However, certain distinctions have been made between contrastive linguistics and typology, such as their scope and granularity (König 2012: 10). Their scope differs not only in number of languages but also the parameters taken into account for the comparison. Hence, the number of languages involved in contrastive linguistics studies is limited (at least two or more) compared to typology studies aiming for all the languages with available documentation. On the other hand, the parameters used to conduct a comparative work can vary under a contrastive linguistics framework while typology only focuses on a few parameters for comparison. As for granularity, contrastive linguistics offers a relatively detailed description of the comparison between languages, with special attention to differences (Gómez González & Doval-Suárez 2005: 8; Gast 2012; Ke 2019: 8; Ramón García 2021: 394). Finally, contrastive linguistics aims eventually for theoretical or practical purposes, such as foreign language teaching and translation (Gast 2012). In short, contrastive linguistics is

a study that combines cross-linguistic comparisons in a small number of languages and in-depth analyses and it is intended to apply for other purposes (Gómez González & Doval-Suárez 2005: 21, 23).

Given that the focus of the dissertation is devoted to the specific structure in the two languages, Mandarin and French, the current research is well-positioned in contrastive linguistics. In addition, since the core of this study concerns the syntax of questions, I adopt specifically comparative syntax as the theoretical framework in this dissertation. Arguably, this framework intends to encompass works comparing two or more languages and highlighting the cross-linguistic differences in syntax (Cinque & Kayne 2008: 3; Haspelmath 2014: 490).

Comparative syntax, however, seems to allow for two different perspectives, functional-typological and formal. Even though both approaches are concerned with the comparison of cross-linguistic structures, they are distinguished by assumptions, goals, methods, and explanations (Haspelmath 2014: 492–494). For the practitioners of the functional-typological approach, the goal of studies is to find general properties of languages. To achieve this goal, detailed descriptions and comparisons are required, but they can be done separately, often with different groups of researchers using different sets of analytical tools. Description works following the functional-typological approach tend to analyse a particular language using descriptive categories, which are possibly varied from one language to another. For example, it does not seem plausible to apply case categories from Latin (e.g., nominative, accusative, etc.) to the analysis of languages with a poor system of case marking, such as English. Accordingly, it has been argued that each language should be analysed on its own terms (Boas 1911). In addition, given that these descriptive categories depend on the inner form of each language, they can also be considered language-specific features. On the other hand, cross-linguistic comparison, which can be reasonably separated from descriptive works following this approach, is much concerned with comparability between languages. In addition, such comparability should be built on the sound, meaning, or gesture rather than grammatical categories (e.g., subject, object, verb), because the former are universal (Haspelmath 2014: 495). Thus, these concepts are closer to conceptual-semantic features of languages, such as agent and patient (see Haspelmath 2010 for more details). In the comparison of word orders, instead of using grammatical categories as notations (e.g., S-V-O and S-O-V), English has agent-

action-patient word order, while Japanese has agent-patient-action order. However, regardless of comparing and describing languages, researchers should bear no a priori assumptions concerning the generality of languages. Consequently, this approach welcomes any explanation with possible factors, including historical, geographical, cultural, cognitive, communicative, etc.

In opposition to the functional-typological approach, the practitioners of the formal approach, particularly those working within generative grammar, tend to identify the result of comparison with innateness. This innateness accounts for the universal categories and features that are used to describe and compare across languages. Taking the example of word order, following the Uniformity Principle claiming that languages are uniform (Chomsky 2001: 2), the generative approach assumes that universal syntactic categories (e.g., subject, verb, and object) can be applied to any language and that these categories have identical sense across languages. The goal of such an approach is to build a formal framework to account for syntactic constructions of any language, given that all languages are innately built in the same categories and features. Table 1 shows the contrast between the two approaches.

Table 1 The contrast of comparative syntax between the two approaches

	Functional-typological	Formal
Goal	Find general properties	Build a formal framework
Assumptions	No a priori assumptions	With universal categories and features
Methods	Description and comparison are two separated procedures	Description and comparison are conducted simultaneously
Explanations	Being open to any possible explanations (e.g., historical, cultural, cognitive, etc.)	Only explanation by innate structures

In my view, despite disagreements in terms of proceeding works of comparative syntax, both approaches are similar in that they aim to uncover the syntactic universals (generalizations) in human language systems. Because the majority of theoretical concepts and terms used in this dissertation are related to generative theory, comparative syntax as applied herein will be placed closer to the formal approach. Particularly,

description and comparison will proceed simultaneously. Also note that I will only adopt the general framework of modern syntax in analyses and discussions, with a limited set of lexical categories, phrase structures, and *wh*-movement concepts, without commitment to the underlying assumptions of the formal approach, such as innateness explanation, universal categories. In the following paragraphs, I present the details of how comparative syntax is applied in formal approach.

A wide range of comparative syntax studies based on the generative framework are presented in the edited volume *The Oxford Handbook of Comparative Syntax* (Cinque & Kayne 2008). They all adopt similar concepts and steps illustrated in Kayne (2013), namely, three types of adequacies for scientific work, *observational*, *descriptive*, and *explanatory*. While, these adequacies represent the steps found in various fields of study, it is detailed below in the specific context of comparative syntax.

Observational adequacy: In comparative work, we should first observe the differences and similarities between the languages that we aim to compare. Yet, the level of closeness between the languages being compared can affect the outcomes and the ease of the observation. Therefore, there is a distinction between micro-comparative and macro-comparative based on the degree of closeness in the compared languages. Arguably, the closer the languages are, the easier their cross-linguistic equivalents seem to be revealed, such as in the case of French and Italian (i.e., a micro-comparative). On the contrary, the farther the languages are, the more difficult the equivalents emerge, such as French and English, compared to French and Italian (Kayne 2013: 137). Despite the different levels of language closeness involved, the main idea is to find syntactic differences and similarities in the languages we intend to compare. A good starting point for comparative syntax is how differences between languages surprise us, whether this surprise is big or small (Kayne 2013: 135). These surprising differences revealed by observation can lead us to the next level of the study.

Descriptive adequacy: In addition to describing the phenomena we observe based on the universal features and categories, it is desirable to formulate linguistic generalizations by observing enough languages, and these generalizations should be applicable for testing on additional languages (Kayne 2013: 135). Such generalizations are akin to the notion of universals in Greenberg (1963) (Croft 2002: 5–6; Cinque & Kayne 2008: 47). The common thread here is the exceptionless feature in generalizations

and universals, which indicates that such generalizations and universals should be applied to all languages. For instance, the Universal 12 put forth by Greenberg (1963) indicates the correlation between basic word orders of languages and the positions of question words. In addition, based on such a universal, another syntactic generalization has advanced for the correlation between the positions of question particles and question words (Baker 1970: 208). I will address more details of both cases in Chapter 2. The above two examples show that a tentative exceptionless principle can be deduced based on thorough observation. This is qualified as a fulfilment of descriptive adequacy.

Finally, *explanatory adequacy* is an ultimate goal for any scientific study: in comparative syntax, it needs to account for the generalizations that are derived from the cross-linguistic comparisons, such as the word order of a clitic object with infinitive in Romance languages (Kayne 2013: 140). Despite their differences, the functional-typological and formal approaches both believe that the explanations behind these generalizations or universals are derived biologically (Croft 2002: 6). The only difference lies in the biological evidence: from evolution in the functional-typological approach, and genetics in the formal approach. From the latter standpoint, in many cases, silent elements as well as syntactic movements are used to explain the cross-linguistic differences that prominently emerge from the comparisons. In addition, sometimes differences observed from cross-linguistic comparisons may be attributed to linguistic parameter settings in one particular language (Cinque & Kayne 2008: 6). As comparative syntax in the formal approach argues for universal categories given that languages are exceptionless, linguistic parameter settings are used to fine-tune the particularities found in languages to maintain the validity of universal categories across languages. Simply put, parameters represent all cross-linguistic differences (Cinque & Kayne 2008: 6). For instance, the ‘pro-drop’ parameter accounts for the fact that supposing every language must have a subject, some languages do allow the unstressed pronominal subject to remain silent, such as in Italian. These explanations, deduced from a generative perspective, suggest that silent elements or syntactic movements are deeply carved in to the human language faculty.

Overall, the comparative syntax framework differs from purely syntactic theory regarding the results it can offer (Cinque & Kayne 2008: 5; Kayne 2013: 139). In the context of purely syntactic theory, the goal is to provide results that require more complicated deductions. While in comparative syntax, the primary focus is to compare

the structures between languages and to describe and explain the differences found in the structures of languages, language-specific parameter settings in certain particularities in languages also receive extra attention. Thus, two goals of comparative syntax can be summarised as follows: first, identify cross-linguistic differences in syntax in addition to describing and delineating the parameters that contribute to such differences. Second, based on these differences, attempt to delineate a general principle of Universal Grammar (UG), which can be applied to any language (Cinque & Kayne 2008: 3). Ultimately, the syntactic landscape of human language faculty might be thoroughly explored (Cinque & Kayne 2008: 55). But once again, this dissertation does not intend to commit to any technical claim related to assumptions of innateness because of the use of this theoretical framework. A fortiori, this dissertation is based on corpus linguistics as the methodology, and there are incompatibilities concerning the concepts between corpus and generative theory, such as the exceptionless feature and “clear cases” (i.e., sentences have an obvious status of grammaticality; Levelt 1972: 21) in generative theory. This feature may be valid in the theoretical context. However, exceptions frequently arise in naturally occurring and attested language use in corpora and examples drawn from corpora may therefore not be as “clear” as those discussed in theoretical work. The discussion of incompatibilities will be further developed at the end of the methodology section.

Before moving on to the methodology, it is necessary to introduce an additional conceptual foundation that underpins the comparison conducted in this study. This foundation is articulated through the notion of the *tertium comparationis*, which provides a framework for defining the key constructs explored in the following section.

1.4. Defining the key constructs through *tertium comparationis*

Comparing languages requires a shared basis for comparison (Krzeszowski 1990: 15; Chesterman 1998: 29; Connor & Moreno 2005: 154; Granger 2010), whether this basis is syntactic, semantic or pragmatic (Egan 2013: 8). This common basis is known as the term *Tertium comparationis*, defined as: the background of sameness against which differences can be viewed and described (James 1980: 166; Johansson 2007: 39). The concept serves as a common platform for contrastive analysis, particularly in studies that incorporate parallel corpora, as emphasised by Johansson (2007: 33–34). By establishing a semantic and pragmatic *tertium comparationis* between the source and translation texts,

contrastive analyses can ensure that the elements under study are genuinely comparable across languages (Hasselgård 2020: 190).

As mentioned in Section 1.3, the three levels of adequacy characterise scientific research across various fields. In the specific context of cross-linguistic studies, theoretical adequacies are operationalised through a classical protocol. Two major steps are identified for contrastive study: (1) description and (2) juxtaposition for comparison (James 1980: 30). Krzeszowski (1990: 35) further extends this by identifying ‘comparison’ as the third step (i.e., contrastive analysis proper). Thus, these three steps constitute a classical protocol for cross-linguistic studies. The first step requires describing each language under study at the appropriate levels (e.g., phonology, lexis, morphology, syntax). The second step involves determining which elements are to be compared (Krzeszowski 1990: 36), a decision that is critical for the validity of the results (Krzeszowski 1990: 15; Lefer 2009: 60). As part of this step, a platform of reference or *tertium comparationis* should be established as the starting point for comparison (Lefer 2009: 58). In other words, cross-linguistic comparison should be grounded in the shared properties of the elements under study (Krzeszowski 1990: 117). These shared properties constitute ‘the third term of comparison’, which lies outside the scope of direct comparison (Lefer 2009: 60), as shown in Figure 1.

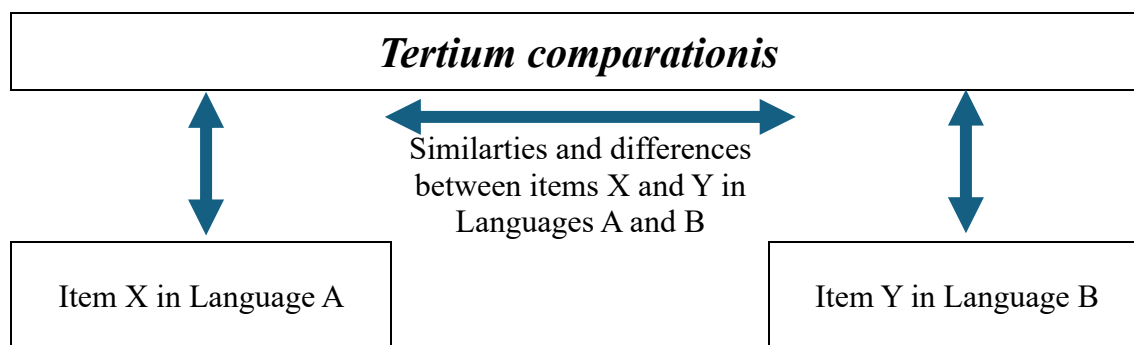


Figure 1 *Tertium comparationis* in contrastive analysis (Lefer 2009: 61)

During the juxtaposition phase, it is essential not only to identify comparable elements, which seems relatively easy to build within parallel corpus than in comparable corpus (Granger 2010; Ebeling & Ebeling 2020: 103), but also to ensure the degree of comparability provided by the corpus, such as matching register and genre (Krzeszowski 1990: 26; Connor & Moreno 2005: 154). Once the linguistic phenomenon under study have been fully described, the final step, comparison, can be undertaken.

Given the importance of *tertium comparationis*, various types of contrastive studies have employed this concept as the basis for cross-linguistic comparison of linguistic phenomena at different levels, including phonological, lexical, syntactic, pragmatics (James 1980: 28). Each type of contrastive study may require a distinct *tertium comparationis*, as demonstrated by James (1980: 69–93, 166–175). For phonology, this is represented by the International Phonetic Alphabet chart and vowel diagram. For lexicology, by a universal set of semantic components. In syntactic studies, three candidates for *tertium comparationis* have been proposed: surface structure, deep structure, and translation equivalence (i.e., semantic-functional equivalence). Translation equivalence is argued to be the best available *tertium comparationis* for contrastive analysis (James 1980: 175), particularly in syntax (Lefer 2009: 66). Following this approach, Cosme (2007: 131–132) applies translation equivalence as the *tertium comparationis* to study two discourse features: clause linking (i.e., coordination vs. subordination) and informational density (i.e., phrasal constituents vs. clausal constituents) in English, French, and Dutch. The study begins with an assumed semantic and/or functional similarity (i.e., coordination) and then examines the different forms that convey that the same meaning or function. Similarly, Lefer (2009: 111) uses semantic equivalence as the *tertium comparationis* for addressing semantic issues in lexicon. To establish the *tertium comparationis*, the author identifies the various meanings expressed by prefixes in English and French, first examining formal similarities (i.e., derivational prefixation). These shared meanings then serve as the basis for cross-linguistic comparison of prefixation.

The present study also adopts semantic-functional equivalence as the *tertium comparationis*, following the steps for establishing a common basis of comparison as outlined in Cosme (2007). Specifically, the study begins with the perceived functional similarities of questions in Mandarin and French, and then examines how this similarity is formally realised in the two languages. Accordingly, the *tertium comparationis* for this syntactic contrastive study is the similarity in semantic and pragmatic functions of questions in Mandarin and French. Semantically and pragmatically, questions seek confirmation regarding a proposition (i.e., polar questions) or request information about an entity (i.e., person, object, places, time in *wh*-questions). Therefore, the shared semantic and pragmatic characteristics of questions in Mandarin and French constitute

the third comparison for the three topics investigated in this study. More specifically, three levels of comparisons are conducted (Krzeszowski 1990: 37): first, semantic comparison of equivalent constructions (i.e., question structure), second, formal/syntactic comparison of equivalent rules (i.e., *wh*-in-situ), and third, comparison of equivalent systems (i.e., question particle), as shown in Figure 2.

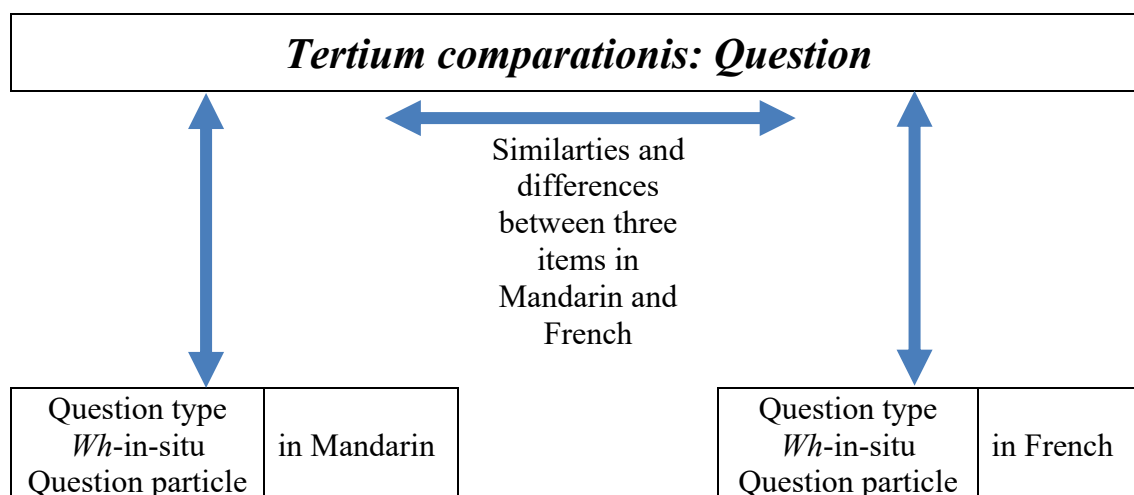


Figure 2 *Tertium comparationis* and three levels of comparisons in the current study

First, question types are distinguished universally based on their semantic properties. Polar questions involve a proposition that requires confirmation with a positive or a negative answer. *Wh*-questions seek information related to the specific question word used, while disjunctive questions typically present two options for the addressee to choose from.

Second, *wh*-in-situ represents a formal similarity between the rules governing question words in Mandarin and French. In Mandarin, this rule functions as a default structural representation, whereas in French it constitutes one of the syntactic options for forming questions.

Finally, question particles reflect syntactic and pragmatic similarities between the particle systems of the two languages, particularly those particles that co-occur with questions. Some particles alter the modality of a sentence, converting declarative sentences into questions, such as *ma* in Mandarin, *est-ce que* in French. Others serve specific pragmatic functions within discourse, linking the current question to preceding context, such as *ne* in Mandarin, *ça* in French. The use of question particles thus contributes to different levels of interrogativity and helps establish contextual coherence.

Overall, the common basis of comparison for the three topics is well-defined through the discussion above, thereby ensuring the reliability of cross-linguistic analysis. By establishing these constructs as a foundation, the research proceeds through the universal scientific steps of observational, descriptive, and explanatory adequacy. The next section presents the tool used to explore and investigate these three topics.

1.5. Methodology: Corpus linguistics

In this dissertation, in addition to the theoretical framework adopted, authentic language use is also essential to the discussion. To have access to authentic product data, instead of an introspection-based approach, electronic corpora serve as a useful tool. These corpora provide actual linguistic behaviour to such an extent an independent discipline is raised, corpus linguistics. This latter is “a whole system of methods and principles that can be applied to answer research questions related to language use and variation in a wide variety of domains of linguistic enquiry and beyond” (Paquot & Gries 2020: V). Arguably, the popularity of this methodology has been seen in various fields of linguistics, including language acquisition, pedagogy, comparative linguistics, and translation studies. However, the use of corpora has greatly diminished as a result of the criticisms voiced by generative practitioners (McEnery, Xiao & Tono 2006: 3). The latter criticize corpus of its “skewness”, namely corpus cannot adequately represent language (McEnery & Wilson 2001: 10). It was not until around the 1980s that corpus regained its territory through the development of technology (Leech 1992: 110; McEnery, Xiao & Tono 2006: 4). Yet, linguists do not agree whether corpus linguistics is a methodology or a theory (Liao & Lei 2017: 1). Herein, no detailed argumentation is discussed, but this dissertation follows the standpoints of McEnery & Wilson (2001: 2), McEnery, Xiao & Tono (2006: 7–8) and Gries (2010). That is, corpus linguistics is considered an approach or a methodology instead of a theory (but see Tognini-Bonelli 2001 regarding corpus linguistics as a full-fledged linguistic discipline), given that corpora can be a tool to be applied in different linguistic fields, for instance when dealing with various syntactic or semantic issues (McEnery & Wilson 2001: 2).

As a tool suitable for any linguistic study, the notion of ‘corpus’ should be clearly defined. A corpus can be defined as a collection of natural language texts for language research (Sinclair 1991: 171; Liao & Lei 2017: 5). During the procedure of building a

corpus, several steps are required based on the purpose of the research, such as designing a corpus, collecting texts, encoding the corpus (McEnery & Hardie 2012: 241). However, there are fundamental issues that need to be taken into consideration when starting to construct a corpus, such as the representativeness of samplings and corpus size (Ädel 2020: 4–8; Reppen 2022: 13–14), machine-readable form, and a standard reference (McEnery & Wilson 2001: 29).

In terms of representativeness, every linguist hopes to fully access the languages investigated. However, in practice, that is rarely possible. To have at least plausible reasoning on one linguistic topic, we can instead have a sample of language variety with less bias but more balance, which becomes a corpus with a full range of variability in a population (Biber 1994).

Following the issue of representativeness of a corpus, researchers often question how large a corpus must be to achieve representativeness. It seems that corpus linguists hold different positions regarding size issues, as noted by Pastor & Seghiri (2010: 120). Two decades ago, McEnery & Wilson (2001: 30) considered corpus with finite size; after a decade, there exist open-ended corpora, meaning that the size of corpora extends as long as the new data are added, such as the example of *COMPARA* corpus (Frankenberg-Garcia & Santos 2014). It seems that currently, the problem is no longer concerned with the size of corpora; instead, the capacity of storage on the cloud becomes important (O’Keeffe & McCarthy 2022: 1). Nowadays, thanks to technological advances, the size issue becomes less acute and large amounts of data are relatively easy to obtain. However, as Sinclair (2004a) points out, regardless of the size of the corpus, it is rather crucial to have clear objectives in the first place. In addition, the target should be easily extracted automatically.

In pursuit of this objective, data must first be made accessible to machines, namely digitalization. The advantage brought by machine-readable form is hard to ignore, given that machine-readable data allows searching for target forms, automatic or semi-automatic extraction and analysis. Researchers can benefit from saving time when carrying out analyses of large amounts of data. Besides, machine-readable corpora can be easily enriched with additional information in the form of annotations or metadata (McEnery & Wilson 2001: 32).

The last aspect is relatively less essential compared to the other three. Once a corpus is built by following the three abovementioned criteria, its availability to other researchers should also be taken into account. To be used in other studies, a corpus must clearly define the representativeness of the languages that it contains. In other words, variations from different studies based on the same corpus could then be attributed to their differences in assumptions or methodologies, but not the corpus itself. Overall, these aspects are essential in that researchers should think through them before constructing their corpus for a particular purpose of research. However, it should be noted that certain conditions need to be fulfilled to be labelled as corpus-based studies. The latter meets the criteria of employing corpus-based methodology, defined as utilizing available electronic corpora and corpus-linguistic tools for analysis (Granger & Lefer 2022).

For the last few decades, given the efficiency brought by corpora, the latter has been increasingly used in various linguistic fields, such as lexicography, second language teaching and learning, translation studies, quantitative linguistics (Liao & Lei 2017: 1–2). In the studies revealing the panorama of corpus-related publications in the past few decades, they have observed how corpus as a methodology is used in linguistic fields. Among them, certain investigate only one specific linguistic domain, such as second language teaching (see Römer 2011), or focus on the publications from one specific area, such as China (see Feng 2006), while only a few are bibliometric analyses among which Liao & Lei (2017) analyse the bibliometric information of corpus-related studies published across countries from 2000 to 2015. Apart from the overall increasing tendency to use corpus as the methodology, they also find that some fields have started to apply such a tool in their research, such as pragmatics, language testing, and psycholinguistics (Liao & Lei 2017: 4–5). Accordingly, a corpus can be designed for different research purposes and serves as a useful tool for that specific research field, which is probably not traditionally related to a corpus. With that being said, given that the focus of this dissertation is concerned with language contrasts, the following paragraphs concentrate primarily on the corpus that is used in this direction of studies.

When studies focus on only certain topics within a language, we can resort to a monolingual corpus, such as COCA (*The Corpus of Contemporary American English*),¹

¹ <https://www.english-corpora.org/coca/>

BNC (*British National Corpus*),² *Academia Sinica Balanced Corpus of Modern Chinese*,³ *Frantext*.⁴ On the other hand, multilingual corpora are needed in cross-linguistic studies, as in the present dissertation, where Mandarin and French are compared. To address topics concerning two or more languages, a multilingual corpus is often used in such cases. Hence, the compilation of parallel corpora (also called *translation corpora* in Johansson (2007: 9) is often envisioned in cross-linguistic studies.⁵ The term refers to collections of source texts and their translations. The latter can be automatically aligned at the level of word or sentence with a high degree of accuracy (McEnery & Wilson 2001: 70). As Lefer (2020: 259) points out, the majority of parallel corpora employed in cross-linguistic studies share two key characteristics. One is that the source and target languages are clearly identified. As concerns the number of source and target languages involved, there are bilingual and multilingual parallel corpora. The other characteristic is that only direct translations are included, indicating that translated texts are not produced based on a pivot language, such as English. In the so-called directional parallel corpus, two types are distinguished based on the direction of translations included. In other words, monodirectional parallel corpus comprises only one translation direction (e.g., from Mandarin to French); whereas bidirectional parallel corpus has both translation directions (e.g., from Mandarin to French and French to Mandarin).

Accordingly, these two characteristics ensure not only that cross-linguistic differences are captured through good correspondences between originals and translations, but also that such differences result from the nature of two compared languages without being affected by pivot languages. Despite this advantage, Johansson (2007: 10) notes that when contrastive studies rely exclusively on parallel corpus data, it is hard to forgo the fact that translation effects may play a role in findings. As translating is a process of connecting two cultures by languages, a tension has been remarked between source and translated texts, namely, from the faithfulness of source languages to the naturalness of target languages (Johansson 2007: 33). In the presence of such tension, unexpected results might emerge in studies based on bidirectional parallel corpora. For

² <https://www.english-corpora.org/bnc/>

³ <https://asbc.iis.sinica.edu.tw/>

⁴ <https://www.frantext.fr/>

⁵ The term might be used differently across studies (Lefer 2020: 257–258), but this issue lies beyond the scope of the present study; see McEnery & Xiao (2007) for further discussion.

instance, an overuse or underuse frequency of certain words and constructions might result from such tension (see the example of tag questions in Johansson 2007: 32). Therefore, he suggests that complementing parallel corpus data with data drawn from bilingual comparable corpora would be helpful to mitigate these effects. A bilingual comparable corpus is a collection of original texts of two or more languages, compiled using the same criteria, such as genres, publication time. One of the pioneers, the *English-Norwegian Parallel Corpus* (ENPC), was built in the 1990s with the aim of combining parallel and comparable corpora, as shown in the model in Figure 3.

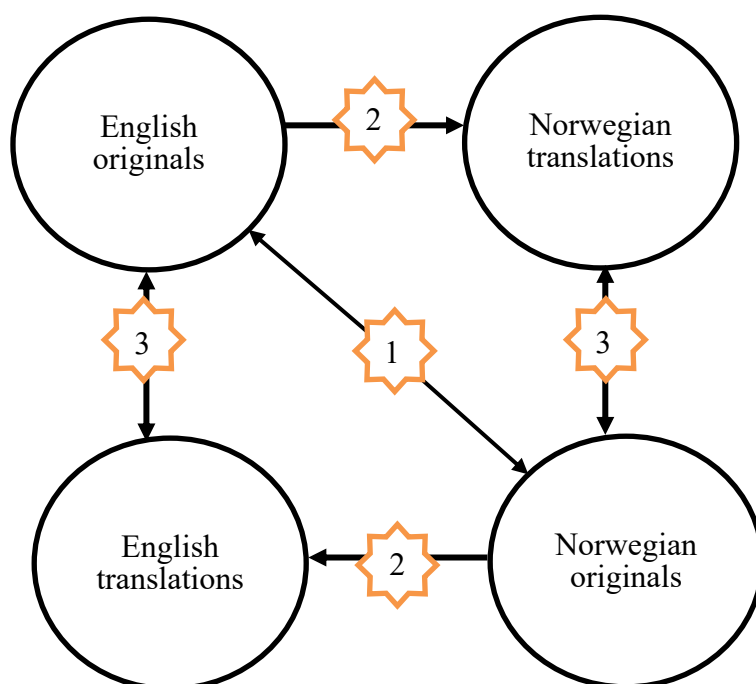


Figure 3 The model of ENPC (adapted from (Johansson 2007: 11))

This model consists of a bilingual comparable corpus (which includes English and Norwegian originals linked by number 1) and a bidirectional parallel corpus (English originals to Norwegian translations and Norwegian originals to English translations; in Figure 3, these are linked with numbers 2, respectively). In addition, other possible combinations exist, such as English originals with English translations, and the same for Norwegian. These are linked with number 3. These latter combinations are more relevant to translation studies. By using this combined model, researchers can observe two language systems from a macro-perspective. As such combined corpora have become the main trend in corpus-based contrastive studies (Granger & Lefer 2020a: 18), and this dissertation will apply this model.

In the model, which includes both comparable and parallel corpora, the sequence of observation depends on the topics and objectives of the study. Generally, there are two approaches (Granger & Lefer 2020b: 171): either starting with comparable and then parallel corpus or the other way around. For the topics dealt with in contrastive analyses, it is usually suggested to start with comparable corpora and then examine parallel corpora, as practiced by Ebeling (2000) for presentative constructions in English and Norwegian (Altenberg 2001: 150; Johansson 2007: 33), in addition to other studies in Granger & Lefer (2020b: 169–170). In the case of Ebeling (2000), the author first describes and highlights the contrast of two different presentative constructions (i.e., full constructions marked with *there* vs. bare constructions without being marked with a dummy subject) in English and Norwegian separately using a bilingual comparable corpus. Then, the two presentative constructions are compared cross-linguistically using a bidirectional parallel corpus. As this model is widely used in contrastive studies and its usefulness is recognised by corpus linguists, I will therefore observe my target of research based on the corpora of this model. Furthermore, I will follow in this dissertation the traditional procedure of contrastive analysis, namely, first comparable then parallel. Chapter 3 will be fully dedicated to the detailed methodology, including corpus design, used in this dissertation.

Before closing this section, I wish to present some final remarks regarding criticisms of using corpus data and some reflections on the choice of a corpus methodology in contrast with a formal theoretical framework.

Leech (1992: 107) underlines the mismatch between corpus linguistics and the Chomskyan paradigm. Chomsky rejected using corpus data in that a corpus can never capture the whole picture of a language, which can be produced with an infinite number of structures (McEnery & Wilson 2001: 9–10). Yet these arguments should not prevent us from analysing a language in the light of the data. It suffices that we acknowledge that the corpora employed in any research are likely to show merely a small part of certain linguistic phenomena in languages. Accordingly, in this dissertation, two sorts of corpora are built to compare questions in Mandarin and French. In addition, I endeavour to control their genres and publication times to ensure a certain degree of comparability in my corpora. Nevertheless, I must acknowledge that it will not be possible to generalise the corpus-based cross-linguistic findings to the whole language system of the two languages.

In addition to the question of the representativeness of corpus data, there is also the divergence in terms of the nature of the data used in analyses. In the formal theoretical framework, only the norms of languages, which follow rules and conventions, are analysed; whereas corpora comprise instances that may deviate from the norms. Such a distinction is made clear in the terms *competence* vs. *performance* (Chomsky 2015: 2). Arguably, *competence* is the knowledge of one's language and is engraved in intuition, whereas *performance* represents actual linguistic behaviour and is accessed through observation. This mismatch also seems obvious in this present dissertation regarding my theoretical framework and methodology. That is, the topics of questions are not only discussed within the formal theory but they are also investigated in the comparable and parallel corpora. However, such a combination might no longer be a mismatch in the linguistic field in that some studies applied with generative grammar have started using corpus, such as in second language acquisition (Lozano 2020: 219–220). Consequently, by presenting two different standpoints in one dissertation, we can see how the same linguistic phenomenon (i.e., questions) can be treated from an interdisciplinary perspective, namely from theoretical analyses and corpus observations. In other words, this dissertation will highlight the gap between theory and authentic language use and aim to converge the two seemingly conflicting perspectives for the topics of questions in both Mandarin and French. Therefore, by combining efforts, I expect that the outcomes “should be more easily understandable and satisfactory to all sides” (Rothman & Slabakova 2018: 436).

1.6. Outline of the dissertation

This dissertation is composed of six chapters. It can be grouped into two main subparts. The first part (Chapter 2) concerns the theoretical perspective on questions. The second part (Chapter 4 and 5) pertains to the examination of questions in real language use through a bidirectional parallel corpus from three approaches: bilingual comparable, bidirectional parallel, and monolingual comparable.

Chapter 2 lays out the theoretical dimensions of this research. In addition to the issue of terminologies (i.e., *interrogative* vs. *question*), I will cross-linguistically review the construct of questions including question types and marking strategies. Then, the focus of sections shifts to the contrast of the two languages studied in this dissertation,

Mandarin and French. Particular attention is drawn to a theoretical review of the three topics: question taxonomies, *wh*-in-situ phenomena, and question particles. This review led to contrastive insights from a theoretical perspective and set the *tertium comparationis* for questions in both languages.

Chapter 3 is concerned with the methodology used in this dissertation. It starts by providing a review of the literature on corpus-based contrastive studies regarding the topics of questions. Given that this study employs a self-built corpus, the main purpose of this chapter is to provide the rationale for the selection of novels in the corpus design and to demonstrate the procedure of data extraction and coding in detail.

Chapter 4 presents the results and engages in discussion related to the three topics of concern in the two languages, which have been discussed in Chapter 2. Each topic will be addressed using the bilingual comparable data. Accordingly, cross-linguistic differences and similarities of questions are revealed based on the observations and analyses of corpora. Chapter 5 adopts both bidirectional parallel and monolingual comparable perspectives, aiming to examine how cross-linguistic findings are reflected in translations and to identify translation effects by comparing translated texts with their non-translated counterparts. In addition, a comprehensive statistical regression model is employed to reassess the findings by incorporating multiple predictors, including register, language, and translation status.

Finally, Chapter 6 presents the general conclusion of this dissertation. It summarises the key findings of the three topics examined through the three analytical approaches. In addition, it highlights the contribution to contrastive and corpus linguistics, as well as to the theory of questions. The chapter also addresses the limitations of the research and proposes promising avenues for future study. The chapter concludes with implications for language pedagogy and learning, aiming to bridge theoretical insights and corpus-based findings and their potential practical applications.

Chapter 2 Formal syntax and semantics of questions

In this chapter, Section 2.1 concerns the use of two terms, *question* and *interrogative*. Both terms are employed in different linguistic studies. Yet linguists have used them in very different ways. Therefore, in this opening section, I will be clear about the stand that I am taking in this dissertation. Having clarified this issue, I will present an overview of the typologies of questions, including strategies of questions used in different languages. In Section 2.2, I will first narrow down to the general construct of questions in the two languages focused on in this dissertation, namely Mandarin and French. The three following sections offer a theoretical review and discussions of the three main topics: question taxonomies in Section 2.3, *wh*-in-situ phenomena in Section 2.4, and question particles in Section 2.5. Finally, Section 2.6 summarises and underlines the key points made in Chapter 2.

2.1. Overview of the construct of questions

This section is devoted to reviewing research that is pertinent to questions. The section begins by clearly illustrating the terminological issue between *question* and *interrogative*. In addition, I explain the principles of the use of these two terms in this dissertation. After addressing the terms, I will shift to the typological research on question types, followed by a discussion of the cross-linguistic marking strategies used to form these question types.

2.1.1 Question vs. Interrogative

It has been argued by some that the emergence of language in humans is partly driven by the need for social interaction, such as cooperation (Seyfarth, Cheney & Bergman 2005: 266). In the process of achieving social interaction, different speech acts are applied interchangeably in communication. These acts seem to be widely shared in human languages. From a functional standpoint, five categories of speech acts are traditionally differentiated (Austin 1962): representatives (or assertives) (i.e., to commit the speaker to the truth of the expressed proposition), directives (i.e., they are attempts by the speaker to get the hearer to do something and questions are a species of this speech act), commissives (i.e., to commit the speaker to some future course of action),

expressives (i.e., the speaker is trying to convey the truth of the expressed proposition which is presupposed), and declarations (i.e., the speaker brings about a state of affairs specified in the propositional content by saying in effect) (Searle 1976). This classification defined by speech functions underlines the intention or purpose of the utterance of a speaker toward the addressee. On the other hand, from the formal perspective, three basic sentence types are recognised: declaratives, interrogatives, and imperatives (Velupillai 2012: 345). In addition to these three types, exclamatory sentences are sometimes added (Quirk et al. 1985: 803). Accordingly, these two different ways of distinguishing speech acts are somewhat correlated (Siemund 2001: 1011) in that declaratives represent statements and interrogatives denote questioning acts. However, König & Siemund (2007) point out that it is too simplistic to merely link one form with one function. Interrogative sentences, for example, can perform other functions, such as suggestions in *Why don't you buy some stocks?* in addition to questioning (König & Siemund 2007: 283).

Of relevance to this dissertation, I am primarily concerned with interrogative sentence that have a true questioning function, rather than an indirect function, such as suggesting. However, the terms *interrogative* or *question* occur so frequently in many studies of questions that some of them have underlined the importance of properly distinguishing between these two terms (see Siemund 2017: 17–18). Huddleston (1994) has made a clear distinction between *question* and *interrogative*. The term *interrogative* is one of the clausal types, and it contrasts with declarative and imperative in terms of morphosyntactic structures. For instance, declarative and interrogative clauses can be distinguished from their subject-verb order (i.e., inversion). In addition, these clause types are mutually exclusive since it is not possible to have *interrogative* and *declarative* in the same clause. Therefore, the term *interrogative* is more relevant to the syntactic structure of a clause.

Concerning the syntactic structure of interrogative clauses, two types of interrogatives are distinguished: *closed interrogative* and *open interrogative*. A closed interrogative represents an interrogative clause marked by subject-auxiliary inversion, whereas an open interrogative has interrogative words. In addition, such a syntactic contrast of *open* and *closed* can also be used to define the term *question*, which is based on distinguishing a set of answers from a semantic perspective (Huddleston 1994: 419).

In addition to a set of answers, questions can be defined semantically in a broad sense as “a sentence type that seeks information from someone being treated as knowing by someone who is unknowing” (Stivers 2010: 2776). In other words, depending on the content of questions and a set of answers, questions can generally be subdivided into three types. First, *polar questions* (also known as *yes-no* questions) seek confirmation and define answers to *yes* or *no*, namely with a closed set of answers. Secondly, *alternative questions* (also called disjunctive questions) seek the information within the choice provided, yet this type of question falls under the category of closed interrogative, given that the choices given seem limited. However, Her, Che & Bodomo (2022: 8) argue that it is misleading to consider disjunctive questions having a closed set of answers. Finally, *constituent questions* (also known as content questions or *wh*-questions) contain a question word and expect information related to that specific question word, which is a set of open-ended answers.

Table 2 illustrates the distinction between interrogatives and questions, in addition to the examples from Huddleston (1994: 416, 419).

Table 2 Distinction between *interrogative* and *question*

Linguistic level	Formal	Semantic	
Term	<i>Interrogative</i>	<i>Question</i>	Examples
Sub-categories	Closed	Polar	<i>Are you ready?</i>
		Alternative	<i>Is it a boy or a girl?</i>
	Open	Constituent	<i>Whose hat is this?</i>

Thus, regardless of the classifications of speech acts and differences between the two terms *interrogative* and *question*, both call our attention to the distinction between form and function, or grammar and speech acts. In this dissertation, I will primarily use *question* as the cover term for form and function. As for the sub-categories of questions, I will use the terms *polar questions*, *disjunctive questions*, and *constituent questions* in the related discussion below.

2.1.2 Typology of questions and question types

Over the past few decades, a number of typological studies of questions have been carried out. One of the most representative pioneering typology studies of questions

appears to be Ultan (1978). In this latter study, questions are systematically examined in a sample of 79 languages. The author focuses on the features of different question types and provides the statistical data on the distribution of each feature across the 79 languages. In addition to the classic study done by Ultan (1978), Hölzl (2018: 55) provides a list about forty works representing a series of important studies concerning questions in typology from the 1990s onwards. These studies cover general to more specific topics related to questions, such as the grammaticalization of question particles in polar questions (Bencini 2003), turn-taking systems (Stivers et al. 2009), or question marking in alternative questions (Mauri 2008). Besides their different scopes of research, the number of languages involved in the studies is also varied, ranging from less than ten languages (Huang et al. 1999, as cited in Hölzl 2018: 65) to nearly two thousand (Idiatov 2007). Among these studies, some are restricted to linguistic areas, such as European languages, Australian languages, African languages, or languages in Taiwan, while others aim for a general cross-linguistic overview of questions without focusing on certain languages. Regardless of the purpose and the number of languages being covered, the first and most fundamental issue is to distinguish question types. Identifying question types at the outset provides a clearer and broader scope for determining the focus of studies. Only after deciding the target, it is possible to further observe the internal syntactic structures of these different question types. Therefore, this section focuses on question types based on an overview of typological works that foster a more general discussion across languages. This cross-linguistic overview will further serve as a basis for me to build upon when I later narrow the discussion to the two languages studied in this dissertation.

From the discussion of the two terms in Section 2.1.1, it is possible to formally categorise two types of questions and semantically three types. Such a divergence of question types is also found in the typology works which I will present in the following paragraphs.

Generally, questions are categorised into two basic types: polar questions and constituent questions (Greenberg 1963: 63; Schulze 2007: 255). However, three types of questions are more frequently cited (Ultan 1978: 43, Sadock & Zwicky 1985: 179, Siemund 2001, König & Siemund 2007, Stivers & Enfield 2010, Dixon 2012): polar questions as in (2.1), constituent questions as in (2.2), and disjunctive questions as in (2.3).

(2.1) *Does a platypus lay eggs?*

(2.2) *What is a platypus?*

(2.3) *Is a platypus a mammal or a bird?* (Siemund 2001: 1010)

These question types are distinguished in terms of the answers expected or the purpose of questioning. On the one hand, polar questions expect a confirmation with either *yes* or *no* as an answer. On the other hand, constituent and disjunctive questions look for concrete information with a set of possible answers, which can be at least two choices or any possible answers to that specific question word.

Despite the fact that disjunctive questions seem to be independent of the other two question types, it is noted that they are frequently subsumed in polar questions (Siemund 2001: 1012). Therefore, it turns out that two question types (i.e., polar and constituent) are the basics of typology. This can be attributed to the fact that disjunctive questions appear less varied across languages in terms of their forms (König & Siemund 2007: 291–292). However, Hölzl (2018: 65) holds a different view in that disjunctive questions have not only diverse forms, but these are distinctive compared to the other two question types. The forms of disjunctive questions will be discussed in Section 2.1.3.3, which deals with the marking strategies of disjunctive questions.

In addition to the controversy surrounding the classification of disjunctive questions as a separate question type, debate has arisen over the attribution of disjunctive questions. Most studies label them as polar questions. Yet, as pointed out by (König & Siemund 2007: 291–292), even though polar and disjunctive questions have a lot in common, they are different in terms of the answers expected. On the other hand, disjunctive questions are akin to constituent questions in terms of the nature of the information they are seeking. Accordingly, some might argue that disjunctive questions are a type of constituent question. This issue of how disjunctive questions are categorised will be further addressed in Section 2.3.1 when discussing the dichotomy of questions.

While two and three question types appear to be in consensus, or at least recognised in most typological works, Hölzl (2016; 2018) suggests a new typology of questions in order to eventually fully cover questions represented across languages. This new typology of questions includes focus questions and tag questions in addition to the other three question types mentioned above.

Focus questions are similar in form to polar questions. In English, they rely on phonological differences as a marker of focus. If we compare the following two questions in (2.4), the word in bold indicates the phonological stress, showing a narrow focus on the constituent *you* in (2.4b). In other words, the focus of this question is restricted to this constituent. Accordingly, the question in (2.4b) could be reformulated as *Is that you who wants coffee?*.

(2.4) a. *Do you want coffee?*

b. *Do **you** want coffee?* (Hölzl 2018: 57)

Besides phonological marking, some languages use morphological markers alongside the constituents in focus, such as in Yuwan (Japonic language), Slavey (Na-Dene language), and other languages mentioned in Dixon (2012: 395–396).

Tag questions shown in examples (2.5) and (2.6) are short questions attached to declarative sentences (Mithun 2012). The form of tag questions can be an inverse polarity compared to the declarative sentence it attaches to, as in the example (2.5). Also, they can be invariant, as in the example (2.6).

(2.5) *It's on the left, **isn't it**?* (Mithun 2012: 2165)

(2.6) *You want coffee, **right**?* (Hölzl 2018: 60)

However, its status is always problematic in that it has been argued for a separate question type (Hölzl 2016), while others consider it as one marking strategy for polar questions (Ultan 1978; Sadock & Zwicky 1985; Siemund 2001). It seems to me that the identity issue of tag questions is much easier to deal with than the one of disjunctive questions. Tag questions are essentially a form of polar questions, with the difference being that tag questions have explicitly shown the bias toward the polarity of the answers.

From the review above, the number of question types may vary from two to five based on different perspectives. However, I think in typology, there are so far two basic types of questions, that is, polar and constituent. This is by no means to deny the existence of focus questions and tag questions. Yet, they should be considered as sub-types or one of the marking strategies of polar questions rather than a separate question type because they are still polar questions in nature. The effect of focus comes from a pragmatic need. Similarly, tag questions (mostly) appended to a declarative clause are often biased regarding the answer expected (Siemund 2001: 1015), as do polar questions. Therefore,

their labels denote pragmatic functions rather than the syntactic or semantic nature of questions.

As illustrated by focus questions and tag questions, different strategies for marking questions may exist cross-linguistically. The following section is devoted to the marking strategies for questions. More specifically, I will examine how questions can be distinguished from declarative sentences through the use of different linguistic devices, ranging from phonological to syntactic. Some of these marking strategies are shared across question types, while some are exclusive to a particular type of questions.

2.1.3 Typology of questions and marking strategies for questions

Languages seem to have at least two question types as discussed in Section 2.1.2. Yet, the ways of marking questions can vary from one language to another. Overall, three linguistic levels involve in marking strategies for questions: prosodic, morphological, and syntactic levels (Luo 2013: 1). According to the typological literature (Ultan 1978; Siemund 2001; König & Siemund 2007; Schulze 2007; Dixon 2012; Dryer 2013b; Hölzl 2018), prosodic marking comprises intonation and pitch. Morphological marking includes particles, verbal morphology (including affixes and clitics), or tags. Regarding syntactic marking, one is disjunctive constructions, which seem to be dedicated to disjunctive questions. The other is different word orders that are usually involved to distinguish questions from declaratives. Table 3 summarises the different marking strategies that are frequently mentioned in the literature. Yet, it is important to note that these different strategies are not necessarily shared cross-linguistically.

Table 3 Marking strategies for questions

Prosodic	Morphological	Syntactic
	Particles	
Intonation	Verbal morphology	Disjunctive construction
Pitch	Tags	Word order change
	Question words	

Despite the fact that these marking strategies are frequently mentioned in the important typological works of questions (Greenberg 1963; Ultan 1978; Sadock & Zwicky 1985; Siemund 2001; König & Siemund 2007; Velupillai 2012; Hölzl 2018), there are other marking strategies that are exclusively used in one particular language,

such as a superficially pronominal question words used in German polar questions (Schulze 2007: 255). In the following discussion, I will only focus on the marking strategies presented in Table 3 and start with the marking strategies for polar questions. In fact, most of them can be used in any question type, except for question words, tags and disjunctive structure. Question words are reserved for constituent questions. Tags are not attested in constituent questions (Siemund 2001: 1019). Disjunctive structures are distinctive to disjunctive questions.

2.1.3.1 Marking strategies for polar questions

Dryer (2013b; see Table 4) describes the distributions of marking strategies that are employed for polar questions across languages (with 955 languages in total) based on a survey of typological studies on questions. Note that languages might use more than one strategy to mark their polar questions (Sadock & Zwicky 1985: 181), such as English uses intonation and word orders; Pirahã (Mura; Brazil) uses both particles and verbal morphology (Dryer 2013b).

Table 4 Cross-linguistic marking strategies for polar questions

Marking Strategies	Number of languages
Particles	585
Intonation only	173
Verbal morphology	164
Word order	13

According to this typological survey, particles seem to be the most frequently used marking strategy in polar questions. And indeed, if intonation is the only possible marking strategy, the cross-linguistic distribution comes in the second place. However, as languages can resort to multiple strategies simultaneously, if counting all the possible strategies that one language can use to mark polar questions, **intonation** would be the most widespread strategy across languages (Ultan 1978: 45; Siemund 2001).

Most languages use rising intonation at the end of the sentence (Sadock & Zwicky 1985: 181; Dixon 2012: 394). Such a prevalence also corroborates the Universal Principle 8 proposed by Greenberg, namely that the distinctive phonological feature of polar questions is often found at the end of the sentence rather than at the beginning (Greenberg 1963: 63). In addition, some minor differences are found among the languages using

rising intonation at the end of sentences. Ultan (1978: 46) shows eight general types of terminal contours based on their placements and degrees. For instance, Iraqi Arabic has a slight terminal rise with length, Vietnamese has a higher pitch at the ultima, while Hebrew has a rise toward the last stressed vowel (see also Dixon 2012: 394–395 for other examples). Despite the various minor differences, the rising intonation still characterises polar questions by its connotation of uncertainty, indecision, and hesitation (König & Siemund 2007: 292).

Nevertheless, the terminal rising contour of polar questions is not shared across all languages. For instance, Fanti (Niger-Congo) and Grebo (Niger Congo) employ a falling contour in questions (König & Siemund 2007: 292) and Sesotho (Bantu) uses a lowered pitch (Paroz 1946: 208, as cited in Dryer 2013b). Therefore, the absolute correlation between rising intonation and polar questions has been challenged (cf. Huddleston 1994: 428–429).

On the final point about intonation, Dryer (2013b) notes that at least 173 languages use intonation alone to distinguish questions from declarative sentences, such as colloquial Italian (Maiden et al. 2014: 147). In other words, in those languages, intonation is the only way to transform declarative sentences into questions, in contrast to other languages having multiple marking strategies for polar questions at their disposal, such as particles and word order change (Sadock & Zwicky 1985: 181).

The second most frequently used marking strategy for polar questions is adding a **particle** (Ultan 1978: 51; Siemund 2001: 1013). By adding such an element, declarative sentences are turned into questions, such as 嗎 *ma* in examples (2.7) in Mandarin.

(2.7) a. Declarative sentence

你	是	他	哥哥。
<i>nǐ</i>	<i>shì</i>	<i>tā</i>	<i>gēge</i>
2SG	COP	3SG	brother

‘You are his elder brother.’

b. Question

你	是	他	哥哥	嗎？
<i>nǐ</i>	<i>shì</i>	<i>tā</i>	<i>gēge</i>	<i>ma?</i>
2SG	COP	3SG	brother	PTCL

‘Are you his elder brother?’

(Luo 2013: 6)

One language can even have multiple particles. For example, Korean uses different particles to show different levels of formality, such as formal, polite, intimate and plain

(Siemund 2001: 1013; Pak 2008: 123–124). In Mandarin, different particles are also allowed to occur in polar questions, as shown by 啊 *a* and 吧 *ba* in example (2.8). But only the particle 吧 *ba* functions similarly with 嗎 *ma* as an indicator of polar questions. Instead of denoting a question, 啊 *a* has the distinctive pragmatic function of reducing the potential face-threatening effect (Fang & Hengeveld 2022: 897). This issue will be addressed in more detail in Section 2.5.

- (2.8) 你 是 他 哥哥 啊/吧 ?
 nǐ shì tā gēge a/ba?
 2SG COP 3SG brother PTCL
 ‘Are you his elder brother?’ (Luo 2013: 2)

It is also worth noting that the position of particles is often discussed in typological works. Their positions somehow largely relate to the word order of propositions. As formulated in Universal Principle 9 (Greenberg 1963: 64), particles are found in the final position when the languages are verb-final, whereas particles tend to be initial in verb-initial languages. However, this generalization seems not to be completely validated across languages. For instance, Persian is a verb-final language, but particles are found in the initial position (König & Siemund 2007: 294).

Following intonation and particles, **verbal morphology** comes in third place. This category can represent a wide range of devices for polar questions, such as affixes, clitics, and verb reduplication (Luo 2013: 6). Even though it would be inaccurate to classify these devices in the same category (Utan 1978: 52), in the following discussion, they will be included under the same label as verbal morphology. In addition to facilitating the review, another reason is that the languages explored in this dissertation do not use such marking strategies.

Compared to the use of particles, **affixes** are relatively rare. The well-known cases come from polysynthetic languages, such as West Greenlandic and Eskimo. They use suffixes to indicate the transition from declarative sentences to questions, as shown in (2.9).

- (2.9) West Greenlandic

- a. *neri-vutit*.
 eat-IND.2SG.PST
 ‘you ate’

b. *neri-vit*

eat-INT.2SG.PST

‘Did you eat?’

(König & Siemund 2007: 299)

Unlike affixes, question **clitics** are freely combined with any constituent in a sentence. Generally, clitics are attached to the verb as the Latin *ne* in examples (2.10). Likewise, clitic *ne* can be attached to other constituents of different word classes, such as nouns in (2.10b) and adjectives in (2.10c). This marking strategy can be related to the function of focus question because the constituent attached to the clitic becomes the focus of the question.

(2.10) Latin

a. *Est=ne puer bonus?*
COP=CLIT boy good
‘Is the boy good?’

b. *Puer=ne bonus est?*
boy=CLIT good COP
‘Is it the boy who is good?’

c. *Bonus=ne puer est?*
good=CLIT boy COP
‘Is it good that the boy is?’

(Sadock & Zwicky 1985: 182)

Finally, **verb-reduplication** is common in the Tibeto-Burman languages, such as Yi language: *la*³³ means ‘come’; whereas *la*⁴⁴*la*³³ means ‘Come?’ (Luo 2013: 7).

Despite the fact that it is not always easy to distinguish between particles and verb morphology, Dryer (2013b) observes that 15 languages use both marking strategies in the same question, provided that these two strategies can be differentiated, such as in the language Pirahã (Mura).

Regarding the syntactic level, the **word order change** (i.e., inversion) seems to be even rarer when compared to other marking strategies. Such a marking strategy is said to be restricted to Germanic, Romance, Slavic languages, and Finno-Ugric languages (Ultan 1978; Schulze 2007: 257). Also, it is present in two Austronesian languages (i.e., Palauan and Manggarai) and two languages in South America (i.e., Warekena and Hup) (Dryer 2013b).

In general, this marking strategy is concerned with the position of the verb, as evidenced in (2.11) by the different positions of the German verb *trinkt* ‘drinks’ between the declarative sentence in (2.11a) and the question in (2.11b). Typically, the finite verb

moves to the initial position of the sentence, while in English, only auxiliary verbs and modals undergo displacement. Without both elements, *do*-support is otherwise needed in English, as in (2.12).

(2.11) German

a. *Der Lehrer trink-t das Wasser.*
 ART teacher drink-3SG ART water
 ‘The teacher drinks the water.’ (Dryer 2013b)

b. *Trink-t der Lehrer das Wasser ?*
 drink-3SG ART teacher ART water
 ‘Is the teacher drinking the water?’ (Dryer 2013b)

(2.12)

- a. *John phoned me yesterday.*
 b. *Did John phone you yesterday?*

Given that inversion is to switch the order between subject and verb, or more precisely, to have the verb precede the subject, this marking strategy can only occur in languages with SOV and SVO word orders.

In addition to these marking strategies surveyed in Dryer (2013b), no relevant statistical data is proposed regarding tags, yet this latter has been widely discussed in the literature as a marking strategy (see also Section 2.1.2). As a marking device, **tags** are added to another sentence. Hence, tags are often found after a declarative sentence (also called an anchor in (Axelsson 2011: 794), as shown in example (2.13).

(2.13) *It is snowing, isn't it?*
 anchor question tag

The tag in example (2.13) consists of a repeated subject and an inversion. This kind of tag is also labelled as variant tag because it changes syntactically depending on the anchor, including tense, person, number, and gender (Axelsson 2011: 795). However, this sort of tag is rarely found in languages other than English (Sadock & Zwicky 1985: 183). Most languages have a fixed form of tag (Dixon 2012: 392), also known as invariant tag. They attach *yes* or *no/not* to anchors. In fact, English also has invariant tags, such as: *It's snowing, no?*.

Regardless of the form of tags, the polarity of tags is often the opposite of its anchor. As shown in example (2.13), if the anchor is positive, the tag is negative. Such an anchor-tag pairing is used more frequently than the one of a negative anchor appended with a positive

tag (e.g., *You don't like this actor, do you?*) (Sadock & Zwicky 1985: 183; König & Siemund 2007: 297). In addition to polarity-based tags, other tags are distinguished, such as alternative tags (e.g., 是不是 *shì bú shì* 'is not is' in Mandarin), disjunctive tags (e.g., *oder* 'or' in German), pronominal tags (e.g., *was* 'what' in German), and interjectional tags (e.g., *huh*). (cf. Moravcsik 1971 and Axelsson 2011).

Finally, it has been noted that tags and particles are closely related in both meaning and distribution (Siemund 2001: 1015). However, tags are different from particles in these two aspects. As far as meaning is concerned, tags are biased towards acquiring the expected answers from the addressee. As for the distribution, tags do not seem to have any association with word order, they are more independent compared to particles (König & Siemund 2007: 297).

On top of these more widely known marking strategies for polar questions, some languages present questions without adding any phonological, morphological, or syntactic markers. For example, instead of adding markers, four languages in Dryer's (2013b) survey remove the declarative morpheme to obtain polar questions. More surprisingly, Mixtec (Chalcatongo) is a language without any marking strategy for polar questions. In other words, its polar question is identical to its declarative sentence. Likewise, Yéli Dnye, a language spoken on Rossel Island, shows no distinction between the two clausal types (Dryer 2013b).

2.1.3.2 Marking strategies for constituent questions

Compared to the number of studies devoted to the marking strategies of polar questions, work dedicated to constituent questions seems to be relatively scarce. There might be two reasons for this. One is that the majority of marking strategies applied in polar questions are assumed to be shared with constituent questions, except tags (König & Siemund 2007: 299). Another reason is that marking strategies are often optional in constituent questions (Siemund 2001: 1018). Although it may seem overly simplistic to conflate the marking strategies of two question types (Siemund 2013: 249), in reality, the marking strategy for constituent questions does have similarities to those of polar questions.

According to Ultan (1978), only few languages mark their constituent questions with **intonation**. But many of them use the same intonation pattern for both polar and

constituent questions (Dixon 2012: 389–390), such as rising intonation in Tariana (Arawak, Brazil). In addition to intonation, **suffixes** might also be the same in both question types, as *vit* attested in West Greenlandic in (2.14). This suffix is identical to the one shown in its polar question in example (2.9b).

(2.14) West Greenlandic

su-mik neri-vit?
 what-INS eat-Q.2SG
 ‘What did you eat?’

(Dixon 2012: 386)

Particles also appear in constituent questions, such as 呢 *ne* in Mandarin in example (2.15) and *ka* in Japanese in example (2.16).

(2.15) 你 要 去 哪裡 (呢)?
nǐ yào qù nǎlǐ (ne)?
 2SG want go where PTCL
 ‘Where do you want to go?’

(2.16) Japanese

Dare-ga kimasu (ka)?
 who-NOM come PTCL
 ‘Who is coming?’

Yet, the fact that particles are one of the marking strategies for constituent questions is still debatable. As König & Siemund (2007: 300) point out, given that these particles are optional, the realization of constituent questions depends on the **question words**, such as *where* or *who*. These question words can clearly convey the goal of questioning without the aid of particles. Besides, from a functional point of view, it seems redundant to use two different elements to represent the same sentence type, although this double-marking phenomenon does exist when marking polar questions in Pirahã (Mura).

Despite the fact that particles do not seem to functionally play a role in question marking, they are considered as one of the marking strategies for constituent questions in some studies (Li & Thompson 1981: 305; Cheng 1991; Huang, Y.-H. Audrey Li & Li 2009: 237). Moreover, after advancing the hypothesis of an association between particles and the position of question words (cf. Baker 1970: 207), the identity of particles in constituent questions has been raised and discussed continuously ever since (cf. Cheng 1991 and Bruening 2007), especially in Mandarin, which has a variety of particles. This

will be addressed in more detail in Section 2.5, where I will focus on the particles used for questions.

The change of word order is another marking strategy shared by both question types, suggesting that word order alone cannot reliably distinguish constituent questions from polar questions; rather, the presence of a question word remains the primary morphosyntactic difference between them. In particular, when a language employs word order change to mark constituent questions, it is likely that the same strategy is also applied to polar questions, as observed by Greenberg (1963:65). However, word order change can be interpreted from two perspectives. One concerns the positional change of the question word itself, while the other involves inversion between the subject and the verb. If this strategy refers to changing the position of question words, it encompasses broader range of languages (König & Siemund 2007: 301), since placing question words in clause-initial position is cross-linguistically common (Siemund 2001: 1019), such as in Finnish in (2.17).

(2.17) Finnish

- a. *Maija ottaa omenaa*
 Maija take.3SG apple.PAR
 ‘Maija is taking an apple.’

- b. *Mitä Maija ottaa?*
 what.PAR Maija take.3SG
 ‘What is Maija taking?’

(König & Siemund 2007: 301)

By contrast, if word order change is restricted to subject-verb inversion, the number of languages covered is considerably smaller. Only certain a small group of languages (e.g., Germanic and Romance languages) employ this type of inversion as question marking. As shown the French example in (2.18), the subject *tu* ‘you’ and the second person singular form of the verb *dis* ‘say’ are reversed.

- (2.18) *Que dis-tu?*
 what say-2SG
 ‘What do you say?’

Given the correlation in the use of word order change in constituent and polar questions, Greenberg (1963: 65) formulates Universal Principle 11, stating that inversion occurs only in languages where question words are normally clause-initial. He further specifies that inversion in polar questions is possible only if it also occurs in constituent questions.

This observation shows an implicational asymmetry of using inversion. In addition, it indicates that subject-verb inversion is not specific to one single question type, but operates at the level of question marking more generally.

2.1.3.3 Marking strategies for disjunctive questions

The third question type (i.e., disjunctive questions) has received much attention in typological studies, despite the fact that its status as an independent question type is still debatable. Before delving into this issue, I prefer to set aside their marking strategies from the other two question types because its marking strategies indeed differ from the other two question types (Hölzl 2018).

The very basic marking strategy for disjunctive questions is to adjoin two choices, as shown in Amis (Austronesian) in example (2.19): two choices, *ma-tayal* ‘work’ and *ma-fúti* ‘sleep’, are provided as two options, and they are juxtaposed one to the other.

(2.19) Amis

<i>Ma-tayal</i>	<i>kisu</i>	<i>ma-futi?</i>
AGT.FOC-work	2SG.NOM	AGT.FOC-sleep
‘Are you going to work or sleep?’		

(Huang et al. 1999: 651, as cited in Hölzl 2018: 65)

This kind of **juxtaposing** seems to be available in all languages (Mauri 2008: 30). However, some languages not only juxtapose two sentences but also attach **particles** to both sentences, as shown the example (2.20) in Korean.

(2.20) Korean

<i>Wuli-ka</i>	<i>ka-l-kka-yo?</i>	<i>salam-ul</i>	<i>ponay-l-kka-yo?</i>
1PL-NOM	go-PRS-PTCL-POL	person-ACC	send-PRS-PTCL-POL
‘Shall we go, or shall we send someone?’			

(Mauri 2008: 37)

In addition to particle, two juxtaposed phrases can be linked with the **disjunctive morpheme**, such as a connective. For example, 還是 *háishì* ‘or’ in Mandarin in example (2.21). The use of connectives in disjunctive questions also occurs in Italian, French, German, etc. (Mauri 2008: 28).

(2.21) 你 喜歡 飯 還是 麵?
nǐ xǐhuān fàn háishì miàn?
 2SG like rice or noodle
 ‘Do you like rice or noodles?’

In most cases, the form of connectives remains identical in declaratives and questions, while some languages do have two different connectives, such as in Finnish, Polish, Basque, and Marathi (Mauri 2008: 28; Hözl 2018: 65). In the following Finnish examples (2.22), there is a contrast of connective *tai* ‘or’ in declarative in (2.22a) vs. *vai* ‘or’ in questions in (2.22b).

(2.22) Finnish

a. Declarative

joskus ovi avaut-ui tai ikkuna paiskaut-ui
 sometimes door open-3SG.PRF or window become-3SG.PRF
kiinni
 closed
 ‘Sometimes a door opened or a window slammed.’

b. Question

mene-mme-kö koulu-un huomenna vai jää-mme-kö koti-in
 go-1PL-Q school-ILL tomorrow or stay-1PL-Q home-ILL
 ‘Do we go to school tomorrow or do we stay at home?’ (Mauri 2008: 27)

Another disjunctive morpheme is *or not*. This is added at the end of the question, such as example (2.23) in Amele (Trans-New Guinea).

(2.23) Amele

ini jalan ke Rawamangun atau bukan?
 this road to Rawamangun or not
 ‘Is this the road to Rawamangun, or not?’ (Dixon 2012: 399)

The English translation of example (2.23) happens to be a representation of the disjunctive question in European languages (Hözl 2018: 66). The first option has an **inversion** of the verb encoded with a rising **intonation**, while the second option of the question is linked with a connective encoded with a falling intonation. It seems to me that this is one of the reasons why disjunctive questions are often considered a sub-type of polar questions.

The examples above show two general ways of marking disjunctive questions, either with covert or overt connectives. Interestingly, another distinctive way has often been mentioned in the literature, namely, the disjunctive-negative structure (König & Siemund 2007: 297–298), also known as the **A-not-A structure** (Li & Thompson 1981: 535; Huang, Y.-H. Audrey Li & Li 2009: 244). Such a structure is well-known in Mandarin and other Asian languages, in addition to languages in Papua New Guinea (König & Siemund 2007: 297).

This structure is said to have omitted the connective and is similar to the example (2.19), in which there is a covert connective. Yet, the structure A-not-A should involve two choices, one positive and the other negative, as illustrated in example (2.24). This latter consists of a positive part of the verb 喜歡 *xǐhuān* ‘like’ along with its negative part 不喜歡 *bù xǐhuān* ‘not like’ as two choices. Besides the basic form of the A-not-A structure, Mandarin has other variants (Huang, Y.-H. Audrey Li & Li 2009: 245). A more detailed description of the A-not-A structure will be given in Section 2.2.1.

- (2.24) 他 喜歡 不 喜歡 這 本 書？
 tā xǐhuān bù xǐhuān zhè běn shū?
 3SG like NEG like DEM CL book
 ‘Does he like this book or not?’

2.1.4 Summary of typology of question types and marking strategies

In this section, I have primarily dealt with two issues. The first issue was about the terms *question* and *interrogative*. The term *question* concerns more the aspect of semantics and is related to the set of answers, while *interrogative* concerns the form and is used to differentiate sentence types with declarative and imperative. Understanding the distinction between these two terms is intended to facilitate the reading of previous works, especially when it is necessary to grasp the concepts to which the author is referring. In this dissertation, I do not distinguish between the two terms, as such a distinction would not particularly affect the discussion. Therefore, the term *question* is used as a cover term for all relevant discussions.

The second issue dealt with the characteristics of questions. I provided an overview of typological works concerning question types and each of their marking strategies. From a typological perspective, polar and constituent questions are undeniably the two basic question types. Other question types, such as focus questions, seem to be included in polar questions in that they are formally identical but differ in that they have an additional pragmatic function. In terms of marking strategies, the two basic question types in fact share most of their marking strategies phonologically, morphologically, and syntactically. On the other hand, disjunctive questions, whose status as independent question type has been long debated, are distinctive in their structures. However, before clarifying this unresolved issue for the exact status of disjunctive questions, I discussed its marking strategies separately from the other two question types because of its

structural uniqueness. The summary table (Table 5) below shows how these marking strategies are used in different question types. A grey background indicates that the strategies are shared among different question types. The symbol “V” indicates that the strategy can be used in that question type, whereas the symbol “X” means such a strategy is not attested. The symbol “V?” indicates that the use of the strategy is debatable. Finally, the symbol “?” indicates that the use of the strategy is not mentioned in the literature.

Table 5 Summary of question types and their marking strategies

	Polar questions	Constituent questions	Disjunctive questions
Intonation	V	V	V?
Particle	V	V?	V
Verbal inflection	V	V	?
Tag	V	X	?
Question words	X	V	X
Word order change	V	V	V?
Connectives (overt/covert)	X	X	V
A-not-A	X	X	V

In addition to formal discussions of questions, typological studies extend their interest to the usage of questions. Enfield, Stivers & Levinson (2010) have conducted cross-linguistic research in ten languages (i.e., Ṭākhoe Hailom (Khoisan language family), Danish, Dutch, English, Italian, Japanese, Korean, Lao, Tzeltal, Yéll Dnye) on how question types are used and distributed in question-response systems. Three question types (i.e., polar, constituent, and disjunctive questions) are taken into account in the coding scheme. As their results show, among the ten languages investigated, polar questions are the most frequent type. In addition to empirical study, the theory of social economics of question is put forth by Levinson (2012). This theory predicts that polar questions are the most frequent question type used in daily language because polar questions reduce the face-threatening effect from the perspective of the speaker. It is also the most economical way to use in the question-response system.

After providing a cross-linguistic overview of question types and marking strategies for questions, along the same line of thought, in what follows, I will narrow down the scope and focus only on the two languages investigated in this dissertation, that is, Mandarin and French. Questions type will be firstly addressed, followed by presenting the marking strategies applied in the two languages.

2.2 Questions in Mandarin and French

As shown Section 2.1.1, the number of question types are varied across typology studies, as summarised in Table 6. Most scholars distinguish between two or three question types.

Table 6 Overview of question types in typology research

Number of question type	References
2	Greenberg 1963, Payne 1997 Schulze 2007, Velupillai 2012, Dryer 2013b
3	Ultan 1978: 43, Sadock & Zwicky 1985: 179, Siemund 2001, König & Siemund 2007, Stivers & Enfield 2010, Dixon 2012
4	Hölzl 2018: 56
5	Hölzl 2016: 17–18

The same issue of categorizing questions also appears in individual languages. In English, for example, the dominant view is to distinguish between three question types: polar questions, constituent questions, and disjunctive questions (Quirk et al. 1985; Biber et al. 1999; Stivers 2010: 2773). By contrast, although some linguists such as Siemund (2017) and Sadock (2012: 103), recognise the existence of disjunctive questions, they argue that there are only two question types in English. Accordingly, disjunctive questions are considered as a sub-type of polar questions. The issue of determining a taxonomy of question types also seems to be found in the two languages studied in this dissertation. In the following two sub-sections, I will present an overview of question types and their marking strategies in Mandarin and French, respectively. This will set the stage for the discussion of the dichotomic view of questions in Section 2.3.

2.2.1. Question types and marking strategies in Mandarin

Classifications of question types in Mandarin vary from two to four types. However, the most common classification is the following three-way distinction: polar questions, constituent questions, and disjunctive questions (Huang, Y.-H. Audrey Li & Li 2009), as shown in examples (2.25) respectively.

- (2.25) a. 你 認識 他 嗎？ (Polar question)
nǐ rènshì tā ma?
2SG know 3SG PTCL
'Do you know him?'
- b. 你 想 跟 誰 去？ (Constituent question)
nǐ xiǎng gēn shéi qù?
2SG want with who go
'Who do you want to go with?'
- c. 你 想 看 電影 還是 打 麻將？ (Disjunctive question)
nǐ xiǎng kàndiànyǐng háishì dǎ májiàng?
2SG want see movie or play majiang
'Do you want to see a movie or play majiang?'

The distinction between the three question types in Mandarin follows the same principles mentioned in typology research. That is, questions are categorised based on the set of answers they trigger. Each question type has its own marking strategy (or strategies). The polar question as in (2.25a) expects the answer *yes* or *no* and it is often marked with a particle, such as 嗎 *ma* in this case, while other particles are also possible. Particles will be covered in detail in Section 2.5. The next example (2.25b) is a constituent question, whose answers can be any of all possibilities that meet the criteria for the question word in that question, such as 誰 *shéi* 'who'. The last example in (2.25c) is a disjunctive question. The answer expected is from either of the choices given in the question, namely, 看電影 *kàn diànyǐng* 'see a movie' or 打麻將 *dǎ májiàng* 'play majiang'.⁶

Regarding marking strategies in Mandarin, in addition to **rising intonation** in polar questions (Liou 2020: 84), we can mention four others, including **particles**, **question words**, **disjunctive structures**, and **tags** (Li & Thompson 1981: 520–521). Three of them have been illustrated in examples (2.25). As question words and particles will be

⁶ The *Majiang* is a sort of tile-based game.

addressed separately in Sections 2.4 and 2.5, I concentrate here only on tags and the special disjunctive structure (i.e., A-not-A) in the following paragraphs.

There are three forms of tags in Mandarin: V-not-V, V-particle, and Neg-V-particle, as illustrated in example (2.26). In addition, only a limited range of verbs are allowed in these forms, such as 是 *shì* ‘be’, 有 *yǒu* ‘have’, 對 *duì* ‘true’, 好 *hǎo* ‘good’.⁷ They are originally discourse verbs which are used to respond to a statement or an inquiry of interlocutors (Hsin 2016: 76). Such a constraint on the choice of verb might be attributed to the highly lexicalised nature of tag forms in Mandarin (Siemund 2001: 1016).

- (2.26) 張三 已經 走 了, 是-不-是? / 是嗎? / 不是嗎?
zhāngsān yǐjīng zǒu le shì-bú-shì? / shì-ma? / bú-shì-ma?
 zhangsan alreadygo ASP COP-NEG-COP/ COP-PTCL/ NEG-COP-PTCL
 ‘Zhangsan has gone, hasn’t he? / is it? / isn’t it?’

These three Mandarin tag forms differ in their degree of presupposition. The Neg-V-particle has the strongest presupposition concerning the anchor (i.e., the declarative sentence preceding the tag), such as 不是嗎 *bú-shì-ma* in example (2.26), whereas the V-particle has a relatively weak degree in presupposition, such as 是嗎 *shì-ma*. By contrast, as the structure of the V-not-V contains two choices, it therefore does not have any degree of presupposition compared to the other two tag forms (Hsin 2016: 73). In other words, questions with the V-not-V tag form are unbiased towards the answers. Based on the different levels of presupposition presented in the different Mandarin tag forms, I wish to argue that V-not-V should effectively be regarded as a disjunctive structure for disjunctive questions, despite its occurrence in tags. Typologically, a tag is seen as a marking strategy exclusively for the questions seeking confirmation (i.e., polar questions) (Siemund 2001: 1019; Hsin 2016: 68–69). It is reasonable that it would pair as such, given the presuppositional nature of the polar question in comparison to the other two question types. However, the V-not-V form, when treated as a tag form, seems to be an exception to this rule, since it has no presupposition. It is fully characterised syntactically and semantically as a disjunctive question. Syntactically, V-not-V is identical to a disjunctive structure without a connective. Semantically, the V-not-V form includes a set of answers with two choices. Therefore, I argue that only the V-particle and Neg-V-particle are the

⁷ Note that it has been argued that adjectives in Mandarin are one type of verbs, namely intransitive stative verbs (cf. Paul 2010 and Masini & Simone 2014 for further discussion).

true forms of tags and that the V-not-V form is a disjunctive structure per se that happens to appear in the same position as the other tag forms.

In Mandarin, the V-not-V form (more commonly known as A-not-A as a cover term for such structure, and later referred to as such), as in 要不要 *yào-bú-yào* ‘want-not-want’ in (2.27), is also widely mentioned as a distinctive marking strategy for disjunctive questions.⁸ Accordingly, disjunctive questions with the A-not-A form have even been considered to be a separate question type along with the other three (Tang 1981; Cheng 1984).

- (2.27) 你 要-不-要 離開？
 nǐ *yào-bú-yào* *líkāi?*
 2SG want-NEG-want leave
 ‘Do you want to leave’ (Cheng 1984: 90)

However, Huang, Y.-H. Audrey Li & Li (2009: 238) argue that the A-not-A structure is regarded as one type of disjunctive question instead of an independent category. Hence, the A-not-A structure is simply considered as one sort of disjunctive question that omits the connective, 還是 *háishì* ‘or’, as shown in (2.28).

- (2.28) 張三 買 還是 不買 書？
 Zhāngsān *mǎi* ~~*háishì*~~ *bù-mǎi* *shū?*
 Zhangsan buy or NEG-buy book
 ‘Does Zhangsan buy or not buy books?’ (Huang, Y.-H. Audrey Li & Li 2009: 244)

Yet, the syntactic behaviour of the A-not-A is in fact different from the disjunctive questions with 還是 *háishì* ‘or’ (McCawley 1994: 186–189; Huang, Y.-H. Audrey Li & Li 2009: 246–248). The most obvious difference concerns the order of the two conjuncts. Generally, the positive option tends to precede the negative one, as in (2.28). The reversed order is prohibited in an A-not-A structure, as in (2.29), but it is allowed if the two conjuncts are linked by the connective 還是 *háishì* ‘or’.

- (2.29) * 張三 不買 買 書？
 Zhāngsān *bù-mǎi* *mǎi* *shū?*
 Zhangsan NEG-buy buy book
 Intended for: ‘Does Zhangsan buy or not buy books?’

⁸ The use of both terms is pointed out in Tang (1981: 232, 270). Given that the cases discussed in A-not-A structure is often concerned with verbs, the term V-not-V is also attested. In fact, the more general and common term is A-not-A structure in that other word classes also occur in a such structure, such as prepositions, adverbs (Gasde 2004: 294; Huang 2010: 2). Accordingly, I use A-not-A as a cover term.

Another difference lies in embedded clauses. When the A-not-A is used as a sentential subject in an embedded clause, it is not possible to obtain a direct-question reading, as shown in example (2.30). This constraint does not impact the 還是 *háishì* ‘or’ structure.

- (2.30) * [他 來-不-來] 比較 好 呢?
 * *tā lái-bù-lái bǐjiào hǎo ne?*
 3SG come-NEG-come more better PTCL
 Intended for: ‘Is it better that he comes or that he doesn’t?’
 (Huang, Y.-H. Audrey Li & Li 2009: 246)

The distinctiveness of the A-not-A structure also results from the fact that the other two variants of the A-not-A can be distinguished: V-not-VP and VP-not-V, as shown respectively in examples (2.31) and (2.32): the VP stands for 買書 *mǎi shū* ‘buy books’, a verb 買 *mǎi* ‘buy’ with its complement object 書 *shū* ‘book’.

- (2.31) 你 買 不 買 書? (V-not-VP)
nǐ mǎi bù mǎi shū?
 2SG buy NEG buy book
 ‘Do you buy books or not?’
 (2.32) 你 買 書 不 買? (VP-not-V)
nǐ mǎi shū bù mǎi?
 2SG buy book NEG buy
 ‘Do you buy books or not?’

One of the variants, the VP-not-V, seems to be the most constrained by two principles: Lexical Integrity and Preposition stranding (Huang, Y.-H. Audrey Li & Li 2009: 248). Regarding the principle of lexical integrity, the second part of the VP-not-V (i.e., the verb following the negation) should always be complete as a lexical item, as shown in examples (2.33). In (2.33a), the verb following the negation is incomplete because it lacks the second syllable of the verb. However, it becomes acceptable if it is the first part of VP-not-V (i.e., the verb preceding the negation) and lacks the second syllable of the verb, as in (2.33b).

- (2.33) a. * 你 今天 高興 不高__?
 * *nǐ jīntiān gāoxìng bù-gāo__?*
 2SG today happy NEG-hap__
 Intended for: ‘Are you happy or not?’
 b. 你 今天 高 不 高興?
nǐ jīntiān gāo bù gāoxìng
 2SG today hap__ NEG happy
 ‘Are you happy or not?’

The example (2.33b) is further classified as a V-not-VP structure. Therefore, the constraint of lexical integrity is only required in the disjunctive questions with the VP-not-V structure.

As for the second principle, the preposition is not allowed to strand in the position following the negation, as shown in examples (2.34).

- (2.34) a. * 你們 明天 從 南站 出發 不 從 ?
 * *nǐmen míngtiān cóng Nán-zhàn chūfā bù cóng?*
 2PL tomorrow from South-Station depart NEG from
 Intended for: ‘Will you depart from South Station tomorrow or not?’
- b. 你們 明天 從 不 從 南站 出發 ?
nǐmen míngtiān cóng bù cóng Nán-zhàn chūfā?
 2PL tomorrow from NEG from South-Station depart
 ‘Will you depart from South Station tomorrow or not?’

Not only do the two variants (i.e., VP-not-V and V-not-VP) differ in terms of their syntactic behaviour, both structures are also derived from different syntactic mechanisms. To obtain the VP-not-V structure, an anaphoric ellipsis is applied, while for V-not-VP, reduplication applies (cf. Huang, Y.-H. Audrey Li & Li 2009: 252–253 for more technical details).

Besides the A-not-A structures with variants, another disjunctive structure (i.e., VP-Neg) is also noted (Cheng, Huang & Tang 1996), as shown in examples (2.35): 不 *bù* in (2.35a) and 沒 *méi* in (2.35b). Note that in this case the negation is changed according to the aspectuality class of the main verb (Huang, Y.-H. Audrey Li & Li 2009: 259).

- (2.35) a. 他 買 書 不 ?
tā mǎi shū bù?
 3SG buy book NEG
 ‘Does he buy books or not?’
- b. 你 吃 了 飯 沒 ?
nǐ chī le fàn méi?
 2SG eat ASP rice NEG
 ‘Have you eaten or not?’ (Huang, Y.-H. Audrey Li & Li 2009: 258)

These structures might be taken to be the reduced form of VP-not-V, but they differ in two aspects (Huang, Y.-H. Audrey Li & Li 2009: 258–259). As illustrated in (2.36), first, VP-Neg is not allowed to occur in an embedded clause, as in (2.36a). Second, the particle 呢 *ne* cannot be attached to the VP-Neg structure, as in (2.36b), but it does to VP-not-V.

- (2.36) a. * 我 不 知道 他 買 書 不。
 * wǒ bù zhīdào tā mǎi shū bù
 1SG NEG know 3SG buy book NEG
 Intended for: ‘I don’t know if he buys books or not.’
- b. * 他 買 書 不 呢？
 tā mǎi shū bù ne?
 3SG buy book NEG PTCL
 Intended for: ‘Does he buy books or not.’

(Huang, Y.-H. Audrey Li & Li 2009: 259)

Despite the differences between the VP-not-V and VP-Neg structures, the latter is still classified as one of the A-not-A structures, given that this structure is akin to A-not-A in terms of form, semantics, and pragmatics (Huang, Y.-H. Audrey Li & Li 2009: 259).

Overall, disjunctive questions in Mandarin have various marking strategies, such as 還是 *háishì* ‘or’, the A-not-A structures including its variants (VP-not-V, V-not-VP, VP-Neg). Thus, it seems plausible and justified to consider disjunctive questions as a third question type. However, it is not uncommon to argue that two question types are sufficient and to categorise disjunctive questions as one of two types, such as in Tang (1984), Li (2007), and Her, Che & Bodomo (2022). These authors have argued for a taxonomy of questions that distinguishes between polar and constituent questions. However, they have different opinions about the classification of disjunctive questions. This question will be addressed in Section 2.3 after reviewing French question types and marking strategies.

2.2.2. Question types and marking strategies in French

In French, the same issue regarding the classification of question types recurs. Research on French recognises either two or three question types. Of these, taxonomies that comprise three question types are the most common (Beyssade 2006; Riegel, Pellat & Rioul 2009; Larrivée & Guryev 2021). They are illustrated in examples (2.37).

- (2.37) a. *As-tu parlé à Paul ?*
 have-2SG speak.PTCP PREP Paul
 ‘Have you spoken to Paul?’
- b. *À qui as-tu parlé ?*
 PREP who have-2SG speak.PTCP
 ‘Who have you spoken to?’

- c. *Vous habitez à Londres ou à Paris ?*
 2PL live PREP London or PREP Paris
 ‘Do you live in London or Paris?’ (Larrivée & Guryev 2021: 11)

Although the criteria of distinction among the three question types in French are consistent with the mainstream typological perspective presented in this chapter, different terminologies are used in the literature. A polar question like the one in example (2.37a) is also known as *question totale* ‘total question’ and *question fermée* ‘closed question’. These terms indicate that the subject of questioning covers the whole proposition, and the answers expected are relatively closed-range options, either *yes* or *no*. The constituent question in example (2.37b) is labelled as *question partielle* ‘partial question’ and *question ouverte* ‘open question’, given that the focus of the questioning is not on the entire proposition but on the question word, *qui* ‘who’ and the range of possible answers seem to be larger compared to the binary answer typical of a polar question. Finally, a disjunctive question is presented in (2.37c), in which provides two options of location. The sort of nomenclature (i.e., closed vs. open) echoes the distinction made in Huddleston (1994: 419–420). However, to remain consistent with the terminology applied in this dissertation, I will opt for the terms polar, constituent and disjunctive questions.

Among the marking strategies observed across languages, French mostly uses three of them: **intonation**, **inversion**, and **particle**⁹ (Harris 1998: 237; Guryev 2017: 24; Druetta 2018: 23). Except for intonation that is commonly seen in polar questions, the other two approaches are both available to any question type in French. Yet, behind this great flexibility, there are also constraints and subtleties.

As regards constraints, it is noted that inversion and the particle *est-ce que* are in fact incompatible with each other (Ultan 1978: 48; Siemund 2001: 1017). This incompatibility can be attributed to the fact that the particle *est-ce que* retains its inverse structure after grammaticalization into a particle (Druetta 2011). This can also account for some hypercorrection cases: the occurrence of two marking strategies (i.e., inversion and particle *est-ce que*) is attested (Larrivée & Guryev 2021: 12).

⁹ From the literature, there are two question particles: *est-ce que* and *-ti/-tu*. Compared to the widespread use of *est-ce que* particle, *-ti/-tu* particles are limited in certain regional variants, especially Canadian French (cf. Auger & Villeneuve 2021 for more details). Given that the variant analysed in this dissertation is primarily French from France, only the *est-ce que* particle will be addressed in the discussion.

As for intricate subtleties, the inversion in French involves three variants: Subject-clitic inversion (Drijkoningen 2010), Stylistic inversion (Richard Kayne & Pollock 1978; Lahousse 2006), and Complex inversion (Rizzi & Roberts 1989). These variants have attracted some interest in linguistics, especially under the generative framework (Kayne & Pollock 2001; Lahousse 2003b; 2014), since they pertain to the movement of constituents. I do not go into any technical detail about the movement in inversions, only the examples are presented in (2.38) to illustrate three kinds of inversion.

- (2.38) a. *Est-il parti ?* (Subject-clitic inversion)
 be-3SG leave.PTCP
 ‘Has he left?’
- b. *Où est allé Romain ?* (Stylistic inversion)
 where be.3SG go.PTCP Romain
 ‘Where did Romain go?’
- c. *Le prix a-t-il augmenté ou baissé ?* (Complex inversion)
 ART price have-t-3SG increase.PTCP or decrease.PTCP
 ‘Has the price increased or decreased?’

The difference between these inversions lies in the characteristics of the subject. Subject-clitic inversion refers to questions with a pronominal subject, such as *il* ‘he’ in example (2.38a). Stylistic inversion is also known as nominal inversion because of the presence of nominal subject in the question, like *Romain* in example (2.38b). The last type of inversion is known as complex inversion. As the term shows, the complexity comes from the fact that two constituents cross-reference each other, *le prix* ‘the price’ and *il* ‘it’ in example (2.38c). Moreover, it is this cross-reference that has led to many debates on the issue of true subjects in French linguistics literature (cf. Muller 2007). Despite their apparent complexity, the three types of inversion can be applied to almost any question type. That said, stylistic inversion is incompatible with polar questions (Guryev 2017: 27). This nuance suggests that different question types also have their own peculiarities in the use of these marking strategies. In the following paragraphs, I will discuss how each question type aligns with these strategies.

Although polar questions are not compatible with stylistic inversion, they can use intonation, *est-ce que* particle, and the other two inversions (Hansen 2001), as illustrated in examples (2.39).

- (2.39) a. *Elle va au cours ?*
 3SG go PREP class
 ‘Is she going to class?’
- b. *Est-ce qu’ elle va au cours ?*
 PTCL 3SG go PREP class
 ‘Is she going to class?’
- c. *Va-t-elle au cours ?* (Subject-clitic inversion)
 go-t-3SG PREP class
 ‘Is she going to class?’
- d. *Marie va-t-elle au cours ?* (Complex inversion)
 Mary go-t-3SG PREP class
 ‘Is she going to class?’ (Larrivée & Guryev 2021: 12)

Word order often distinguishes declaratives from questions in French. In example (2.39a), intonation in spoken language or punctuation in written texts plays a role in marking questions, since the word order of the question is identical to the declarative sentence (i.e., *Elle va au cours*. ‘She is going to class.’). Accordingly, this form of polar questions is also known as *questioning declaratives* (Abeillé, Crabbé & Marandin 2012) or *declarative questions* (Zhang 2014). Other marking strategies are related to morpho-syntax. Example (2.39b) adds the *est-ce que* particle in front of a declarative sentence. Examples (2.39c) and (2.39d) illustrate two kinds of inversion adopted for polar questions.

The variety of constituent question structures is highlighted in the literature, ranging from eight (Guryev 2017: 1) to up to sixteen variants (Quillard 2001; Coveney 2011; Larrivée & Guryev 2021). Firstly, three types of inversions are shown in (2.40) in addition to *est-ce que* particle, as in (2.41).

- (2.40) a. *Où va-t-elle ?* (Subject-clitic inversion)
 where go-t-3SG
 ‘Where does she go?’
- b. *Où va Marie ?* (Stylistic inversion)
 where go Mary
 ‘Where does Mary go?’
- c. *Où Marie va-t-elle ?* (Complex inversion)
 where Mary go-t-3SG
 ‘Where does Mary go?’ (Larrivée & Guryev 2021: 12)
- (2.41) *Où est-ce qu’ elle va ?*
 where PTCL 3SG go
 ‘Where does she go?’ (Larrivée & Guryev 2021: 12)

Secondly, some variants are related to the position of question words. Basically, question words tend to be fronted in French. In that case, the word order following the question word can be either reversed, as illustrated by examples (2.40), or identical to a declarative sentence, as in (2.42).

- (2.42) *Où elle va ?*
 where 3SG go
 ‘Where does she go?’ (Larrivée & Guryev 2021: 12)

When question words are found *in-situ*, the question has the same word order compared to its declarative counterpart except that the object is replaced with the question word, as in (2.43). More details related to French question words *in-situ* are discussed in Section 2.4.4.

- (2.43) *Elle va où ?*
 3SG go where
 ‘Where does she go?’ (Larrivée & Guryev 2021: 12)

If a question word is used as a subject, it is naturally placed in the initial position, as *qui* ‘who’ in example (2.44).

- (2.44) *Qui va au cours ?*
 who go PREP class
 ‘Who is going to the class?’ (Larrivée & Guryev 2021: 13)

Finally, other marginal structures are also discussed in the literature, as shown in (2.45). One is in the cleft structure. This structure reinforces the focus on the question word, such as *où* ‘where’ in (2.45a). Two related variants are also proposed but they are considered non-standard, as in (2.45b) and (2.45c).

- (2.45) a. *C’ est où qu’ elle va ?*
 DEM COP where COMP 3SG go
 ‘It is where that she is going?’
 b. *Où c’ est qu’ elle va ?*
 where DEM COP COMP 3SG go
 ‘It is where that she is going?’
 c. *Où qu’ elle va ?*
 where COMP 3SG go
 ‘It is where that she is going?’ (Larrivée & Guryev 2021: 13)

Another marginal structure is the infinitive structure (Riegel, Pellat & Rioul 2009: 278; Coveney 2011: 115). This structure only accepts question words functioning without a

subject because the real subject is implicit and syntactically covert. Also, the real subject can only be the person who addresses the question or a more generic referent as in (2.46).

- (2.46) *Que faire ?*
 what do.INF
 ‘What (should I/we) to do?’ (Coveney 2011: 115)

Constituent questions behave more diversely than polar questions in terms of structural variants. The same gap of structural diversity can also be observed when compared with disjunctive questions in French. In French, the marking strategies used in disjunctive questions are often identical to those for polar questions, including inversion, and the *est-ce que* particle. This happens to such an extent that disjunctive questions are often considered polar questions, as illustrated in examples (2.47) and (2.48).

- (2.47) *Est-ce un bien, est-ce un mal ?*
 COP-DEM ART good COP-DEM ART bad
 ‘Is it good, is it bad?’ (Coveney 2011: 113)

- (2.48) *Est-ce que je continue ou est-ce que j’arrête ?*
 PTCL 1SG continue or PTCL 1SG stop
 ‘Do I go on or do I stop?’ (Coveney 2011: 113)

Example (2.47) can be considered a juxtaposition of two choices without the connective *ou* ‘or’. It can also be the same structure as two polar questions with inversion. Likewise, example (2.48) illustrates the same situation but with the *est-ce que* particle joined by a connective *ou* ‘or’. To be more disjunctive-like, the second part of the question can even be reduced to a negative constituent, as in (2.49).

- (2.49) *Est-ce une vipère ou pas ?*
 COP-DEM ART viper or NEG
 ‘Is it a viper or not?’ (Riegel, Pellat & Rioul 2009: 680)

This reduced form *ou pas* ‘or not’ is however occasionally seen as an extra component appended to polar questions (see the case of Fijian language in Dixon 2012: 399). In general, when speaking of appended components to polar questions, it is the tag form, which is rarely mentioned in the French literature on French questions. Tag forms in French include *oui* ‘yes’, *non* ‘no’, *hein* ‘huh’, *c’est ça* ‘that’s it’, *n’est-ce pas* ‘isn’t it’, *d’accord* ‘OK’ (Debaisieux & Boulton 2007: 41; Abeillé, Crabbé & Marandin 2012). These tags appending to polar questions request confirmation from the addressee. Since French tags are not formally or pragmatically special compared to other languages, they

are mentioned only because of the formal similarity with the reduced negation form in disjunctive questions. Therefore, I will not discuss French tags any further here.

As mentioned above, these disjunctive questions are still very often regarded as a combination of two polar questions (Coveney 2011: 113). The literature commonly puts these two types of questions in the same category, identifying disjunctive questions as a sub-type of polar questions. However, some studies point out the fact that even though the two question types are formally identical (in fact, they are not to a certain extent), they are supposed to trigger different answers (Beyssade 2006: 185; Abeillé & Godard 2021: 1402). In the next section, I address the core issue of question type and aim to provide a more comprehensive point of view toward a plausible taxonomy of questions for both Mandarin and French.

2.2.3. Summary

Section 2.2 have zoomed in on the relevant issues for the two languages under scrutiny here, namely Mandarin and French. Table 7 summarises the classification of question types in the two languages.

Table 7 A summary of question types in Mandarin and French

Number of question type	References in Mandarin	References in French
2 (polar and constituent)	Tang 1984, Li 2007, Her, Che & Bodomo 2022	Wagner & Pinchon 1962, Coveney 2011
3 (polar, constituent, disjunctive)	Li & Thompson 1981, Tiee 1986, Huang, Y.-H. Audrey Li & Li 2009, Wang & Sun 2015	Harris 1998, Beyssade 2006, Riegel, Pellat & Rioul 2009, Guryev 2017, Druetta 2018, Abeillé & Godard 2021: 1402
4 (polar, constituent, disjunctive, A-not-A)	Tang 1981	

Based on this summary table, it appears that three question types are probably the most common categorization in the literature because of the different marking strategies

applied to distinguish three question types. However, the disjunctive questions are often categorised as polar questions due to their formal similarities. Some studies advocate a two-way distinction of question types, both from a typological point of view (cf. Table 6) and from a monolingual perspective, as shown in Table 7.

However, the issue of the taxonomy of questions does not merely stop here. In the two-way distinction, the polar question is understood to contrast with the constituent question in terms of its forms in addition to its relationships with our cognitive model that is to convey the meaning of an event or an object in the world at any given time (Kearsley 1976: 364). As regards the classification of disjunctive questions, disagreement arises. As mentioned earlier, disjunctive questions have been mostly categorised as a subtype of polar questions. Nevertheless, another proposition has been advanced, namely that disjunctive questions and constituent questions are part of the same category. Accordingly, in the following section, I will look at two different arguments for the classification of disjunctive questions and argue for a dichotomic view of questions, which is theoretically more plausible. Following my arguments, a sub-section will further illustrate the plausibility of such a dichotomy based on empirical evidence gathered from the grammaticality judgement tasks in Mandarin and French.

2.3 The dichotomic view of questions in Mandarin and French

This section aims to first provide an overview of the different arguments put forward in typology research and monolingual linguistics (for English and the two languages studied here) on how to categorise disjunctive questions. Then, through empirical evidence from Mandarin and French, I aim to consolidate one of the plausible arguments. The goal of this section is on the one hand to clarify how disjunctive questions should be categorised and to examine whether such a dichotomic view of questions is supported by empirical evidence.

2.3.1 The delineation of a disjunctive question

The most widespread argument is that disjunctive questions and polar questions belong to the same category, given their similarities in certain aspects, such as intonation and syntactic structures.

Rising intonation is generally regarded as one of the marking strategies for polar questions. This strategy is also found in disjunctive questions. Hence, the typical pattern

of intonation in disjunctive questions is rising on the first conjunct and falling on the second or the last (Siemund 2001: 1012).

(2.50) *Do you want to come or not?*

In example (2.50), the first conjunct is presented as a form of polar questions with rising intonation, while the second part shows a shortened form of a disjunction encoded with a falling intonation.

In addition to intonation, disjunctive questions can also be assumed as an extended version of polar questions (Jespersen 1933: 304).

(2.51) *Do you want to come or don't you want to come?*

Example (2.51) can be analysed as a polar question attached to a negative counterpart. As mentioned previously, some studies do analyse disjunctive questions as a series of two (or more) polar questions juxtaposed and connected linked by a connective (cf. Coveney 2011 and Abeillé & Godard 2021: 1410 for French; Li 2007 for Mandarin). Interestingly, from another perspective, polar questions can be examined as a type of disjunctive questions (Karttunen 1977: 383, 391–392), which undergoes ellipsis or is with a covert disjunctive structure. Besides, the disjunctive structure seems to be considered one of the marking strategies for polar questions (Siemund 2001: 1016; König & Siemund 2007: 297–298). In Mandarin, unbiased polar questions use the A-not-A structure, compared to the biased polar questions with particles. Furthermore, the connective 還是 *háishì* ‘or’ can link a series of polar questions (Li 2007: 124–126). Similarly in English, the connective *or* is also attested in polar questions, as in (2.52).

(2.52) *Are you free on Tuesday **or** Wednesday?*

However, if example (2.52) is interpreted as a polar question with a rising intonation toward the end of the question, this contour is different from the typical intonation of disjunctive questions, which falls at the end. In other words, the connective *or* in polar questions is distinguished from the one used in disjunctive questions, despite their identical forms. In fact, such a distinction of different connectives is found in other languages. In Mandarin, two connectives are differentiated by their lexical forms. For instance, only 還是 *háishì* ‘or’ appears in disjunctive questions to link options, whereas 或是 *huòshì* ‘or’ occurs in polar questions without implying any choice.

With the similarities in intonation and structure, polar questions and disjunctive questions are formed as a group of ‘closed interrogative’ (Huddleston 1994), as opposed to ‘open interrogative’, which includes only constituent questions. The reason for such a grouping is that they both have common syntactic environments. For instance, the inversion in ‘closed interrogative’ is without any trigger (i.e., question words) and this type of ‘interrogative’ can be embedded with subordinators like *if* and *whether* (Huddleston 1994: 419).

Despite the large resemblance in intonation and form between polar and disjunctive questions, it is unreasonable to attribute these disjunctive structures to the strategy used in polar questions. For the Mandarin A-not-A structure, since it contains two choices, it implies no bias in the stance of the speaker. In that regard, there is no doubt that such a structure is for disjunctive questions. In my opinion, the fuzzy boundary in phonology and syntax between polar and disjunctive questions is one of the reasons why they have often been categorised together. In reality, they are particularly different in the purpose of questioning. Studies have clearly pointed out that disjunctive questions seek answers that are different from those for polar questions (Jespersen 1933: 304; Siemund 2001: 1012; Beyssade 2006: 185; König & Siemund 2007: 291–292), namely, polar questions expect the binary answers *yes* and *no*; whereas disjunctive questions expect information, which is already provided as choices in the question. Consequently, some studies argue that disjunctive questions belong to the same category as constituent questions.

This argument is mostly evidenced from a semantic perspective by the observation of their expected answers. That is, as opposed to the logical polarity of *yes* and *no* expected by polar questions, disjunctive questions should trigger exact information.

(2.53) *Is it raining or is it snowing?* (Sadock & Zwicky 1985: 179)

In example (2.53), the inquirer is waiting for an answer with exact information about the weather. The same concept can be found in constituent questions.

(2.54) *How is the weather?*

The expected answers for example (2.54) can range from the smallest set, such as two choices presented in example (2.53), to a set of unlimited possibilities. Accordingly, both question types are grouped together in that they all seek information as their answers (Sadock & Zwicky 1985: 179; Luo 2013: 9–10). The only difference lies in the number

of possible answers: two (or more) can be listed in disjunctive questions, whereas in constituent questions, an unlimited number of answers is expected only if the scope of the answers falls in the universe of discourse between locutors (Tang 1984: 383). Semantically, using a disjunctive structure is to convey a set of substitutable possibilities (Mauri 2008: 25), while asking a constituent question is to expect a set of propositions expressed by possible answers to it (Karttunen 1977: 387). Besides, given that both types of question present their set of answers either explicitly in disjunctive questions or implicitly in constituent questions, they have no presupposition when questioning. In other words, they are unbiased in terms of their answers, which is in sharp contrast with polar questions (Li & Thompson 1981: 550; Li 2007: 117).

Along with their shared semantics, both question types can also be marked with the same particle, as shown in (2.55). In Mandarin, 呢 *ne* particle occurs in disjunctive in (2.55a) and constituent questions in (2.55b). However, the 呢 *ne* particle is not allowed in polar questions, as in (2.55c).

- (2.55) a. 外面 下雨 還是 下雪 呢?
wàimiàn xiàyǔ hái shì xià xuě ne?
 outside rain or snow PTCL
 ‘Is it raining or snowing outside?’
- b. 今天 天氣 怎麼樣 呢?
jīntiān tiānqì zěnmeyàng ne?
 today weather how PTCL
 ‘How is the weather today?’
- c.* 今天 天氣 好 呢?
jīntiān tiānqì hǎo ne?
 today weather good PTCL
 Intended for: ‘Is it good, today’s weather?’

Note that disjunctive questions with A-not-A variants other than VP-Neg can also take the 呢 *ne* particle, as in (2.56).

- (2.56) 外面 有-沒-有 下雨 呢?
wàimiàn yǒu-méi-yǒu xià yǔ ne
 outside have-NEG-have rain PTCL
 ‘Is it raining outside?’

Therefore, regardless of the form presented in disjunctive questions, both disjunctive and constituent questions can be marked identically by the 呢 *ne* particle. In addition to the

(2.57) a.* 你們 在 討論 [高-不-高興 的 那個 人] ?
 nǐmen zài tāolùn [gāo-bù-gāoxìng de nàge rén]?
 2PL ASP discuss hap-NEG-happy DE DEM person
 Intended for: ‘Are you discussing the person who is happy or the one who
 isn’t?’
 (Huang, Y.-H. Audrey Li & Li 2009: 256)

(Huang, Y.-H. Audrey Li & Li 2009: 263)

To summarise, even though grouping polar and disjunctive questions seems to be the most welcomed and widely recognised, I intend to argue that it is more plausible to categorise disjunctive questions with constituent questions from both syntactic and

semantic points of view. One may argue that the illustrations above concern mostly Mandarin instead of French. Indeed, such a claim seems to be rare for French. This might be because disjunctive questions do not attract much attention in French linguistics and instead have long been regarded as one special question type on the continuum between polar and constituent questions (Druetta 2021: 91–92). In other words, disjunctive questions can be either polar or constituent questions, depending on the perspectives taken. In fact, some studies argue that different forms of disjunctive questions should be treated differently (Stenström 1984: 153; Minga et al. 2020: 739). Nevertheless, I still wish to argue that disjunctive questions should be categorised with constituent questions, be they in French or Mandarin. In the next sub-section, I will resort to empirical evidence to check whether my claim holds against data gathered from native speakers of the two languages investigated.

Before moving to the next section, one might argue against the necessity of categorizing disjunctive and constituent questions as one question type. After all, both question types are different to some extent, such as the A-not-A structure in disjunctive questions or question words in constituent questions. Admittedly, it is still possible to distinguish the two question types from certain angles. However, beyond the three-way distinction based on forms and sets of answers, I think the two-way distinction is more advantageous from two perspectives. First, the two-way distinction of question types shows its plausibility from semantic and syntactic perspectives. Second, in the light of philosophical principle, Occam's razor, if a binary distinction can be made and be justified empirically, such a distinction should be favoured (Her, Che & Bodomo 2022). Consequently, the aim here is to present the classification of question types from different linguistic layers, as shown in Table 8. Syntactically, although a three-way distinction is possible, disjunctive and constituent questions indeed share a certain number of syntactic features. Semantically, questions can be distinguished by the number of the set of propositions that they have for interlocutors to choose from, namely a single set of propositions vs. multiple sets of propositions (Bhatt & Dayal 2020: 1124). Pragmatically, based on the nature of the answers expected, questions can be distinguished as confirmation-seeking or information-seeking (Dixon 2012: 377; Her, Che & Bodomo 2022).

Table 8 Classification of question types based on linguistic levels

Linguistic level	Classification of question types		
Pragmatic	Confirmation-seeking	Information-seeking	
Semantic	Singleton set of propositions	A non-singleton set of propositions	
Syntactic 1	Polar questions	Constituent questions	
Syntactic 2	Polar questions	Disjunctive questions	Constituent questions

2.3.2 Empirical evidence of taxonomy of question in Mandarin and French

Based on the discussion in Section 2.3.1, disjunctive and constituent questions should be grouped together, as both differ from polar questions at the morphosyntactic, semantic, and pragmatic levels. In other words, a binary distinction between polar questions and non-polar (i.e., disjunctive and constituent) questions is theoretically well-founded. This dichotomy has been supported in analyses of several languages, including Taiwan Mandarin, Xiang and English (Her, Che & Bodomo 2022), Taiwan Southern Min (Hsiao & Her 2021), and Paiwan (Austronesian) (Huang & Her 2024). Additional empirical evidence further reinforces this two-way distinction of question types, such as grammaticality judgement tasks conducted in Mandarin (Hsu, Chen & Her, submitted) and in French (Hsu, Allasonnière-Tang & Her, in preparation).

Following the syntactic criteria proposed by Her, Che & Bodomo (2022), Hsu, Chen & Her (submitted) adopt three syntactic features to distinguish question types: *particle* (*ma* vs. *ne*), *adverb* (*nándào* vs. *dàodǐ*), and *question form* (direct vs. indirect question), as shown in Table 9.

Table 9 Distinguishing features and conditions

Distinguishing features	Conditions	
Particle	① 嗎 <i>ma</i>	② 呢 <i>ne</i>
Adverb	③ 難道 <i>nándào</i>	④ 到底 <i>dàodǐ</i>
Question form	⑤ direct question	⑥ indirect question

Each syntactic condition further exhibits grammatical contrasts across different question types, as illustrated in Table 10.

Table 10 Grammaticality prediction of distinguishing features and conditions

Distinguishing features	Conditions	Polar	Wh	Disjunctive
Particle	嗎 <i>ma</i>	✓	X	X
	呢 <i>ne</i>	X	✓	✓
Adverb	難道 <i>nándào</i>	✓	X	X
	到底 <i>dàodǐ</i>	X	✓	✓
Question form	Direct	✓	✓	✓
	Indirect	X	✓	✓

(grammatical conditions are marked with a symbol ✓, while ungrammatical conditions are marked with X and a grey background)

Based on the grammaticality contrasts observed under these syntactic conditions, a timed grammaticality judgment task was conducted to determine whether these syntactic features empirically support the proposed dichotomy of questions. Both grammaticality judgments and reaction times were recorded as dependent measures.

The descriptive results of the judgement responses are presented in Figure 4. Statistically, there is a significant difference between the acceptance rates of grammatical and ungrammatical conditions ($\chi^2(1) = 150.56, p < .0001$).

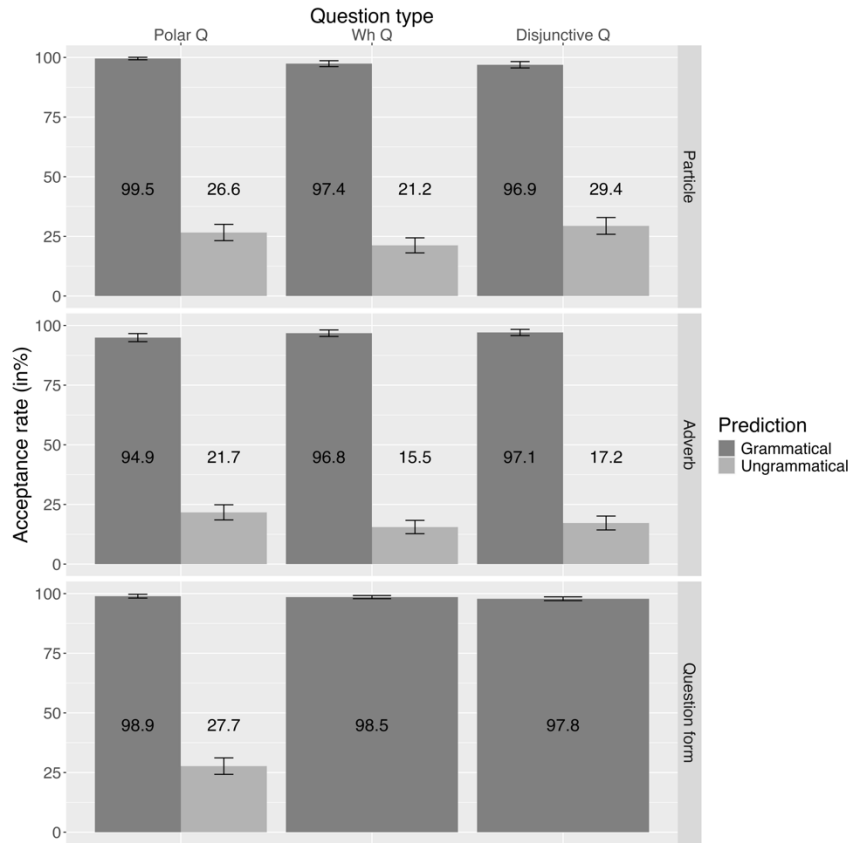


Figure 4 Acceptance rate to target conditions in Mandarin grammaticality judgment task

As for reaction times, Figure 5 presents the descriptive results. The statistical analysis indicates that reaction times for grammatical conditions are significantly shorter than those for ungrammatical ones ($\chi^2(1) = 24.02, p < .0001$).

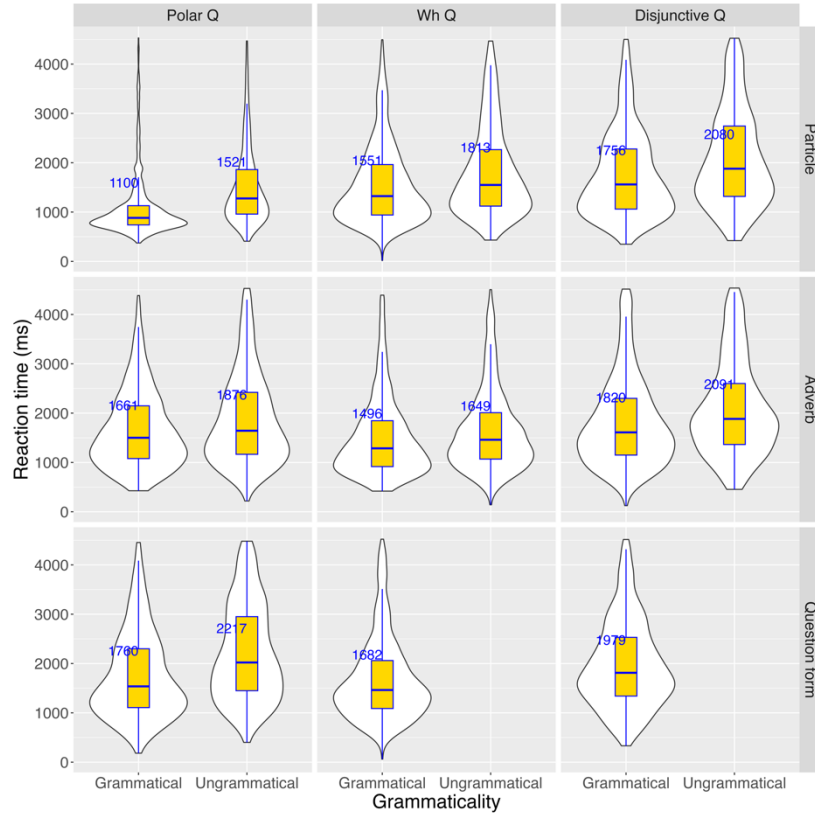


Figure 5 Reaction time to target conditions in Mandarin grammaticality judgment task

These experimental results provide empirical support for the proposed dichotomy of questions types in Mandarin. On the one hand, the acceptance rates indicate that the three syntactic features reliably predict grammaticality. On the other, the reaction time data further strengthens the robustness of the grammatical contrasts observed in interaction between syntactic conditions and question types. It is worth noting that a semantic distinction between question types is implicitly reflected in the syntactic contrasts revealed by the data.

Given that the dichotomy of question types has been argued cross-linguistically, Hsu, Allasonnière-Tang & Her (in preparation) conducted a similar experiment in French. Building on the design of the Mandarin experiment and incorporating the morpho-syntactic characteristics of French, two syntactic features were selected for the French experiment: *adverb* and *question form*, as shown in Table 11.

Table 11 Distinguishing features and conditions

Distinguishing features		Conditions	
Adverb	① <i>Serait-il possible que</i>	② <i>En fin de compte,</i> <i>À la fin</i>	
Question form	③ Direct question	④ Indirect question	

The selection of adverbs was based on *Grand dictionnaire Ricci de la langue chinoise* (Institut Ricci 2001) and a bidirectional parallel corpus of contemporary Mandarin and French novels. This approach enabled the identification of adverbial phrases that contrast in semantic function within questions, as illustrated in Table 12.

Table 12 Fundamental semantic contrast of the adverb in French

Fundamental Semantic contrast	
Casting doubt on the truth hood of one proposition	Forcing a choice among two or more propositions
<i>Serait-il possible que</i> ‘is it possible that’	<i>En fin de compte</i> ‘in the end’ <i>À la fin</i> ‘in the end’

According to the theory of the dichotomy of questions (Her, Che & Bodomo 2022), Hsu, Allasonnière-Tang & Her (in preparation) argue that these adverbial phrases are associated with different question types. Specifically, the phrase *serait-il possible que*, which cast doubt on the truth hood of one proposition, is compatible only with polar questions. In contrast, the adverbial expressions *en fin de compte* and *à la fin* are to prompt the addressees to select one piece of information among two or more propositions, thus occurring with disjunctive and *wh*-questions, which involve multiple propositions.

With respect to the question form in French experiment, French does not exhibit any grammaticality contrasts for this syntactic feature across question type. In contrast to Mandarin, all French question types can form indirect questions. For polar questions, the complementiser *si* is typically used to construct an indirect question. However, *est-ce que* is also attested in indirect polar questions. The acceptability of this variant appears to vary depending on the educational background of the speaker. Based on these observations,

Hsu, Allasonnière-Tang & Her (in preparation) argue that French may be undergoing a shift in its taxonomy of question types. Specifically, if native speakers accept *est-ce que* as a complementiser in indirect polar questions, polar questions can no longer be syntactically distinguished from *wh*- and disjunctive questions. In this case, French may be said to have only one syntactic question type, namely, information-seeking questions. Conversely, if native speakers only accept *si* in indirect polar questions and reject *est-ce que*, this suggests that polar questions remain syntactically distinct from other question types.

Accordingly, the four basic conditions were crossed with three question types, as shown in Table 13.

Table 13 Grammaticality prediction of distinguishing features and conditions

Distinguishing features	Conditions	Polar	<i>Wh</i>	Disjunctive
Adverb	<i>Serait-il possible que</i>	✓	X	X
	<i>En fin de compte,</i> <i>À la fin</i>	X	✓	✓
Question form	Direct	✓	✓	✓
	Indirect with a complementiser	✓ (<i>si</i>)	✓	✓ (<i>si</i>)
	Indirect <i>est-ce que</i>	X	NA	X

(grammatical conditions are marked with a symbol ✓, while ungrammatical conditions are marked with X and a grey background)

The results of judgement task are shown in Figure 6. A significant difference is found between the acceptance rates of grammatical and ungrammatical conditions ($\chi^2(1) = 26.12, p < .0001$), with the exception of the adverb condition in disjunctive questions ($z = -2.047, p = .6155$).

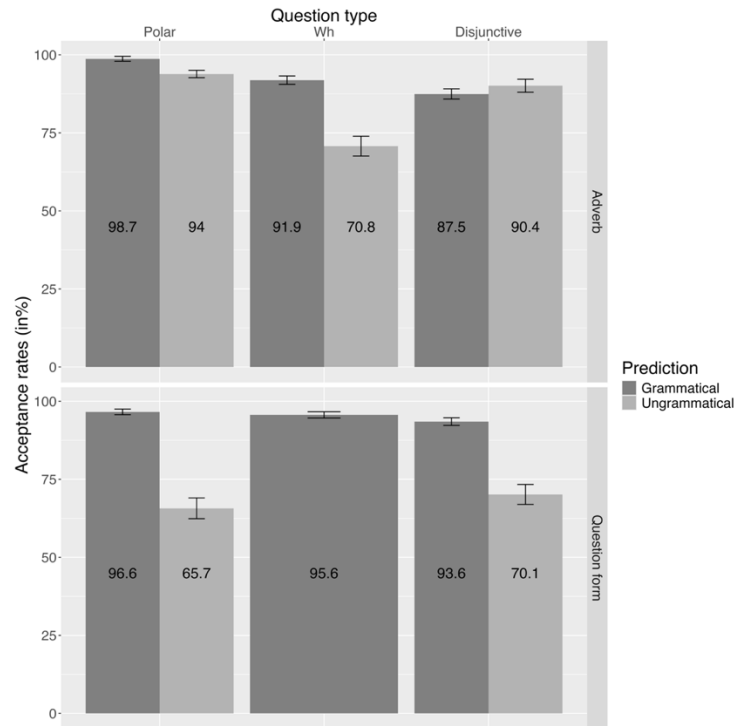


Figure 6 Acceptance rate to target conditions in French grammaticality judgment task

Although the overall difference is statistically significant, ungrammatical conditions, particularly in the adverb condition of polar and disjunctive questions, receive relatively high acceptance rates. In the case of *wh*-questions, the acceptance rate for ungrammatical condition also exceeded 70%. The high acceptance of ungrammatical condition in stark contrast to the results from the Mandarin experiment (see Figure 4), where acceptance rates for ungrammatical conditions are consistently below 30%. These findings suggests that, unlike Mandarin, French does not exhibit a clear morphosyntactic distinction between polar questions and other question types.

Regarding the acceptance rate of the complementiser *est-ce que* across educational levels, the overall analysis no statistically significant difference ($\chi^2(5) = 8.21, p = .145$). Nevertheless, Figure 7 illustrates predicted acceptance rates by question types across different educational levels.

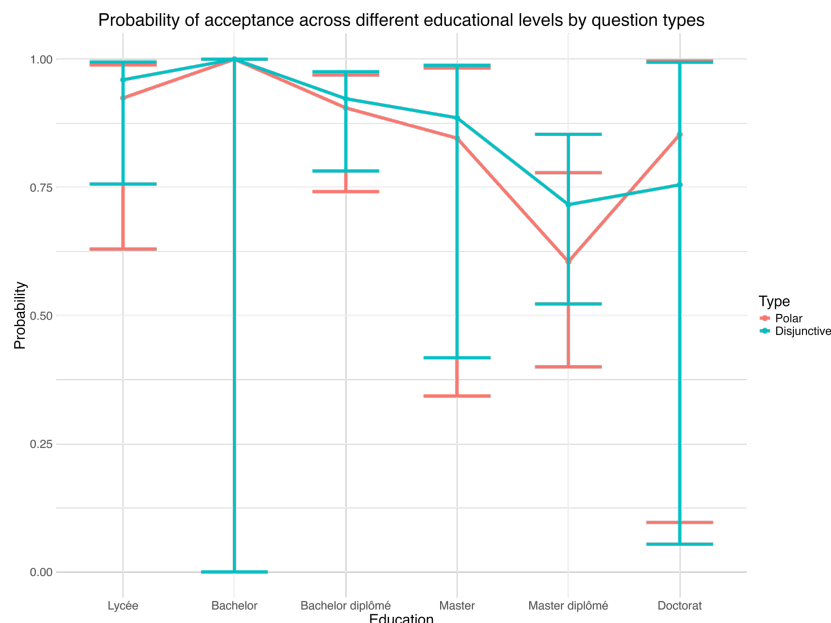


Figure 7 Probability of acceptance rate across different educational levels

A decreasing trend is observed from high school to PhD level. In other words, native speakers with higher educational attainment (e.g., Master's or PhD degree) tend to reject the use of *est-ce que* in indirect polar and disjunctive questions more than those with only a high school degree.

The overall reaction time between grammatical and ungrammatical conditions also shows a significant difference ($\chi^2(1) = 5.2, p < .05$). However, in the adverb condition of polar questions, the reaction time for the ungrammatical condition is slightly shorter than for grammatical ones, and this difference is not statistically significant (z value = 2.28, $p = .4476$), as shown in Figure 8.

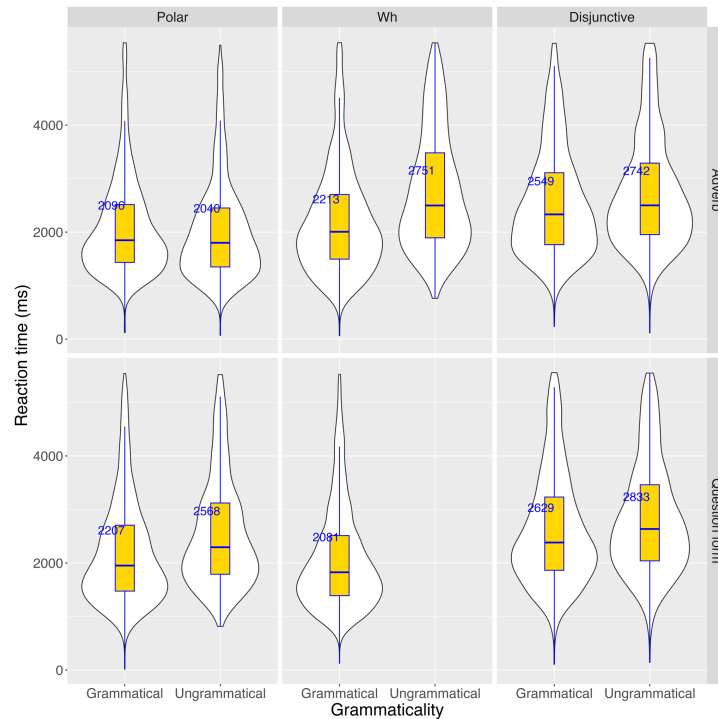


Figure 8 Reaction time to target conditions in French grammaticality judgment task

The results from the French grammaticality judgment task suggest that French may lack morphosyntactic distinction between polar questions and other question types. However, a tendency toward a dichotomy of question types appears to emerge among speakers with higher levels of education. This emerging pattern may account for the significant differences observed in both acceptance rates and reaction times.

2.3.3 Summary

This section discussed the taxonomy of questions in Mandarin and French based on the empirical evidence from grammaticality judgement tasks. The results indicated that **the two languages differ in their morphosyntactic distinction of two question types** (i.e., confirmation-seeking and information-seeking questions). In Mandarin, these two question types are distinguishable through the use of adverbs and question forms. By contrast, French does not exhibit a clear morphosyntactic distinction between these question types, despite the semantic contrast of the selected adverbs. However, a tendency toward a dichotomy of question type emerges among French speakers with higher educational levels, such as those holding Master's or PhD degrees. This suggests that although French does not fully distinguish polar questions using morphosyntactic elements, a nascent dichotomy may develop in speakers with advanced education.

Crucially, this cross-linguistic comparison of theoretical and experimental results provided a foundation for interpreting the distributional patterns of question types observed in authentic data. The findings presented here will serve as a basis for the corpus-based observations and analyses.

2.4 Question words

2.4.1 Question words and their syntactic positions

Question words, also known as *wh-words* or *interrogative pronouns/words* (Hölzl 2018: 76), are distinctive in their marking strategy, given that they only occur in constituent questions. This marking strategy is universal (Ultan 1978: 53; Mackenzie 2009: 1133), except for one exceptional language from the Chapacuran family, which uses demonstratives as question words (Hölzl 2016: 19).

Constituent questions use different question words to indicate the specific content being sought. The eight canonical question words, distinguished typologically in Dixon (2012: 407), are *who*, *what*, *where*, *when*, *why*, *how*, *which*, and *how many/much*. However, not all languages have such a wide range of question words (Ultan 1978: 53; Hölzl 2018: 77). Question words are often categorised based on their word class (Diessel 2003: 637; Velupillai 2012: 358) and semantic scope (Diessel 2003: 640; Mackenzie 2009: 1132), even though their forms may differ across languages. Table 14 presents the categorization of eight canonical question words mentioned above. Previous studies have proposed different word classes and semantic scope subcategories, but that is not the focus of this section (cf. Hölzl 2018: 81, 85 for more details).

Table 14 Categorization of question words (Dixon 2012: 407; Diessel 2003: 637, 640)

Question words	Word class	Semantic scope
<i>who</i>	Pronoun	person
<i>what</i>		object
<i>where</i>		place
<i>when</i>	Adverb	time
<i>why</i>		cause
<i>how</i>		manner
<i>which</i>	Determiner	selection
<i>how many/ much</i>	Quantifier	quantity

Given the variety of question words in terms of their forms, word classes, and semantic scopes, different dimensions of question words are investigated, such as polysemy and complexity including the number of phonemes and word length (cf. Heine, Claudi & Hünemeyer 1991: 55–59 and Mackenzie 2009). In addition, studies have also extended the scope to the structural level, such as the syntactic position of question words. This dissertation will focus on this latter topic.

Under the generative theory framework, the position of question words in syntax is often a topic of discussion. Languages are categorised into three groups based on their neutral constituent questions, namely constituent questions without conveying additional pragmatic purpose as in rhetorical questions: (1) fronting languages, which require question words in clausal-initial position; (2) *in-situ* languages, which have question words in the same position as the constituent questioned; and (3) optional fronting languages, which have both possibilities (Siemund 2001: 1019).

According to the typological survey conducted by Ultan (1978: 54), fronting languages have a larger proportion (i.e., 73.4% of 53 languages), compared to the other two groups. However, in a larger scale typological work by Dryer (2013c), the group of fronting languages only account for 29% ($n = 264$), while 68% of languages ($n = 615$) do not obligatorily have initial question words. The difference between these two studies may be due to the sample of different language families and the criteria used to categorize the position of question words. Certain language families in Europe, North Africa, and areas around the Philippines tend to place their question words initially. In contrast, question words in most African and the mainland Asian language families, including Japanese,

does not occur in the initial position. Ultan (1978) does not clearly explain their categorization criteria. However, determining the position of question words in a language can be relatively difficult, especially when deciding between *in-situ* and optional fronting (Cheng 1991; Siemund 2001: 1020). The difficulty is demonstrated with Scottish Gaelic (Siemund 2001: 1020). This language places question words initially in a specific structure (i.e., relative clause), although it seems to be an *in-situ* language.

Both surveys present different statistical results regarding the position of question words across languages; however, they both indicate that geographical distribution and basic word order of a language account for the position of question words. These two factors are related in the sense that geographical distribution illustrates the clusters of language families, and each language family has a high probability of sharing similar or identical word order. The correlation between basic word order and the position of question words has been highlighted (König & Siemund 2007: 294–295). This correlation has been formulated as Universal principle 12 by Greenberg (1963: 65). According to this principle, languages with VSO word order tend to place their question words initially. In languages with SOV word order, question words are found *in-situ*. However, this correlation is not present in languages with SVO word order, indicating that both fronting and *in-situ* are used in SVO languages (Siemund 2001: 1020).

Based on the data collected in Greenberg (1963), in addition to basic word order, Baker (1970: 207, 208) argues that question words can only be fronted when the polar question in such a language is marked by a particle in the initial position. This generalization implicitly rules out the situation where a language places its question words initially while it has a particle in the final position. Accordingly, this generalization predicts the position of question words in languages based on the position of particles in polar questions. Nevertheless, this generalization has been challenged by counterexamples in languages such as Hopi, Bahasa Indonesia, and Hindi (Cheng 1991: 19).

Although the basic word order and position of particles in polar questions can somewhat predict the position of question words, such correlations are not absolute (Bruening 2007: 143 ff.) Likewise, counter-evidence can be found in French, which has a particle (i.e., *est-ce que*) in the initial position of polar questions, while allowing question words to be *in-situ* (e.g., *Tu manges quoi?* Literal translation: ‘You eat what?’).

In the same vein, Mandarin is considered an *in-situ* language, yet its question words may undergo an implicit movement. In other words, the question words in Mandarin only move on the level of Logical form and such a movement is not perceived on the morpho-syntactic level (e.g., word order). This section primarily considers the *in-situ* phenomenon in both Mandarin and French. First, I present the question words in both languages, followed by separate sections dedicated to their respective syntactic mechanism of *in-situ*. Finally, a summary concludes this section from a contrastive perspective.

2.4.2 Question words in Mandarin and French

Both Mandarin and French have the canonical question words to fulfil the specific information-seeking. This section mainly focuses on the **eight canonical question words in Mandarin and French**, including their positions in questions. Additionally, some special uses or combinations of question words are also addressed.

In Mandarin, the question words are as follows: 誰 *shéi* ‘who’, 什麼 *shénme* ‘what’, 哪裡 *nǎlǐ* ‘where’, 為什麼 *wèishénme* ‘why’, 怎麼 *zěnmé* ‘how’, 什麼時候 *shénme shíhòu* ‘when, what time’, 哪-*classifier nǎ-CL* ‘which one’, 幾 *jǐ* ‘how many’, 多少 *duōshǎo* ‘how many/ much’ (Li & Thompson 1981: 522; Cheng 1984: 90; Huang, Y.-H. Audrey Li & Li 2009: 260; Yip & Rimmington 2021: 160). These words can be classified into different word classes and semantic scopes, as shown in Table 15.

Table 15 Categorization of question words in Mandarin

Question words	Word class	Semantic scope
誰 <i>shéi</i> 'who'	Pronoun	person
什麼 <i>shénme</i> 'what'		object
哪裡 <i>nǎlǐ</i> 'where'	Adverb	place
什麼時候 <i>shénme shíhòu</i> 'when, what time'		time
為什麼 <i>wèishénme</i> 'why'		cause
怎麼 <i>zěnmē</i> 'how'		manner
哪- <i>classifier nǎ-CL</i> 'which one'	Determiner	selection
幾 <i>jǐ</i> 'how many' 多少 <i>duōshǎo</i> 'how many/ much'	Quantifier	quantity

Mandarin also has a type of question word used as a verb, namely, 幹嘛 *gànmá* 'to do what' (Ulan 1978: 53). Although it is frequently used (Li & Thompson 1981: 525), it is not a question verb per se, but rather a form composed of the verb 幹 *gàn* 'do' and the phonetically reduced question word 什麼 *shénme* 'what'.

In Mandarin, question words take on diverse forms and are commonly used in *in-situ*, meaning that they are substituted by the expected answer in the same position. However, the position of these *in-situ* question words can vary within a clause depending on their semantic scope and syntactic role (e.g., subject, object).

Firstly, the **pronouns** 誰 *shéi* 'who' and 什麼 *shénme* 'what' can function as both subjects and objects. Their syntactic roles are determined by their positions, as shown in (2.58) and (2.59).

(2.58) a. 誰 吃 了 蘋果? (subject)
shéi chī le píngguǒ?
 who eat ASP apple
 ‘Who ate the apple?’

b. 你 喜歡 誰? (object)
nǐ xǐhuān shéi?
 2SG like who
 ‘Who do you like?’

(2.59) a. 什麼 是 愛? (subject)
shénme shì ài?
 what COP love
 ‘What is love?’

b. 你 吃 什麼? (object)
nǐ chī shénme?
 2SG eat what
 ‘What do you eat?’

The question word 什麼 *shénme* ‘what’ can also modify a noun phrase, such as 什麼東西 *shénme dōngxī* ‘what thing’, 什麼人 *shénme rén* ‘what person’ (Li & Thompson 1981: 523). Additionally, this construction based on 什麼 *shénme* ‘what’ is sometimes used to replace the question words 誰 *shéi* ‘who’ with 什麼人 *shénme rén* ‘what person’; 哪/哪裡 *nǎ/nǎlǐ* ‘where’ with 什麼地方 *shénme dìfāng* ‘what place’. Question words can be formed not only by a single constituent but also by a combination of two constituents (Velupillai 2012: 357).

Secondly, a series of adverbs, including 哪裡 *nǎlǐ* ‘where’, 什麼時候 *shénme shíhòu* ‘when, what time’, 為什麼 *wèishénme* ‘why’, and 怎麼 *zěnmē* ‘how’ are typically found in the adverbial position. However, they differ in terms of their adverbial categories, such as manner, sentential, locational, and temporal. Accordingly, their position in the clause may vary from one to another (Cinque 1999).

The locational and temporal adverbs are placed either at the beginning or end of a sentence. For example, see (2.60) and (2.61).

(2.60) a. 哪裡 有 賣 蜂蜜?
nǎlǐ yǒu mài fēngmì?
 where have sell honey
 ‘Where do people sell honey?’

- b. 他 去 哪裡?
tā qù nǎlǐ?
 3SG go where
 ‘Where does he go?’
- (2.61) a. 什麼 時候 可以 玩 手機?
shénme shíhòu kěyǐ wán shǒujī?
 what moment can play cellphone
 ‘When can (we) play cellphone?’
- b. 他 (的) 生日 是 什麼 時候?
tā (de) shēngrì shì shénme shíhòu?
 3SG DE birthday COP what moment
 ‘When is his birthday?’

Another question form 哪(兒) *nǎ(r)* for ‘where’ is not included. On the one hand, this question word is rarely used in Taiwan Mandarin. On the other hand, this word might be confused with 哪 *nǎ*, which is used as a question word for selection. In addition, it is worth noting that the question word for time, 什麼時候 *shénme shíhòu* ‘when, what time’ is a composed form of a modifier 什麼 *shénme* ‘what’ and a noun 時候 *shíhòu* ‘moment’. Unlike *when* in English, there is no exclusive question word for asking about time in Mandarin.

The other two question words 為什麼 *wèishénme* ‘why’ and 怎麼 *zěnmě* ‘how’ are used to seek the cause and manner respectively. These adverbs can be found in the initial and preverbal positions, as shown in the following examples in (2.62) and (2.63).

- (2.62) a. 為什麼 他 不 開心?
wèishénme tā bù kāixīn?
 why 3SG NEG happy
 ‘Why he is not happy?’
- b. 他 為什麼 不 開心?
tā wèishénme bù kāixīn?
 3SG why NEG happy
 ‘Why he is not happy?’
- (2.63) a. 怎麼 他 不 開心?
zěnmě tā bù kāixīn?
 how 3SG NEG happy
 ‘How come he is not happy?’

- b. 你 怎麼 來 學校？
 nǐ *zěnmē* *lái* *xuéxiào?*
 2SG how come school
 ‘How/Why do you come to school?’

The question word 怎麼 *zěnmē* ‘how’ has two different meanings: cause and manner. When this question word is found in the initial position as in (2.63a), only the meaning of cause is acceptable. Accordingly, 怎麼 *zěnmē* functions as 為什麼 *wèishénme* ‘why’ in this question. However, when 怎麼 *zěnmē* is in the preverbal position, its interpretation might be ambiguous. In other words, both interpretations of *how* and *why* are acceptable (Li & Thompson 1981: 523). Two interpretations can be disambiguated by adding an aspect marker and modals (Tsai 2008: 84–85), as shown in (2.64).

- (2.64) a. 你 怎麼 來 了 學校？
 nǐ *zěnmē* *lái* *le* *xuéxiào?*
 2SG how come ASP school
 ‘Why do you come to school?’
- b. 你 怎麼 可以 來 學校？
 nǐ *zěnmē* *kěyǐ* *lái* *xuéxiào?*
 2SG how can come school
 ‘How come you could come to school?’
- c. 你 可以 怎麼(樣) 來 學校？
 nǐ *kěyǐ* *zěnmē(yang)* *lái* *xuéxiào?*
 2SG can how come school
 ‘How can you come to school?’

In comparison to the ambiguous example (2.63b), the addition of the aspectual marker 了 *le* in (2.64a) ensures a causal interpretation and excludes the reading of manner. In addition, the contrast between (2.64b) and (2.64c) suggests that two different interpretations of 怎麼 *zěnmē* ‘how’ can be distinguished according to its position with modals, such as 可以 *kěyǐ* ‘can’ in this case. Furthermore, disambiguation can also be achieved by using alternative forms of 怎麼 *zěnmē*, such as 如何 *rúhé* and 怎麼樣 *zěnmeyàng*. However, only the preverbal 怎麼樣 *zěnmeyàng* has a manner reading (Tsai 1994a: 125–126).

Thirdly, the question word for selection is composed of 哪 *nǎ* and a noun phrase with a classifier, such as 哪個人 *nǎ-ge rén* ‘which person’, 哪個地方 *nǎ-ge dìfāng* ‘which place’ (Huang, Y.-H. Audrey Li & Li 2009: 260). Their syntactic positions are different.

For instance, 哪個人 *nǎ-ge rén* ‘which person’ has the same meaning as *who* and occupies the positions available for pronouns. 哪個地方 *nǎ-ge dìfāng* ‘which place’ is used as a locational adverb, like *where*, and is allowed to appear in the positions of that category of adverbs.

Finally, in terms of the quantity, Mandarin uses both 幾 *jǐ* ‘how many’ and 多少 *duōshǎo* ‘how many/ much’. Both words function as modifiers followed by other constituents, such as 幾顆蘋果 *jǐ kē píngguǒ* ‘how many apples’, and 多少錢 *duōshǎo qián* ‘how much money’. This type of question word occurs in the position of pronouns.

Overall, question words in Mandarin are typically placed according to their syntactic roles in clauses (e.g., subject, object, adverb). In addition, question words are usually found in the same position where the requested information would substitute. This makes Mandarin an *in-situ* language. Among the question words, adverbial question words have more flexibility in terms of position. However, this flexibility can lead to ambiguous interpretations when encountering 怎麼 *zěnmē* ‘how’ in the preverbal position. It can have interpretations for both cause and manner. Finally, the question words in Mandarin are formally identical to indefinite pronouns (e.g., *whoever*, *whatever*) (Li & Thompson 1981: 527–531; Huang, Y.-H. Audrey Li & Li 2009: 292–293; Dixon 2012: 401). In non-interrogative contexts, the question words mentioned above can be used (e.g., 誰醉了就罰十塊錢 *shéi zuì le jiù fá shíkuàiqián* ‘Whoever gets drunk will be fined ten dollars’). However, this is not directly relevant to the present study (cf. Y.H Li 1992 and Lin 2014: 186–193 for more details).

The canonical question words in French are categorised within the basic word classes and semantic scopes, as shown in Table 16 (Riegel, Pellat & Rioul 2009: 383–385; Abeillé & Godard 2021: 1405).

Table 16 Categorization of question words in French

Question words	Word class	Semantic scope
<i>Qui</i> 'who'	Pronoun	person
<i>Que/ Quoi</i> 'what'		object
<i>Où</i> 'where'	Adverb	place
<i>Quand</i> 'when'		time
<i>Pourquoi</i> 'why'		cause
<i>Comment</i> 'how'		manner
<i>Quel</i> 'which'	Determiner	selection
<i>Combien</i> 'how many/ much'	Quantifier	quantity

Question words in French are classified as either simple or complex forms (Riegel, Pellat & Rioul 2009: 384). Simple forms have single constituent, while complex forms are compositions of two constituents, as previously shown in Mandarin. Table 16 displays simple form of question words.

As a fronting language, question words in French are placed initially, which can make it difficult to determine the syntactic role of the question word *qui*. This ambiguity arises when analysing whether ***qui* 'who'** is a subject or a complement that is required to be fronted (Abeillé & Godard 2021: 1111), as illustrated in (2.65).

- (2.65) *Qui choisira Marie ?*
 who choose-FUT Mary
 'Who will choose Marie?'
 'Whom will she choose?'

The example (2.65) can be interpreted as *Qui choisira Marie* 'Who will choose Mary?', in which the question word *qui* 'who' is the subject. On the other hand, it can also be

interpreted as *Qui choisira-t-elle?* ‘Whom will she choose?’, in which the question word *qui* ‘who’ is the complement of the verb *choisir* ‘choose’, and *qui* ‘who’ moves to the initial position. Accordingly, the context becomes essential to solve such ambiguity. In addition to the context, the subject-verb agreement can be another solution, as in (2.66).

- (2.66) **Qui** choisira ces enfants?
 who choose.FUT DEM children
 ‘Who will choose those children?’

The verb form used in this example is conjugated with the third person singular noun, *qui* ‘who’ in this case, which clarifies that the question word in (2.66) is the subject.

Contrasting with the question word *qui* ‘who’, the question word for the object has two forms, namely **que** and **quoi** ‘what’. Although they have the same semantic scope, they differ in their phonetic and syntactic characteristics (Druetta 2018: 36; Abeillé & Godard 2021: 986, 1111–1112), as shown in (2.67). The question word *que* is phonetically weak and cannot be placed at the end of the sentence (Boucher 2010a: 61), as in (2.67b). Furthermore, the word *que* ‘what’ syntactically depends on verbs and is only used to inquire about direct objects. Consequently, it imposes the postposition of pronominal and nominal subjects as in (2.67a) and (2.67c).

- (2.67) a. **Que** mangez-vous ?
 what eat-2PL
 ‘What do you eat?’
 b. **Vous* mangez **que** ?
 2PL eat what
 Intended for: ‘What do you eat?’
 c. **Que** mange Paul ?
 what eat Paul
 ‘What does Paul eat?’

The word *quoi* is phonetically strong and is an allomorph of *que* (Boucher 2010a: 61; Coveney 2011: 128), as in (2.68). Its phonetic nature makes it occur mostly at the end of sentences, as shown in (2.68a). It can only be found in the initial position when preceded by a preposition, as shown in (2.68b), or when modified, as in (2.68c).

- (2.68) a. *Vous* mangez **quoi** ?
 2PL eat what
 ‘What do you eat?’

- b. *À quoi penses-tu ?*
 PREP what think-2SG
 ‘What are you thinking about?’
- c. *Quoi d’autre pourrait vous satisfaire ?*
 what PREP other can.COND 2PL satisfy
 ‘What else could satisfy you?’ (Abeillé & Godard 2021: 1112)

The distribution of the question words *que* and *quoi* is syntactically complementary (Coveney 2011: 128). It is often simplified as *que* is typically placed before verbs, and *quoi* after verbs, as shown in (2.67a), (2.67c) and (2.68a). However, this rule may not apply in cases such as (2.68b) and (2.68c). *Quoi* appears preverbally if the subject is omitted from questions (Blanche-Benvéniste 1997: 133), as in (2.69), in the infinitive structure in (2.69a) or present participle in (2.69b).

- (2.69) a. *Quoi faire ?*
 what do.INF
 ‘What to do?’ (Blanche-Benvéniste 1997: 133)
- b. *En quoi faisant ?*
 APRT what do.APRT
 ‘By doing what?’ (Blanche-Benvéniste 1997: 133)

On the other hand, the question word *que* appears to have more restricted use in the initial position, as shown in examples (2.70). It becomes ungrammatical when separated from the verb, as in (2.70a) and (2.70b) (Coveney 2011: 128). Additionally, the question word *que* cannot function as a subject by itself in the initial position, as shown in (2.70c). However, it becomes grammatical if I add a particle *est-ce qui*, as in (2.70d) (Munaro & Pollock 2008: 561).

- (2.70) a.* *Que les autres font-ils ?*
 what ART other do-3PL
 Intended for: ‘What the others do?’ (Coveney 2011: 128)
- b.* *Que tu vois ?*
 what 2SG see
 Intended for: ‘What do you see?’ (Blanche-Benvéniste 1997: 133)
- c.* *Que flotte dans l’eau ?*
 what float PREP ART water
 Intended for: ‘What is floating in the water?’ (Shlonsky 2017: 351)
- d. *Qu’est-ce qui flotte dans l’eau ?*
 what PTCL float PREP ART water
 ‘What is floating in the water?’

Regarding question words in the category of adverbs (i.e., *où* ‘where’, *quand* ‘when’, *pourquoi* ‘why’, *comment* ‘how’), Abeillé & Godard (2021: 1113) argue that *où* ‘where’ should be classified as a preposition. However, I will not delve further into their argumentation distinguishing adverbs from prepositions as it is of little relevance here (cf. Abeillé & Godard 2021: 753–754 for further discussion). Traditionally, these four question words are categorised as adverbs because they are invariant (Riegel, Pellat & Rioul 2009: 385). In addition, adverbials such as time, place, reason, and manner can substitute for them. Also note that the *comment* ‘how’ can also function as an attribute, as in (2.71) (Grevisse & Goosse 2007: 487).

- (2.71) *Comment* *est* *ton* *fiancé ?*
how COP POSS.2SG fiancé
‘How is your fiancé?’

As for their syntactic structures, it is noteworthy that they exhibit more limitations on the types of inversion when compared to other question words. As illustrated in Section 2.2.2, constituent questions allow for three types of inversion, while polar questions do not permit stylistic inversion. Interestingly, the question word *pourquoi* ‘why’ is like polar questions in that it also does not allow for stylistic inversion, as shown in (2.72) (Korzen 1990: 60; Coveney 2011: 129; Abeillé & Godard 2021: 1114).

- (2.72) * *Pourquoi* *sont* *partis* *les* *autres ?*
why COP leave.PTCP ART other
Intended for: ‘Why did the others leave?’ (Coveney 2011: 129)

Likewise, the unacceptability of stylistic inversion is probably demonstrated by the causal meaning of the question word *comment* ‘how’ (Cornulier 1974: 141), as in (2.73), where the question from A is answered as if it was seeking the cause.

- (2.73) A: ? *Comment* *a* *survécu* *Jean ?*
how have survive.PTCP John
Intended for: ‘How come did John survive?’
B: *C’* *est* *qu’* *il* *est* *résistant.*
DEM COP COMP 3SG COP resistant
‘It is because he is resistant.’ (Cornulier 1974: 141)

The incompatibility may be because both question words have a causal feature and are syntactically in the same position as other causal adverbs, outside the scope of the proposition (cf. Cornulier 1974; Korzen 1990 for further explanation).

Unlike invariant question words used as adverbs, the question word *quel* ‘which’ varies depending on the gender and number of the following nouns (Abeillé & Godard 2021: 1112, 2312). There are four different forms of *quel* which are distinguished only in writing because of their identical pronunciation. These forms are *quel*, *quels*, *quelle*, and *quelles* (Druetta 2018: 38), as in examples (2.74). The question word *quel* in (2.74a) agrees with the masculine noun *jour* ‘day’. The question word *quelle* in (2.74b) agrees with the feminine noun *température* ‘temperature’.

- (2.74) a. ***Quel*** *jour* *sommes-nous* ?
 which day COP-1PL
 ‘What day is today?’
- b. ***Quelle*** *est* *la* *température* ?
 which COP ART temperature
 ‘What is the temperature?’

As shown in the two examples above, the question word *quel* functions not only as a determiner as in (2.74a) but also as an adjective as in (2.74b). These two functions of *quel* differ in terms of the nature of the entities being questioned. Essentially, both functions of *quel* ‘which’ inquire about a set of entities that is presupposed in the discourse (Abeillé & Godard 2021: 1112–1113, 2312). However, as in (2.75), the determiner *quel* is used for questions regarding the type in (2.75a), whereas the adjective *quel* is used more for properties, as in (2.75b) (Abeillé & Godard 2021: 2312).

- (2.75) a. ***Quel*** *arbre fleurit en* *hiver*? (Quel determiner)
 which tree bloom PREP winter
 ‘Which tree blooms in winter?’
- b. ***Quel*** *est le nom d’* *entreprise* ? (Quel adjective)
 which COP ART name PREP company
 ‘What is the name of the company?’

Another difference between the determiner *quel* and the adjective *quel* lies in their preference for inversion (Grevisse & Goosse 2007: 494–495). The determiner *quel* favors complex inversions, while the adjective *quel* favors stylistic and subject-clitic inversions. This tendency in the adjective *quel* is due to its similarity to *que* ‘what’ as a weak form in syntax, where no intervention is allowed between the question word and the verb. In addition, both adjective *quel* and *que* ‘what’ do not appear in the final position (Abeillé & Godard 2021: 1112).

The word *quel* can also be used in combination with definite articles (i.e., *le*, *la*, *les*) to form complex question words. Likewise, they vary accordingly based on gender and

number of definite articles (e.g., *laquelle* for feminine singular and *lesquels* for masculine plural). Sharing the same semantic scope with the single-form *quel*, the complex forms have an extra deictic function derived from definite articles. This function refers to the nominal elements in the preceding context (Riegel, Pellat & Rioul 2009: 385). Accordingly, complex forms agree with their anaphoric nominals in number and gender (Abeillé & Godard 2021: 1110). As shown in (2.76), the feminine form of *laquelle* refers to its anaphoric feminine noun *plusieurs robes* ‘many dresses’. Note that the number does not agree here because of the context.

- (2.76) *J’ ai plusieurs robes. Laquelle dois-je mettre ?*
 1SG have many dress which.of.them should-1SG put
 ‘I have many dresses. Which one of them should I put on?’

Finally, the question word ***combien*** ‘how much/many’ functions as a quantifier, seeking information about quantity rather than intensity or frequency, as in (2.77). The requested information is provided in the form of number or quantity adjectives (e.g., *few*, *some*, *half*).

- (2.77) ***Combien*** *il nous faut ?*
 how.much/many 3SG 1PL need
 ‘How much do we need?’

On the other hand, *combien* ‘how much/many’ is also categorised as an adverb of degree, as compared to others, such as *beaucoup* ‘many’, *peu* ‘little’, and *tant* ‘so much’ (Abeillé & Godard 2021: 2274). Like the construction of *quel* determiner, the question word *combien* ‘how much/many’ can also be followed by nouns, but they are required to be introduced by a preposition, as in (2.78). Therefore, the requested quantity is indicated by this nominal structure, for example, *combien d’huile* ‘how much oil’.

- (2.78) ***Combien*** *d’ huile il nous faut ?*
 how.much/many PREP oil 3SG 1PL need
 ‘How much oil do we need?’

The brief overview aims to help understand the position and syntactic nature of question words in French. Except the case where question words used as subjects, French question words normally have tendency to move to the initial position (Grevisse & Goosse 2007: 502). The structure can be marked with different types of inversions or remain unchanged (i.e., a declarative word order), as demonstrated in Section 2.2.2. Note that detailed discrepancies of the type of inversion may occur. For instance, *pourquoi*

‘why’ seems to have difficulty with stylistic inversion, while *quand* ‘when’ might be unacceptable when followed by a declarative word order (e.g., *quand ça commence?* Literal translation: ‘When it starts?’) (Coveney 2011: 130). In fact, various factors (e.g., verb form, type of pronouns) also play a role in determining the matches between question words and type of inversion (cf. Grevisse & Goosse 2007: 492 for more details). Overall, the main characteristic of French constituent questions is the fronted question words. However, French has another option besides fronting its question words (Abeillé & Godard 2021: 1110), namely *wh*-in-situ. This is one of the variants of constituent questions in French, and it seems to be popular in spoken language (Grevisse & Goosse 2007: 502). It has been frequently attested in modern French since the late 19th century (Coveney 2011: 125; Guryev 2017: 99).

To conclude this section, the question words in Mandarin and French belong to similar word classes and semantic scopes, but the distribution of question words is systematically different in both languages. In Mandarin, question words are placed according to their syntactic role, such as the subject in the initial position and the object towards the end of the clause. Additionally, the word class of question words determines the placement of question words used as adverbs, such as 怎麼 *zěnmě* ‘how’ and 為什麼 *wèishénme* ‘why’. In contrast, French question words are unanimously fronted regardless of their word class or syntactic role. However, the phenomenon of *in-situ* question words appears to be developing in the contemporary French. Therefore, the following two sections will discuss the syntactic mechanism of *wh*-in-situ in Mandarin and French, respectively. This analysis will enable us to explore the similarities and differences in the syntactic processes of *in-situ* question words.

2.4.3 *Wh*-in-situ in Mandarin

Generative grammar in the past several decades has heavily relied on its principles and rules to account for the diversity of linguistic phenomena. One of the essential operations in this field is syntactic movement (Li & Wang 2018: 1277). Following the development of the Minimalism program, movement is now considered a ‘last resort’ and should only take place for formal reasons (Lasnik 2002: 434). Despite the various frameworks of generative syntax, the concept of movement remains relevant when discussing related topics.

As stated at the beginning of Section 2.4, question words are generally categorised into three tendencies based on their positions in questions: fronting, *in-situ*, and both. The discussion of questions in languages, such as English, French, and other fronting languages, often involves the concept of movement. Their question words are placed in the initial position as in (2.79), while their answers are usually postverbal as in (2.79).

(2.79) a. *What* does he eat?

b. He eats a *hamburger*.

Accordingly, these question words appear to undergo movement, also known as *wh*-movement in generative terms.

However, linguists have noted that some languages, such as Mandarin and Japanese, do not follow the same rule as fronting languages. In these languages, question words remain in their original position, meaning that the concept of movement is not relevant. From a generative perspective, these *in-situ* languages do not involve any syntactic movement of question words (Cheng 1991: 19). However, if generative grammar aims to use its principles to explain languages cross-linguistically, it may not be comprehensive and accurate to claim that no *wh*-movement is applied in *in-situ* languages. In fact, such movement is required to denote a proposition with interrogative reading. In other words, question words must be moved to the specifier position of complementiser phrase (CP) to take the scope of the whole proposition. Consequently, linguists, such as Huang (1982), Watanabe (1992; 2001), Cheng & Rooryck (2000), Simpson & Bhattacharya (2003), Tomioka (2007), Shlonsky (2017), Bayer & Cheng (2017), Shlonsky, Villata & Franck (2020), have attempted to explain the syntactic mechanism of *in-situ* languages, and to see if these languages also adhere to the same principle to obtain an interrogative reading.

Several theories are put forth for Mandarin (Lin 2014: 193; Cheng 2009; Pan 2022), including the quantificational theory (Huang 1982) and the unselective binding theory (Aoun & Li 1993; Tsai 1994b). These theories differ fundamentally in terms of movement. The purpose of this section is to provide an overview of how theories adopt and explain the use of *in-situ* question words in Mandarin. The focus is not on comparing theories or proposing new accounts for *wh*-in-situ in Mandarin.

The quantificational theory assumes that every question word is a quantifier. The interrogative interpretation relies on the movement of question words to the initial position.

(2.80) 張三 看見 了 什麼？
Zhāngsān kànjiàn le shénme?
 Zhangsan see ASP what
 ‘What did Zhangsan see?’

(2.81) [什麼_i [張三 看見 了 *t_i*]]?
 [shénme_i [Zhāngsān kànjiàn le *t_i*]]?
 What Zhangsan see ASP

Accordingly, the question word 什麼 *shénme* ‘what’ in (2.80) moves to a higher position to obtain the interrogative interpretation and the scope of the proposition, as shown in (2.81). Unlike English, which overtly moves question words in the Syntax level, Mandarin relies on a covert movement in the Logical Form level (i.e., a level where represents the interpretation of a sentence (Pan 2022: 2), also known as the LF movement (Huang 1982; Huang, Y.-H. Audrey Li & Li 2009: 261–273).

The LF movement assumption in Mandarin is deduced from the selection requirement of matrix verbs and the constraints on the distribution and interpretation of constituent questions. These requirements and constraints likewise impose on question words in English, which undergo overt movement.

The selection requirement states that different matrix verbs select different types of complements. There are three groups of matrix verbs. The first group includes the verbs *think* and *believe*, which cannot have questions as their complements (e.g., **John thinks what Mary bought*). In other words, the entire proposition should be interpreted as a direct question (e.g., *What does John think Mary bought?*). Second group, the verbs *wonder* and *ask* need questions as complements and the propositions can be understood as indirect questions (e.g., *John wonders what Mary bought.*). Finally, the verbs *remember* and *know* can be followed by both types of complements (e.g., *John knows what Mary bought. / What does John know Mary bought?*) (Huang 1982: 255; Huang, Yen-Hui Audrey Li & Li 2009: 262; Lin 2014: 181).

Mandarin is similar to English in this respect, but with a difference in syntactic structures. English is explicit, while Mandarin is implicit in interpretation because of the *in-situ* question words, as shown in (2.82).

(2.82) a. 張三 相信 誰 買 了 書？
Zhāngsān xiāngxìn shéi mǎi le shū?
 Zhangsan believe who buy ASP book
 ‘Who does Zhangsan believe bought books?’

- b. 張三 問 我 誰 買 了 書。
Zhāngsān wèn wǒ shéi mǎi le shū.
 Zhangsan ask 1SG who buy ASP book
 ‘Zhangsan asks who bought books’
- c. 張三 知道 誰 買 了 書 (?)
zhāngsān zhīdào shéi mǎi le shū(?)
 Zhangsan know who buy ASP book
 ‘Who does Zhangsan know bought books?’
 ‘Zhangsan knows who bought books’ (Huang 1982: 254)

One possible explanation for this similarity is the assumption of movement in Mandarin. Therefore, the observed phenomenon in the selection requirement can be attributed to the syntax movement in English and LF movement in Mandarin.

Constraints in the distributions of different types of question words and interpretations of constituent questions have also been studied. Special attention has been given to question words used as adjuncts, such as 為什麼 *wèishénme* ‘why’. This *wh*-adjunct cannot be extracted from syntactic islands, which are structures that prevent a word inside them from being extracted. These islands often contain complex noun phrases (e.g., a noun phrase modified by relative clauses), coordinate structures, sentential subject, and left branches (e.g., a noun phrase on the left branch of another larger noun phrase) (Boeckx 2012: 5) In other words, these structural domains limit certain syntactic operations (Ross 1967: 487; Boeckx 2012: x). Accordingly, syntactic islands impose constraints on forming direct questions of *wh*-adjuncts in Mandarin, as illustrated in the two examples in (2.83) (Huang 1982: 525–526).

- (2.83) a. 你 想 知道 誰 為什麼 買 了 書 ?
nǐ xiǎng zhīdào shéi wèishénme mǎi le shū ?
 2SG want know who why buy ASP book
 ‘Who is the person x such that you wonder why x bought books?’
- b. 你 想 知道 誰 怎麼 買 了 書 ?
nǐ xiǎng zhīdào shéi zěnmē mǎi le shū
 2SG want know who how buy ASP book
 ‘Who is the person x such that you wonder how x bought books?’

Both examples in (2.83) only permit direct question reading of the question word 誰 *shéi* ‘who’, but not of the question words 為什麼 *wèishénme* ‘why’ and 怎麼 *zěnmē* ‘how’. This constraint is due to the *wh*-island. In other words, the interpretation of asking *why*...or *how*...is never obtained once *wh*-adjuncts are placed inside a syntactic island.

The asymmetry between adjuncts and arguments is further highlighted by the constraint on *wh*-adjuncts. *Wh*-arguments seem to be easier to move out from *wh*-islands in English, as in (2.84). Moreover, no island effect is even detected in Mandarin, as in (2.85). Both *wh*-arguments have a grammatical interpretation in direct questions.

(2.84) ? *What_i did you wonder [how to fix t_i]?*

(Huang, Y.-H. Audrey Li & Li 2009: 263)

(2.85) 你 想 知道 誰 買 了 什麼 ?
 nǐ xiǎng zhīdào shéi mǎi le shénme?
 2SG want know who buy ASP what

‘What is the thing x such that you wonder who bought x?’

‘Who is the person x such that you wonder what x bought?’ (Huang 1982: 525)

To address the asymmetry between arguments and adjuncts in the two languages with their movement settings, the Empty Category Principle (ECP) is attributed to both overt and covert movements. This principle necessitates an empty category, such as a trace left by a moved *wh*-element, to be well-governed by a lexical head or an antecedent. If *wh*-elements are arguments, the empty category (i.e., trace) can be properly governed by a lexical head. However, adjuncts need to be governed by antecedents, and this government appears impossible when crossing syntactic island boundaries. Therefore, *wh*-adjuncts would never receive a direct question interpretation if they were located in syntactic islands.

The asymmetry is also observed within *wh*-adjuncts in Mandarin. Huang (1982: 527–528) has acknowledged that the island effect disappears once the *wh*-adjuncts are paraphrased as prepositional phrases, such as 為了什麼原因 *wèile shénme yuányīn* ‘for what reasons’ and 用什麼 *yòng shénme* ‘in what manner/ by what means’. In addition, 哪裡 *nǎlǐ* ‘where’ and 什麼時候 *shénmeshíhou* ‘when’ can also remain intact in the island. The asymmetry within adjuncts may be attributed to the contrast between adjuncts that are noun phrase (NP) (e.g., 為了什麼原因 *wèile shénme yuányīn* ‘for what reasons’) and non-NP adjuncts (e.g., 為什麼 *wèishénme* ‘why’) (Huang 1982: 530).

Another asymmetry is found with the adjunct 怎麼樣 *zěnmeyàng* ‘how’. It can be saved from island effects (Lin 2014: 184) and be used as complements of direct questions for almost any verbs that can take direct questions as complements. However, 為什麼 *wèishénme* ‘why’ can only follow ‘semantically-bleached’ verbs (e.g., 說 *shuō* ‘say’) (Lin

1992; Lin 2014: 184–185). In my view, the *wh*-adjunct 怎麼樣 *zěnmeyàng* ‘how’ can also be categorised as an adjunct NP. The word 樣 *yàng* refers to the type of an entity and is also associated with deictic pronouns, such as 這樣 *zhèyàng* ‘this kind’. Therefore, the asymmetry between 怎麼樣 *zěnmeyàng* ‘how’ and 為什麼 *wèishénme* ‘why’ is still related to the nominal property of adjuncts.

To account for the asymmetry within *wh*-adjuncts, Tsai (1994a: 125) argues for the referentiality of adjuncts in addition to the nominal property. *Wh*-adjuncts can be distinguished by their different readings. The question word 怎麼 *zěnmé* ‘how’ has manner and means/instrumental meanings, while 為什麼 *wèishénme* ‘why’ has reason and purpose meanings. The asymmetry found within *wh*-adjuncts can be attributed to the different meanings of each question word. Referential adjuncts, such as 怎麼 *zěnmé* ‘how’ with instrumental meaning and 為什麼 *wèishénme* ‘why’ with purpose, can be extracted from the islands. However, extracting non-referential adjuncts, such as 怎麼 *zěnmé* ‘how’ with manner meaning and 為什麼 *wèishénme* ‘why’ with reason, is difficult.

Huang (1982) argues that *in-situ* question words in Mandarin undergo covert movement, as is observed the overt movement in fronting languages, due to the correlation between fronting and *in-situ* question words highlighted by the selection requirement and island effect. However, this explanation is not entirely satisfactory theoretically or empirically (Huang, Y.-H. Audrey Li & Li 2009: 267). Theoretically, it has been argued that certain constraints present in overt movements are not attested in covert movements. For instance, in Mandarin, arguments can avoid different island effects, whereas in English, this is not possible (Huang, Y.-H. Audrey Li & Li 2009: 267). To account for this discrepancy, Huang (1982) further argues that overt movements in English must follow not only the ECP principle but also the subjacency and constraint of Condition on Extraction domain (CED), while covert movements in Mandarin are only sensitive to the ECP principle. The observed discrepancy between overt and covert movements lies in the fact that they are subject to different constraints.

Empirically, some instances of LF movement do obey subjacency constraints. Therefore, studies support the application of these constraints to both overt and covert movements. In the case of *wh*-arguments in Mandarin that are not affected by islands, studies analyse

them from a different perspective, specifically the pied-pipping mechanism (cf. Nishigauchi 1986 and Fiengo et al. 1988 for more technical details).

If the concept of movement causes disagreements in syntactic manipulation, one alternative is to argue against movement by using unselective binding for *in-situ* question words. Admittedly, there is also a mixed approach that combines LF movement and unselective binding.

This alternative is proposed by Aoun & Li (1993) and Tsai (1994b), namely the **unselective binding theory**. According to this theory, question words do not move in the LF level, which avoids the constraints caused by movements. The argument for no LF movement is based on the interaction between focus words (e.g., *only*) and question words (Aoun & Li 1993: 206). The focus word *only* must be associated with its lexical item in the same domain (Jackendoff 1974; Rooth 1985; Tancredi 1990). In other words, the lexical item associated with focus words cannot be topicalised or moved out of the domain governed by the focus word, as shown in examples (2.86). This is also known as the Principle of Lexical Association (Tancredi 1990, as cited in Huang, Y.-H. Audrey Li & Li 2009: 274).

(2.86) a. **Mary_i, he only likes t_i.*

b. **Who_i does he only like t_i?*

Such a principle of association is applicable in Mandarin as well, as shown in examples (2.87).

(2.87) a. 他 只 喜歡 誰?
tā zhǐ xǐhuān shéi?
 3SG only like who
 ‘Who does he only like?’

b. 他 只 喜歡 在 哪 看書?
tā zhǐ xǐhuān zài nǎ kànshū?
 3SG only like PREP where read
 ‘Where does he only like to read?’

These examples suggest that question words in Mandarin do not move in the LF level. Otherwise, the lexical association cannot be established between the focus word 只 *zhǐ* ‘only’ and its associated question words.

However, to some extent, question words in Mandarin do not function the same way as in English. There is no movement involved in Mandarin. In addition, question

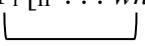
words in Mandarin have no quantificational force in nature (Tsai 1994b: 59) and are not necessarily interpreted as question words. They are lexically undetermined, serving as variables bearing “quantificational variability effects” (Huang, Y.-H. Audrey Li & Li 2009: 275). The role of these variables is determined by their syntactic environments (Cheng 1991: 123; Huang, Y.-H. Audrey Li & Li 2009: 274). In quantificational or conditional settings, question words are indefinite, such as in (2.88) and (2.89). In an interrogative setting (e.g., with question-related markers, such as 呢 *ne* particle), question words are interpreted as interrogative expressions, as in (2.90).

- (2.88) 他 什麼 都 喜歡。
tā shénme dōu xǐhuān.
 3SG what all like
 ‘He likes everything.’

- (2.89) 如果 你 想要 什麼， 我 就 送 你。
rúguǒ nǐ xiǎngyào shénme wǒ jiù sòng nǐ.
 if 2SG want what 1SG then give 2SG
 ‘If you want something, I will buy it for you’

- (2.90) 你 喜歡 什麼 呢？
nǐ xǐhuān shénme ne?
 2SG like what PTCL
 ‘What do you like?’

As question words are determined by the syntactic settings, they must be bound by a syntactic operator. In an interrogative setting, such a question operator (QOP) in (2.91) provides an interrogative interpretation to these words, making them question words. In Mandarin, question particles 嗎 *ma* and 呢 *ne* are considered operators (Aoun & Li 1993: 210–211). The schema below shows the binding relation between an operator and a *wh*-element.

- (2.91) [CP QOP_i [IP . . . *wh*_i . . .]]

 (Tsai 1994b: 17; Huang, Y.-H. Audrey Li & Li 2009: 276)

Tsai (1994b: 18–23) has pointed out English has the same undetermined *wh*-elements. For instance, the English existential-reading indefinite *somewhat* is composed of the existential prefix *some*- and *what*, while the universal-reading indefinite *whatever* is composed of the universal suffix *-ever* and *what*. Therefore, English and Mandarin *wh*-elements have the quantificational variability effects. This effect distinctively pertains to the morphological level in English, whereas it pertains to the syntactic level in Mandarin.

The unselective binding theory shows *in-situ* question words do not require movement. It is particularly relevant in the context of the minimalism program, as it is more economic due to the absence of movement (Tsai 1994b: 17).

While a binding theory may appear to be an ideal way to explain the syntactic mechanism of *wh-in-situ* languages, it must also account for the argument/adjunct asymmetry and selection requirements addressed and resolved within the framework of LF movement. To maintain the selection requirement based on a binding relation, it is essential to identify the question operators in the appropriate specifier position within CP. This enables the question words *in-situ* to be bound by the QOP.

- (2.92) a. [QOP_i [張三 以為 [[李四 買 了 什麼 _i]]]] ?
 [QOP_i [Zhāngsān yǐwéi [[Lǐsì mǎi le shénme_i]]]] ?
 zhangsan think Lisi buy ASP what
 ‘for which x: x a thing, Zhangsan thinks Lisi bought x?’
- b. *[[張三 以為 [QOP_i [李四 買 了 什麼 _i]]]] 。
 *[[Zhāngsān yǐwéi [QOP_i [Lǐsì mǎi le shénme_i]]]] .
 ‘Zhangsan thinks [for which x: x a thing, Lisi bought x].’

(Huang, Y.-H. Audrey Li & Li 2009: 278)

Examples (2.92) demonstrate that the verb 以為 *yǐwéi* ‘think’ cannot take a question as its complement. The entire proposition must take the wide scope (i.e., direct question reading). Therefore, it is reasonable to postulate that a binding relation exists between an operator found in the higher specifier position of CP and the *in-situ* question word 什麼 *shénme* ‘what’. For the LF movement, it is argued that question words are moved covertly. The binding theory, on the other hand, proposes that operators are situated in a higher position to obtain the scope, which depends on the selection requirements of verbs. Therefore, the binding theory suggests that all question words remain *in-situ* in the LF. The scope of interpretation is easily obtained through the binding relation between operators and *wh*-elements.

However, accounting for the argument/adjunct asymmetry in the binding theory, particularly the island effects imposing on *wh*-adjuncts, is more challenging. Two possible solutions have been proposed. One is to admit LF movement is still applied along with the unselective binding theory (Tsai 1994b: 61). This standpoint assumes that *wh*-adjuncts are themselves operators, requiring them to move either overtly in English or covertly in Mandarin. This approach is supported by the fact that *wh*-adjuncts, such as 怎

麼 *zěnmě* ‘how’ and 為什麼 *wèishénme* ‘why’, do not have an existential interpretation, as shown in examples (2.93) (Huang, Y.-H. Audrey Li & Li 2009: 279).

- (2.93) a. * 他 會 為什麼 很 好運 嗎?
 * *tā huì wèishénme hěn hǎoyùn ma?*
 3SG will why very lucky PTCL
 Intended for: ‘Will he get lucky for some reason?’
- b. ? 他 會 怎麼 修車 嗎?
 ? *tā huì zěnmě xiūchē ma?*
 3SG will how fix.cars PTCL
 Intended for: ‘Will he fix cars in some way?’

Accordingly, these *wh*-adjuncts only have an interrogative interpretation. Unlike *wh*-arguments, which function as variables bound by operators, adjuncts need to move in LF in Mandarin. However, this mixed approach encounters some empirical issues. In ‘bare-conditional’ sentences (i.e., conditional sentences without a leading element (e.g., *if*) in the antecedent clause (Cheng & Huang 1996: 127), *wh*-adjuncts seem to be considered as variables and are unselectively bound by an appropriate QOP, as *wh*-arguments do (Huang, Y.-H. Audrey Li & Li 2009: 280–281) (see Cheng & Huang 1996 for further discussion on bare-conditional sentences).

Given the analysis of movement frequently encounters other issues, another solution is to consider the asymmetry with the unselective binding theory. This theory argues that *wh*-adjuncts do not move in LF. However, the operator of *wh*-adjuncts does move, as shown in the following schema, as in (2.94) (Huang, Y.-H. Audrey Li & Li 2009: 281).

- (2.94) [CP QOP_i [IP 他 認為 [CP t_i [IP 李四 為什麼_i 沒 來]]] 呢] ?
 [CP QOP_i [IP *tā rènwéi* [CP t_i [IP *lǐsì wèishénme_i méi lái*]]] *ne*] ?
 3SG think Lisi why NEG come PTCL
 Movement **Binding**

The unselective binding theory appears to provide a better explanation for the *wh-in-situ* phenomenon in Mandarin. This theory aligns with the idea that movement should be a last resort and aims for a more economic explanation. The covert LF movement offers another approach to study *in-situ* question words and highlights several cross-linguistic differences between English and Mandarin. In addition to the two theories presented above which elucidate *wh-in-situ* in Mandarin, other approaches are put forth, such as clausal typing (Cheng 1991), the non-movement cartography template (Yuan

- (2.97) * *La temperature est quelle ?*
 ART temperature COP.3SG which
 Intended for: ‘What is the temperature?’ (Abeillé & Godard 2021: 1112)

Finally, *pourquoi* ‘why’ is rarely found *in-situ*, as in (2.98).

- (2.98) ? *Jean saigne pourquoi ?*
 John bleed why
 Intended for: ‘Why does John bleed?’ (Beyssade 2006: 181)

Pourquoi ‘why’ is not entirely unacceptable compared to the cases in (2.95) and (2.96). Furthermore, *pourquoi* ‘why’ *in-situ* becomes more frequent in spoken French (Coveney 2011: 130), as shown in (2.99).

- (2.99) *Ils votent pour Le Pen pourquoi ?*
 3PL vote for Le.Pen why
 ‘Why do they vote for Le Pen?’ (Coveney 2011: 130)

As for question words functioning as non-subject, *in-situ* seems to be another option in addition to fronted, as in (2.100).

- (2.100)a. *Jean a invité qui ?*
 John have invite.PTCP who
 ‘Who did John invite?’ (Beyssade 2006: 180)
- b. *Tu préfères quoi ?*
 2SG prefer what
 ‘What do you prefer?’ (Abeillé & Godard 2021: 1111)
- c. *Tu es où ?*
 2SG COP where
 ‘Where are you?’ (Fairon, Simon & Grevisse 2018: 889)
- d. *Tu viens quand ?*
 2SG come when
 ‘When do you come?’ (Fairon, Simon & Grevisse 2018: 889)
- e. *Cet argent sera distribué comment ?*
 DEM money COP.FUT distribute.PTCP how
 ‘How will the money be distributed?’ (Grevisse & Goosse 2007: 502)
- f. *Vous avez pris quel train ?*
 2PL have take.PTCP which train
 ‘Which train did you take?’ (Abeillé & Godard 2021: 2312)
- g. *Finalement, on était combien ?*
 finally 1PL COP.IMPARF how.much/many
 ‘Finally, how many of us were there?’ (Abeillé & Godard 2021: 1114)

- h. *Tu dépenses combien de litres [...] ?*
 2SG spend how.much/many PREP litre
 ‘How many litres do you spend?’ (Grevisse & Goosse 2007: 502)

Nevertheless, the tendency to remain *in-situ* varies among question words in spoken French, depending on the syntactic relationship between the question words and verbs (Druetta 2018: 42–43). As explained in Section 2.4.2, *quel* ‘which’ in French has two functions: adjective and determiner. The example (2.97) shows that *quel* ‘which’ adjective is an attribute of the verb and cannot be placed *in-situ*. *Quel* ‘which’ determiner followed by the noun *train* ‘train’ in (2.100f) can be *in-situ*, as *quel train* ‘which train’ is the complement of the verb.

Likewise, as in examples (2.101), question *comment* ‘how’ used as attribute of the verb can be optionally *in-situ* or fronted (Druetta 2018: 41–42), as in (2.101a) and (2.101b). When it is used as an adverbial, it tends to be fronted, as in (2.101c).

- (2.101)a. *Les gens ils sont comment ?*
 ART people 3PL COP how
 ‘The people, how are they?’ (Druetta 2018: 42)
- b. *Comment était l’orchestre ?*
 how COP.IMPARF ART orchestra
 ‘How was the orchestra?’ (Druetta 2018: 42)
- c. *Comment ça s’écrit ?*
 how that REFL write
 ‘How do you write it?’ (Druetta 2018: 42)

Figure 9 adopted and simplified from Druetta (2018: 43) illustrates the tendency of question words to be initial or *in-situ*. This figure represents a continuum of tendency. Therefore, it implies the likelihood of initial placed question words to be found *in-situ*, except for illegitimate cases shown in (2.95), (2.96), and (2.97).

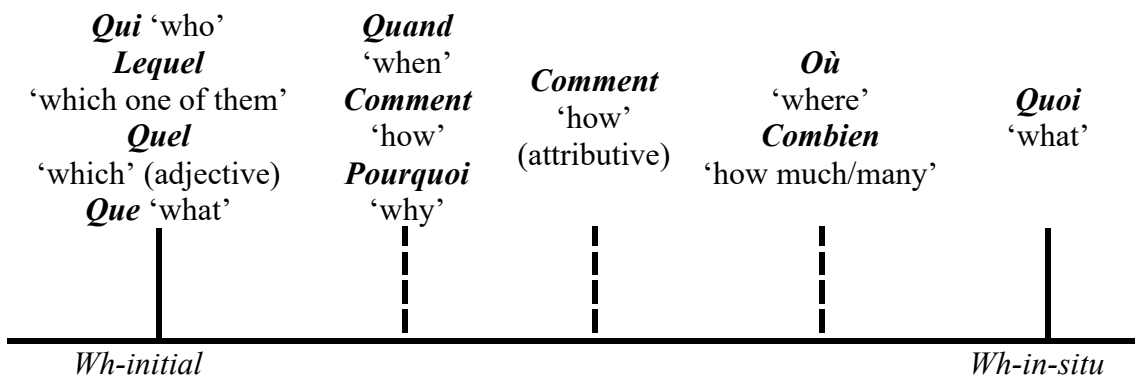


Figure 9 Distributional tendency of question words (Druetta 2018: 43)

The semantic interpretation of question words also plays an essential role in determining distributional tendency. This is particularly relevant for the two adverbial question words, *comment* ‘how’ and *pourquoi* ‘why’.

According to Aoun (1985: 27–28), based on personal communication with Dominique Sportiche, only *comment* ‘how’ with instrumental reading can be found *in-situ*, as shown in examples (2.102). In a question-response pair, *comment* ‘how’ in (2.102a) should be answered by the response in (2.102a’), rather than by that in (2.102a’'). This is because the *in-situ comment* ‘how’ should be interpreted with an instrumental reading.

- (2.102) a A : *Tu as ouvert la porte comment ?*
 2SG have open.PTCP ART door how
 ‘How did you open the door?’
 a’ B: *Avec une clef.*
 ‘With a key.’
 a’’ B: *Lentement.*
 ‘Slowly’ (Aoun 1985: 27)

Similarly, the question word *pourquoi* ‘why’ can only be interpreted with purposive reading when it is *in-situ*, as shown in (2.103).

- (2.103) a A: *Tu es venu pourquoi ?*
 2SG COP come.PTCP why
 ‘Why did you come?’
 a’ B: *Pour étudier la géométrie.*
 ‘To study geometry.’
 a’’ B: *Parce que je suis malade.*
 ‘Because I am sick.’ (Aoun 1985: 27)

However, to reply with a reason as in (2.103a’’) the *in-situ* question word in (2.103a) needs to be paraphrased with *pour quelle raison* ‘for what reason’ (Coveney 2011: 130), as shown in (2.104). This echoes the asymmetry of *wh*-adjuncts in Mandarin (referential vs. non-referential) as mentioned in Section 2.4.3.

- (2.104) *Tu es venu pour quelle raison ?*
 2SG COP come.PTCP PREP which reason
 ‘What is your reason for coming?’

In addition to the distributional tendency of each question word, I am mainly interested in how studies syntactically account for the optional *in-situ* phenomenon in

French. There are mainly three, each distinguished by the involvement of movement or not: covert movement, remnant movement, and unselective binding (Boucher 2010a: 69).

(2.105) [CP Ø [+WH] [IP *Tu* [I' *vas* *où*]]]
 2SG go where

- (2.106)a. ?* *Jean ne mange pas quoi ?*
 John NEG eat NEG **what**
 Intended for: 'What does not John eat?' (Bošković 1998: 46)
- b. * *Il peut rencontrer qui ?*
 3SG can meet **who**
 Intended for: 'Who can he meet?' (Chang 1997: 19)
- c. * *Chaque étudiant a rencontré qui ?*
 each student have meet.PTCP **who**
 Intended for: 'Who did each student meet?' (Chang 1997: 19)
- d. * *Jean et Pierre croient que Marie a vu qui ?*
 John andPeter believe COMP Mary have see.PTCP **who**
 Intended for: 'Whom do John and Peter believe that Mary saw?' (Bošković 1998: 46)

Even though the LF movement in French is limited by the local interventions (also known as weak islands), the movement does not suffer from strong island effects (Shlonsky 2012), such as complex noun phrases as relative clauses, as in (2.107).

- (2.107) *Vous connaissez des gens qui pourraient héberger
 2PL know ART people REL can host
 combien de personnes ?
 how.many PREP person
 ‘You know people who could host how many people?’ (Obenauer 1994: 297)*

However, this immunity to strong island is only experienced with question words used as objects, but not subjects, as shown in the ungrammaticality in (2.108).

- (2.108)* *Il a construit une machine que qui va
 3SG have build.PTCP ART machine REL.OBJ who go
 utiliser ?
 use.INF
 Intended for: ‘He has built a machine that who is going to use?’
 (Obenauer 1994: 297)*

Nevertheless, the local constraints on *wh-in-situ* in French has been argued against theoretically (Mathieu 1999) and empirically (Mathieu 2004; Adli 2006). Both opposing standpoints do challenge the account for the covert movement of *wh-in-situ* in French. Furthermore, in an extreme interpretation of Minimalism, universal grammars do not allow covert movements, including both feature and phrasal movements (Kayne 1998: 183; Poletto & Pollock 2004: 251). Therefore, another kind of movement (i.e., remnant movement) is adopted to illustrate the *wh-in-situ* in French.

The analysis of **remnant movement** argues that *in-situ* question words do move overtly in French. This analysis is based on the observation of the *wh*-clitic doubling in North Italian dialects (cf. Poletto & Pollock 2004: 260–266). In these dialects, it is assumed to have two positions for *wh*-elements, namely, WhP₁ and WhP₂ (Poletto & Pollock 2009: 202).

- (2.109) [WhP₁ Wh^o₁ [FORCEP F^o [GROUNDP G^o [WhP₂ Wh^o₂ [IP ...]]]]]

The former is targeted by clitic question words (e.g., *s* in (2.110)), the latter is for ‘strong’ question words (e.g., *che* in (2.110)).

- (2.110) Illasi (Verona)
S’ a-lo fat che?
 what have-3SG do what
 ‘What has he done?’ (Poletto & Pollock 2009: 235)

According to their analysis, the ‘strong’ question words originated in the lower syntactic position first move to WhP₂ to check the feature. After the movement of ‘strong’ question

words, the remnant movement is triggered to place the rest of the clause, including the clitic question words, to a layer devoted to the expression of shared or presupposed information, namely the GroundP (Poletto & Pollock 2004: 284). Then, the clitic question words raises to WhP₁. Considering the scope of this dissertation, I do not go into any detailed motivation behind these movements, but only focus on the *wh*-doubling and the concept of remnant movement.

In the same vein, French *in-situ* is considered as *wh*-doubling. The only difference is that the position of WhP₁ is occupied by a null operator, but not clitic question words as in Italian dialects. The WhP₂ position is occupied by the question word originating from the Inflection phrase (IP), as shown in the syntactic structure in (2.111).

(2.111) [WHP₁ Ø [FORCEP F° [GROUNDP *tu vas*_j [TOPP [WHP₂ *où*_i [IP t_j t_i]]]]]]]
(Boucher 2010a: 71)

The question word *où* ‘where’ moves to the WhP₂, followed by the remnant movement of the rest of IP *tu vas* ‘you go’ to the GroundP. Both movements leave the traces (i.e., *t_j* and *t_i*) inside the IP. As a result, question words are *in-situ* on the surface, yet they do go through an overt movement. The phenomenon of *wh-in-situ* is in fact attributed to the result of remnant movement of the rest of the clause to a higher syntactic position.

In addition to movements, *wh-in-situ* in French can also be explained without resorting to movement analysis. Boucher (2010b) illustrates the **unselective binding theory** process by taking the *wh*-doubling structure and pragmatic constraints into account. As *in-situ* questions in French are highly constrained pragmatically and they are strongly grounded in contexts, Boucher (2010b: 130–131) argues that French *in-situ* questions use weak D-linked (i.e., discourse-linked; Pesetsky 1987) mechanism. By this mechanism, a covert existential operator is assumed to be inserted through contextual grounding. With this insertion, the null operator as a quantifier in the higher position can bind the question words in the lower position.

In addition to the three accounts mentioned above, the *in-situ* phenomenon in French is also treated from other approaches, such as the prosody licensing in Cheng & Rooryck (2000). They argue that rising intonation, as a polar questions morpheme in their term, is essential to license *in-situ* question words in French. The rising intonation is attested in constituent questions with *wh-in-situ*, but not with particle *est-ce que* and inversion (Cheng & Rooryck 2000: 5). On the other hand, Shlonsky (2017) argues that

wh-in-situ in French is derived from the feature movement rather than phrasal movement along with the account of criterial freezing (i.e., when a nominal moves into a criterial position, such as the positions of *wh*, focus, and topic, this nominal is not allowed to move further; Rizzi 2006: 112; Shlonsky 2017: 359). This latter mainly aims to account for the subject vs. non-subject asymmetry, occurring in the embedded *in-situ* question words.

Even though *in-situ* in French is theoretically constrained by the locality in the morpho-syntactic level, such as negative or modal verbs (Cheng & Rooryck 2000) and by strong island effects imposed on the embedded question words used as subjects, empirical data from corpora and experiments are provided to argue against the local constraints for *wh-in-situ* in French (Adli 2006). Moreover, the theoretical and empirical standpoints also disagree on the optionality of *in-situ* in French. As argued by Cheng & Rooryck (2000), question words in French do not optionally choose to either move or *in-situ*. It is determined by the occurrence of polar questions morpheme (i.e., rising intonation) in syntactic structure (see also Wu 2019). On the other hand, Adli (2006) lends empirical supports to not only overrule the ungrammatical sentences illustrated in theories, but also to argue that the optionality of question words correlates with registers, which is somehow in line with Boucher (2010b: 104), who suggests that *wh-in-situ* in French is pragmatically motivated.

2.4.5 Summary

The commonality of *wh-in-situ* in Mandarin and French have differences in certain aspects. First, **unlike Mandarin as a genuine *in-situ* language, French *in-situ* questions have no question particle to satisfy the question feature checking** (Cheng & Rooryck 2000).

Second, ***in-situ* is an option in French**. Namely, French is situated between English and Mandarin in terms of the positions of question words (Mathieu 1999: 441; Boucher 2010b: 104). English has the question words fronted while Mandarin has them *in-situ*. Given the optionality in French, the question word *qui* ‘who’ used as subject can be ambiguous when it comes to determining its syntactic mechanism, that is, whether such a question word is moved or *in-situ*. It has been argued that question words used as subject must move despite seemingly *in-situ* (cf. Shlonsky 2017 for more discussion). In contrast

to French, in Mandarin, the question word used as a subject 誰 *shéi* ‘who’ is determined *in-situ* as base-generated.

Third, other **detailed syntactic differences of *in-situ* questions between French and Mandarin can also be attributed to the covert LF movement.** From the overview in Sections 2.4.3 and 2.4.4, under the framework of movement, French *in-situ* is more constrained syntactically than Mandarin. Hence, *wh*-arguments and adjuncts are banned to be *in-situ* by local interventions (e.g., negation) in French, but not in Mandarin (Mathieu 2004: 1093, 1095). In addition, even *wh*-arguments are not allowed to be *in-situ* within strong island, such as relative clauses (Obenauer 1994) or embedded questions (Mathieu 1999: 443). However, Mandarin seems to escape from these locality effects, only certain *wh*-adjuncts, such as 怎麼 *zěnmě* ‘how’ with manner reading and 為什麼 *wèishénme* ‘why’ with causal reading, are constrained by strong islands. Table 17 summarises the syntactic constraints related to *in-situ* issues in the two languages.

Table 17 Syntactic interventions of *wh-in-situ* in Mandarin and French
(The sign “V” indicates with intervention effect, “X” indicates no effect)

		Mandarin	French
Weak islands (local interveners)	<i>Wh</i> -argument	X	V
	<i>Wh</i> -adjunct	X	V
Strong islands	<i>Wh</i> -argument	X	V (only with subject)
	<i>Wh</i> -adjunct	V (only with a limited number of adjuncts)	X

Such differences between two languages can be attributed to not only their syntactic properties, but also to the morphology of *wh*-elements and the nature of the variable (Mathieu 1999: 444).

However, it should also be noted that the syntactic constraints argued for *in-situ* questions in French are rejected empirically (cf. Adli 2006). Recently, Zimmermann & Kaiser (2019) have reviewed these syntactic constraints imposed on *wh-in-situ* in French and tended to corroborate mostly syntactic deductions with empirical findings in the corpus of Contemporary Hexagonal French Literary prose writing (cf. Zimmermann & Kaiser 2019: 134–137 for information about the corpus). It turned out that both

ungrammatical and dubious sentences disallowed in theory are in fact licit. Although they refuted the syntactic restrictions toward *wh-in-situ* in French, some of the restrictions remain unattested in the corpus (cf. Zimmermann & Kaiser 2019: 146–147). The absence of such instances is probably due to chance, indicating that it does not entirely support the infelicity of *in-situ* in certain syntactic environments as suggested theoretically. Therefore, the authors provided no firm conclusion with regard to those without empirical evidence and they encouraged future research. In addition, they examined and analysed *in-situ* questions based on morphological characteristics of verbs (e.g., tense, mood, and passive voice) and syntactic properties (e.g., verb classes and subject types). Their results showed that a definite subject might be the only requirement if question words remain *in-situ*.

The discussion on *in-situ* questions appeared to be divisive, which might be attributed to the scope of questions (Boucher 2010b: 124). In other words, some of the theoretically unacceptable *wh-in-situ* are legitimate when they are used as echo questions, or requests for clarification and confirmation. Studies have pointed out that French *in-situ* is not only constrained at the syntactic level, but also pragmatically (Obenauer 1994; Chang 1997: 45; Cheng & Rooryck 2000; Coveney 2011: 136–137). That is, *in-situ* questions in French are allowed to occur with the context of presupposition and they do not accept negative responses. However, it seems to me that pragmatic constraints are less covered from previous studies. Only empirical studies argue against the proposition that negative answers are not allowed (cf. Adli 2006 and Zimmermann & Kaiser 2019). The level of presupposition between *in-situ* questions and *wh-fronted* questions has not yet been discussed in empirical studies.

Throughout the overview of *wh-in-situ* in Mandarin and French, it can be seen that both languages share the asymmetry of adjuncts (referential vs. non-referential) (Tsai 1994a: 131–135). However, when it comes to the syntactic mechanisms of accounting for *wh-in-situ*, especially movement, there are more differences than similarities. While islands have effects on both languages, Table 17 illustrates that the **two languages behave differently when dealing with syntactic islands**. It seems easy to account for the distribution of *in-situ* question words, but it is no easy task to fully capture the *in-situ* syntactic mechanism within the same framework. This idea has also been brought up in a recent contrastive study. Wu (2019) argues for a unified syntactic account for *wh-in-*

situ in the two languages, namely a deletion of the copy in a higher syntactic position (i.e., specifier of CP). Despite the same syntactic derivation of *wh*-in-situ, Mandarin and French have different motivations due to their distinguished requirement in the level of Phonetic Form (cf. Wu 2019: 121–122). I will not get into any detail of the procedure, yet I would point out the fact that differences might occur even when endeavouring to account for the same linguistic phenomenon under a unified syntactic account, be it movement or deletion.

Finally, although *wh*-in-situ is recognised in the nature of Mandarin questions, some contexts do allow for overt *wh*-movement (i.e., *wh*-fronting) in Mandarin (Cheung 2014; Liu 2019; Ren 2023). Moreover, the asymmetry between adjuncts and arguments is documented (Tsai 1994a: 162–163). Given that *wh*-fronting is not directly relevant to this dissertation, I will not discuss this issue further, whether monolingually or cross-lingually. In the following section, I will discuss the topic of question particles, which is noted to be correlated with the position of question words, such as *wh*-in-situ discussed in this section (Baker 1970: 208). The following section will cover the overview of question particle in individual languages. A cross-linguistic account of question particles will be provided as well.

2.5 Question particles

In addition to *wh*-in-situ, Mandarin and French share another similarity in their questions: they use particles. Therefore, the core of this section concentrates on the common ground between the two languages.

Typologically, the particle is considered one of the marking strategies at the morphological level, which also comprises tags, verbal morphology, and question words; see Section 2.1.3 for more details. The topic related to particles has been studied from a typological perspective (Dryer 2013a; Bailey 2010; Bailey 2013; Del Gobbo, Munaro & Poletto 2015). However, from a monolingual perspective, the study of particles seems more developed in Asian languages, while it has not received much attention in European languages (Hancil, Post & Haselow 2015: 5). Taking the example of the two languages studied in this dissertation, scholars in Mandarin have contributed abundantly on particles, such as Li & Thompson 1981, Cheng 1977, Cheng, Huang & Tang 1996, Li 2006, Chu 2009, Simpson 2014, Pan & Paul 2016, Luo & Wu 2017, Fang & Hengeveld 2022, Pan

2021, to just name a few. Compared with Mandarin, French has considerably fewer in-depth studies on particles, such as Munaro & Pollock 2008 and Smirnova & Abeillé 2021. Given the lack of in-depth research on French particles, there is, to the best of my knowledge, hardly any contrastive study of Mandarin and French on this topic. Accordingly, this section first aims to give an overview of the particles in Mandarin and French and then provide a contrastive view.

Although the only focus here is the particles used in the questions, the term *particle* can also refer to other elements, such as interjections, prepositions, and articles. In such circumstances, the particle is a sort of word that is difficult to theoretically categorise into a word class (Hancil, Post & Haselow 2015: 6) insomuch that Greenberg does not include them in his typological generalization principle for word order (Paul 2014: 77–78). Nevertheless, scholars propose a definition of particle as follows: “an invariant element with a grammatical function that does not belong to one of the major grammatical categories” (Bailey 2010: 25). A similar definition is also proposed by Simpson (2014: 156) and Hancil, Post & Haselow (2015: 9). They point out the characteristic of particles as being formal, non-inflecting and monomorphemic (cf. Mulder & Thompson 2008: 183 and Thompson & Suzuki 2012: 668 for other tentative definitions). Hence, this relatively loose definition may contribute to the term encompassing various elements, from prepositions to interjections. Although there is not much similarity between them in terms of grammatical function, each of them usually has a specific grammatical and pragmatic meaning (Del Gobbo, Munaro & Poletto 2015: 359), such as *up* in the verb phrase *give up* (Richards, Platt & Platt 1992: 265).

The particles in the following discussion in Mandarin and French fit well with this terminology. They are **invariant** on the one hand; on the other hand, they not only have **a particular grammatical meaning**, such as modifying the modality of a sentence, but they may **encode other pragmatic functions**, such as showing the relevance to the precedent context.

Therefore, given my primary focus related to questions, I will use *question particle* as a cover term, as defined and applied in Luo & Wu (2017: 81), but without committing to the fact that they can only be used in questions or that they assume the total contribution of questioning, particularly the 呢 *ne* particle in Mandarin (see Section 2.5.1.4). Also, the literature on question particles in Asian languages or in generative framework often uses

the term *sentence-final particle* (Erlewine 2017; Fang & Hengeveld 2022; Paul & Yan 2022) or *utterance-final particle* (Chu 2009; Tseng & Jeong 2020) interchangeably (Hancil, Post & Haselow 2015: 4), given the fixed final position of question particles. In this dissertation, I adhere to a neutral term ***question particle*** in that the particles in French may not be placed finally (cf. Dryer 1992: 102, indicating the particle’s position may vary from one language to another based on the position of the verb).

This section will be organised as follows: I will begin with the frequently used particles in Mandarin and pay special attention to those occurring in questions, such as 嗎 *ma*, 吧 *ba*, 啊 *a*, and 呢 *ne*. I will mainly discuss their syntactic distribution and pragmatic functions. Following the discussion in Mandarin, my attention shifts to the three particles in French, namely, *est-ce que*, *ça*, and *donc*. The discussion will similarly focus on the syntactic and pragmatic levels. Finally, I will provide a summary of the main insights discussed in the section, paying particular attention to the cross-linguistic contrasts that emerge from this literature review.

2.5.1 Particles in Mandarin

In Mandarin, particles are monosyllabic items. They cannot be used independently. However, they add extra structural and functional information to the sentences to which they are attached (Yip & Rimmington 2021: 309). Some information related to particles may be used to differentiate the modality of sentences (e.g., adding the 嗎 *ma* particle transforms a declarative sentence into a question). These particles all occur at the final position of a sentence to the extent that they are often labelled *sentence-final* or *utterance-final* particles in the literature (see above).

Among the various particles mentioned in the Mandarin literature (Chao 1968: 797–814; Li & Thompson 1981: 238; Li 2006: 6; Chu 2009: 282; Pan 2021; Yip & Rimmington 2021: 151; Fang & Hengeveld 2022: 879), six of them are frequently used, namely, 的 *de*, 了 *le*, 嗎 *ma*, 吧 *ba*, 啊 *a*, and 呢 *ne* (Li 2007: 144; Fang & Hengeveld 2022: 879), as shown in (2.112) - (2.117).

- (2.112) 他 會 生氣 的。
 tā *huì* *shēngqì* *de*.
 3SG will angry PTCL
 ‘He will get angry (that is what’s going to happen).’ (Chao 1968: 797)

- (2.113) 晚上 的 餃子 有 點 鹹 了。
wǎnshàng de jiaozi yǒu diǎn xián le.
 night DE dumpling have a.little salty PTCL
 ‘The dumplings I had last night were a little salty!’
 (Fang & Hengeveld 2022: 882)
- (2.114) 你 懂 我 的 意思 嗎？
nǐ dǒng wǒ de yìsi ma?
 2SG understand 1SG DE meaning PTCL
 ‘Do you understand what I mean?’ (Fang & Hengeveld 2022: 884)
- (2.115) a. 你 明白 我 的 意思 吧？
nǐ míngbái wǒ de yìsi ba?
 2SG understand 1SG DE meaning PTCL
 ‘You understand what I meant, right?’ (Fang & Hengeveld 2022: 891)
- b. 我們 走 吧！
wǒmen zǒu ba!
 1PL go PTCL
 ‘Let’s go!’ (Chu 2009: 285)
- (2.116) 哪 本 書 啊？
nǎ běn shū a?
 which CL book PTCL
 ‘Which book is it then?’ (Fang & Hengeveld 2022: 895)
- (2.117) a. 胡飛 買 了 哪 一 本 書 呢？
Húfēi mǎi le nǎ yī běn shū ne?
 Hufei buy ASP which one CL book PTCL
 ‘Which book did Hufei buy?’ (Cheng 1991: 25)
- b. 他們 有 三 條 牛 呢。
tāmen yǒu sān tiáo niú ne.
 3PL have three CL cattle PTCL
 ‘They have three cattle.’ (Li & Thompson 1981: 301)

The first two particles, 的 *de* and 了 *le*, can also be used at the morphological level. For example, 的 *de* particle can be used as a particle in a possessive structure, which links the possessor with its possessee, as in (2.118). The 了 *le* particle is used to modify an action, known as 了₁. It is a perfective aspect marker and indicates the completion of such an action (Fang 2018: 590), as in (2.119).

- (2.118) 他 的 書
tā de shū
 3SG DE book
 ‘his book’ (Chao 1968: 795)

(2.119) 我 回來 了。(了₁)

wǒ huílái le.
1SG back ASP
'I have come back'

(Chao 1968: 799)

By contrast, when the particles 的 *de* and 了 *le* are used at the sentential level, they are added to imply and signal the speaker's attitude. For instance, the 的 *de* particle in (2.112) reinforces the certainty of the speaker on the fact that the other person will surely get angry. Without 的 *de*, the sentence would become neutral, meaning that the speaker enunciates a statement (Chao 1968: 797; Fang & Hengeveld 2022: 881).

The 了 *le* used at the sentential level has two functions, labelled as 了₂ and 了₃. The former 了₂ is a perfect marker (also called mood auxiliary) indicating a new situation or change of state (Li & Thompson 1981: 296–300; Fang 2018: 590, 598), as in (2.120).

(2.120) 火車 馬上 就 開 了。(了₂)

huǒchē mǎshàng jiù kāi le.
train immediately soon leave PTCL
'The train is leaving immediately'

(Li & Thompson 1981: 296)

The latter 了₃ used in (2.113) is a mirative marker (Fang 2018; Fang & Hengeveld 2022: 882–883). Adding the 了₃ particle indicates that the utterance stated is newsworthy and surprising. In addition to the worthiness highlighted, reminding the addressee is also entailed. For instance, in example (2.113), the speaker reminds the addressee of not putting too much salt next time. Although the particle 了 *le* encodes several functions, when 了₂ and 了₃ occur in the final position, as in (2.121), they can merge into one single 了 *le*.

(2.121) 我 餓 了。(了₂₊₃)

wǒ è le.
1SG hungry PTCL

'I have become hungry. (Anything I can eat?)'

(Fang 2018: 598)

This example above shows that a state change began in the past and continues to the moment of utterance. Also, the speaker tries to make the addressee notice the hunger. Given the focus of the dissertation, I will not go into detail on how to disentangle the three 了 *le* particles (cf. Fang 2018 for more information).

The other four particles 嗎 *ma*, 吧 *ba*, 啊 *a*, and 呢 *ne*, are concerned with the sentential level, and their primary function is to modify sentences (see Del Gobbo, Munaro & Poletto 2015: 364–365 for common properties of sentential particles). Consequently, the four particles can be found in different sentence types, such as declarative, interrogative, and exclamative (Li & Thompson 1981: 300–305, 307, 314–317; Li 2006: 29; Fang & Hengeveld 2022). Most importantly, the four particles are all allowed in questions. In particular, the 嗎 *ma* and 呢 *ne* particles are often touched upon when discussing questions, as they distinguish two question types (i.e., confirmation-seeking and information-seeking questions) (cf. Sections 2.3.1 and 2.3.2). While each of these four particles has a different degree of relevance to questions, they are all associated with this type of sentence. Thus, the following subsections will detail their usage in different sentence types in the order of 嗎 *ma*, 吧 *ba*, 啊 *a*, and 呢 *ne*. Note that the following discussion will be accompanied by a number of examples extracted from the literature. However, some studies do not clearly indicate the source of the examples, such as Han (1995) and Zhu (1999). By contrast, other studies are based on the authentic language uses, such as Chao (1968: vii, xvii), Li & Thompson (1981), Kubler & Ho (1984), Chappell (1991) use spoken data, while Chu (1998) and Li (2007) use written data. Moreover, using a large-scale corpus is often seen in the study after 2000, as Li (2006) using the CCRL and Fang & Hengeveld (2020; 2022) using CCL modern corpus.¹⁰

11

2.5.1.1 Ma particle

The ***ma* particle** is considered **a particle for polar questions**, as in (2.114). This particle takes effect when it comes to transforming declarative sentences into polar questions (Han 1995: 114; Li 2007: 119; Pan 2021: 3). Accordingly, the polar question in Mandarin is also termed (*question*) *particle question* or *ma question* in the literature (Li & Thompson 1981: 547; Cheng 1984: 89).

¹⁰ CCRL stands for Chinese Corpus Retriever for Language Teaching and Research created by Department of Computer Science and Technology, Beijing Language and Culture University (Li 2006: 4).

¹¹ The CCL modern corpus is the Peking Corpus compiled by the Center for Chinese Linguistics (CCL) (Zhan et al. 2019). In the study of Fang & Hengeveld (2020; 2022), they use the subcorpus of modern Mandarin, which covers both spoken and written Mandarin in various text types, including novels, plays, stories, TV programs, movies, newspapers, cyber-text, and translated works.

Although the usual character to use is 嗎 *ma* (Chao 1968: 801), other graphic variants also exist, such as 嘛 in (2.122) and 麼 in (2.123), examples retrieved from the *Academia Sinica Balanced Corpus of Modern Chinese 4.0* (CKIP group 1998).

(2.122) 難道 就 這樣 耗 下去 嘛?
nándào jiù zhèyàng hào xiàqù ma?
 don't.tell.me just this waste down PTCL
 'Don't tell me that we are just going to continue like this?'

(2.123) 你 也 認得 她 麼?
nǐ yě rèndé tā ma?
 2SG also recognise 3SG PTCL
 'Do you recognise her too?'

However, the distribution of different graphics is correlated with sentence types. Namely, 嗎 *ma* is used in questions, whereas 嘛 *ma* is found in other sentence types, such as declarative, exclamative, and imperative (Li 2006: 32). This complementary distribution of *ma* in sentences also reflects pronunciation (Kubler & Ho 1984: 76; Li 2006: 67), although Li (2006: 47) does not provide examples with characters.

(2.124)a. *Hóngjiàn qù xuéxiào le ma? ↑*
 Hongjian go school PRF PTCL
 'Did Hongjian go to school?'
 b. *Hóngjiàn qù xuéxiào le ma. ↓*
 Hongjian go school PRF PTCL
 '(Obviously/certainly) Hongjian went to school.' (Li 2006: 47)

Comparing the two sentences in (2.124), the pitches are different on the last syllable (i.e., the constituent *ma*), indicated with an upper and falling arrow. In (2.124a), the sentence ends with a particle associated with a rising pitch. This intonation-particle combination shows that the speaker has no idea whether Hongjian attended school. By contrast, the particle in (2.124b) is associated with a falling pitch, meaning that the speaker is relatively confident that Hongjian went to school, given that this event is contextually self-evident.

Regarding the distinction of characters, although both characters may be used interchangeably, as attested in Modern Chinese (examples (2.122) and (2.123)), even diachronically, the link between characters and sentence types has gradually stabilised (Pan 2021: 3–5), at least concerning their pronunciation. Without going into further detail, I fully acknowledge and agree that the character 嗎 *ma* is for questions, while the character 嘛 *ma* is for other sentence types (especially for declarative sentences), despite

the fact that there is as yet no consensus on the convention of character use, as suggested by Chao (1968: 801).

A distinction is made between two *ma* particles according to sentence types (Kubler & Ho 1984: 76; Chappell 1991: 43–44; Chu 1998): *ma*₁ is a question particle while *ma*₂ is a particle used in other sentence types (also known as a modal marker) (Chappell 1991: 39). The *ma*₂ particle has a presupposition function at the semantic level and an insistence function at the pragmatic level (Chu 1998, as cited in Li 2006: 27), as attested in (2.124b): speakers with impatience and disagreement insist on having their hearers accept a presupposed fact. Given that the *ma*₂ particle has no direct relevance to the discussion of the question particle, my focus will be mainly on the *ma*₁.

As opposed to the two-way distinction of particle *ma*₁ and *ma*₂, Li (2006: 28) suggests a unified *ma* particle. Even though *ma* as a particle used in polar questions seems undebatable, she argues that this particle is merely a reinforcement to the speech act in any sentence type. In declarative sentences, the *ma* particle shows a high degree of the speaker's commitment to the assertion. In other sentence types, such as questions and imperatives, this particle represents a high degree of the speaker's intention to have an action carried out (Li 2006: 31–36). Because the *ma* particle functions in a similar way across sentence types, she argues against the label of question particle *ma* (Li 2006: 29). Similarly, Fang & Hengeveld (2022: 884) contend that the *ma* particle denotes an interrogative force in that polar questions in Mandarin can also be marked by intonation.

In my opinion, one should distinguish between phonological and morphological strategies for question marking, as does the literature in Mandarin (Li 2007: 117–118; Lin 2015: 238) and the typology of questions (cf. Section 2.1.3). Accordingly, the *ma* particle is still legitimate for bearing an interrogative force in polar questions (Simpson 2014: 160), namely a question marker (Paul 2014), while Li (2006: 35) only considers it as a degree marker. In addition, the two marking strategies should be differentiated regarding the cline of interrogativity (Halliday & McDonald 2004: 336). Polar questions ending with *ma* particle are found at the end of the strongest interrogativity, whereas polar questions marked by intonation are the weakest (Li 2007: 121). Likewise, such a difference also reflects on the face-threatening effect both marking strategies are subject to. Due to the weakest interrogativity, a polar question marked by intonation is a declarative form by appearance. Therefore, it is less likely to be treated as a 'canonical'

question. In other words, using polar question with declarative word order can be seen to minimise the face-threatening risk caused by asking a question (Levinson 2012: 23).

Despite such distinction on the level of interrogativity and face-threatening effect, phonetic and morphosyntactic strategies for polar questions in Mandarin are commonly biased compared to the A-not-A structure, which is considered an unbiased question form. Semantically, a polar question comprises either a positive or a negative statement required to be confirmed or denied by the interlocutor. In other words, polar questions are already positively or negatively biased when the enquirer enunciates them (Li 2007: 120).

Given that polar questions are easily found biased when seeking an answer, such a characteristic makes them eligible for rhetorical questions (Li & Thompson 1981: 552; Alleton 1988: 291; Caponigro & Sprouse 2019: 131). In Mandarin, the rhetorical modality is mainly attested in negative polar questions (Chao 1968: 800; Yip & Rimmington 2021: 170–171), as in (2.125). The rhetorical effect can become even more potent by adding an adverb 難道 *nándào* (Alleton 1988: 278; Xu 2012; Yip & Rimmington 2021: 171), as in (2.126).

(2.125) 你 不 怕 老虎 嗎?
nǐ bú pà lǎohǔ ma?
 2SG NEG afraid tiger PTCL
 ‘Aren’t you afraid of tigers?’ (Chao 1968: 800)

(2.126) 你 難道 不 想 家 嗎?
nǐ nándào bù xiǎng jiā ma?
 2SG don’t.tell.me NEG miss home PTCL
 ‘Don’t you really miss your family?’ (Yip & Rimmington 2021: 171)

Overall, the distinction between *ma*₁ and *ma*₂ seems less relevant, since the primary focus herein is the *ma* particle used in the questions. However, I think the distinction of the two *ma* particles is still necessary, phonetically (Li 2006: 29, 47, 67), morpho-syntactically (Kubler & Ho 1984: 76), and pragmatically (Chappell 1991; Chu 1998).

One is used in questions, while the other is used in different sentence types. As for their characters, although the characters of two *ma* particles may not seem to be strictly assigned to one sentence type, as shown in (2.122) and (2.123), to some extent there is a tendency to distinguish the usage of two *ma* particles in writing, as attested in Taiwan Mandarin in Kubler & Ho (1984: 76).

Regarding pronunciation, it is true that such a difference should be attributed to the whole sentence contour (Fang & Hengeveld 2022: 883–884) since particles are usually neutral and de-stressed. However, different sentence contours also influence the intonation of the *ma* particle. Accordingly, the phonetic change under the influence of the sentence contour indirectly distinguishes two *ma* particles.

By summing up the views from Kubler & Ho (1984: 76), Chappell (1991), and Li (2006: 47), pragmatically, the *ma* particle attached to questions with high intonation shows the intention of seeking confirmation from the addressee, while the other occurring in other sentence types with low intonation is to express the impatience or disagreement from the speaker and to highlight the obviousness of the presupposed fact that the interlocutor has neglected.

Finally, the biased nature of polar questions with a *ma* particle or declarative form makes polar questions one of the possible candidates of rhetorical questions. However, they differ on the level of questioning to reduce the threat when facing the addressee.

In addition to the *ma* particle, the 吧 *ba* and 啊 *a* particles are attested in polar questions (Zhu 1999: 237–238; Li 2006: 29; Luo 2013: 39), as in (2.127).

- (2.127) 你 是 他 哥哥 啊/吧?
 nǐ *shì* *tā* *gēge* *a/ba?*
 2SG COP 3SG brother PTCL
 ‘Are you his elder brother?’ (Luo 2013: 6)

They may occur in the same question type, but their occurrence in polar questions could have a varied degree of certain functions and even leading to different effects in pragmatics.

2.5.1.2 Ba particle

Compared to the *ma* particle, the ***ba* particle** seems less problematic in terms of character. However, Chao (1968: 807) distinguishes two *ba* particles derived from different etymology origins. The *ba*1 particle is a fusion of the word 不 *bù* ‘not’ and the 啊 *a* particle, while the *ba*2 particle is derived from the verb 罷 ‘stop, finish’ (Tang 2015: 2). Following Chao (1968), Zhu (1999: 237) agrees that only the *ba*1 particle is related to questioning but that the *ba*2 particle denotes an imperative function. Apart from these two

studies, other scholars seem to agree on a unified *ba* particle. This view is also adopted in the following discussion of the *ba* particle.

In terms of distribution, the *ba* particle occurs in declarative sentences, questions, imperative sentences (Han 1995: 101–102; Li 2006: 30), exclamative sentences, and prohibitive sentences (Fang & Hengeveld 2020: 290), as shown in the following examples in (2.128).

- (2.128) a. 他 在 辦公室 吧。(Declarative)
tā zài bàngōngshì ba.
 3SG PREP office PTCL
 ‘(Probably) He is in his office.’ (Li 2006: 30)
- b. 他 在 辦公室 吧?(Question)
tā zài bàngōngshì ba?
 3SG PREP office PTCL
 ‘He is in his office, right?’ (Li 2006: 30)
- c. 進來 吧!(Imperative)
jìnlái ba!
 come.in PTCL
 ‘(I suggest you) come in!’ (Li 2006: 30)
- d. 也 太 明顯 了 吧!(Exclamative)
yě tài míngxiǎn le ba!
 as.well too obvious PTCL PTCL
 ‘It is so obviously unreasonable as well.’ (Fang & Hengeveld 2020: 290)
- e. 別 掉 了 吧.(Imperative/ Prohibitive)
bié diào le ba.
 do.not sling PTCL PTCL
 ‘Don’t sling (the rail) anymore.’ (Fang & Hengeveld 2020: 290)

It seems that the *ba* particle can occur in any sentence type, but opinions differ regarding questions. **The *ba* particle is generally added in polar questions** (Chao 1968: 807) but not allowed in *wh*-questions and disjunctive questions with A-not-A structures (Li & Thompson 1981: 310). Nevertheless, certain studies provide evidence against such a use, which is limited to polar questions (Chao 1968: 807; Zhu 1999: 237–238; Han 1995: 101–102; Li 2006: 22, 30–31; Fang & Hengeveld 2022: 301), as in the following examples in (2.129).

- (2.129) a. 你 到底 要 幹 什麼 吧 ?
 nǐ dàodǐ yào gàn shénme ba?
 2SG anyway want do what PTCL
 ‘What do you want to do, anyway?’ (Chao 1968: 807)
- b. 一共 多少 錢 吧 ?
 yíngòng duōshǎo qián ba?
 altogether how.much money PTCL
 ‘How much altogether? Please tell me.’
 (Zhu 1999: 237; Fang & Hengeveld 2022: 891)
- c. 到底 誰 去 吧 ?
 dàodǐ shéi qù ba?
 anyway who go PTCL
 ‘Who's going anyway?’ (Han 1995: 101)
- d. 你 說 不 說 吧 ?
 nǐ shuō bù shuō ba?
 2SG speak NEG speak PTCL
 ‘Are you going to tell me or not?’ (Han 1995: 102)

Although the occurrence of *ba* particle in both questions is documented, it is rarely used as such (Han 1995:112). Instead, the *ba* particle occurs mainly in polar questions (Zhao & Sun 2015: 125). Furthermore, very few co-occurrences of *wh*-word 什麼 ‘what’ and the *ba* particle are found in the CCL modern corpus (Fang & Hengeveld 2020: 301). The examples in (2.129) are no longer questions but imperative sentences (Chao 1968; Zhu 1999; Han 1995; Li 2006: 33). Adding the *ba* particle seems to invite the addressee to answer the question (Fang & Hengeveld 2020: 301), it is an imperative (a directive). Example (2.129a) shows a directive meaning of ‘tell me, what do you...’ (Chao 1968: 807), likewise in (2.129b). In addition, adding this particle amplifies the underlying emotion of the speaker. The particle *ba* in (2.129c) and (2.129d) shows the impatience and anger of the speaker to such an extent as to force the hearer to respond if the latter wants to avoid trouble (Han 1995: 111). Therefore, although some studies accept the use of the *ba* particle in *wh*- and disjunctive questions, this may be due to differences in the usage of Mandarin in different regions or differences based on the intuition of the researchers. I think the comparable corpus of Taiwan Mandarin used in the present study can provide another perspective regarding this usage.

Although the occurrence of the *ba* particle in declarative, questions, and imperative seems unequivocal, its functions are highly debatable in the literature, including modal meaning of uncertainty, estimation, soliciting agreement, suggestion, undetermined

intention (Fang & Hengeveld 2020: 285, 2022: 892), hesitation, warning, friendly sarcasm, and politeness (Chu 2009: 286–288). Moreover, there is no consensus on whether this particle functions similarly across different types of sentences. In a declarative sentence, as in (2.128a), the *ba* particle reduces the speaker's commitment to the statement to such an extent that the speaker welcomes any correction or rejection (Han 1995: 104–105). That is, adding the *ba* particle shows the speaker's uncertainty and creates an easy way for the addressee to disagree with that statement (Fang & Hengeveld 2020:284). Given the uncertainty encoded in the *ba* particle, the latter introduces a function of request or advice in imperative sentences, as in (2.128c) (Li 2006: 23, 33).

When *ba* particle occurs in polar question as in (2.128b), the speaker is soliciting agreement from the addressee. It seems that the uncertainty of the speaker is the core property of the *ba* particle, and other functions are derived from it (Chu 2009: 288). However, I think the functions of this particle can be discussed from different angles in a discourse. From the speaker's point of view, the particle adds the speaker's uncertainty to the sentence. Standing from the perspective of the interaction, the solicitation of agreement may be regarded as a hint of turn-taking. Therefore, two primary functions of the *ba* particle are often mentioned: (1) to seek agreement from the hearer and (2) to show the uncertainty of speakers (Kendrick 2018: 18–19).

However, few studies argue for a unified function of the *ba* particle, such as downgrading the speaker's epistemic position. When using the *ba* particle, the speaker demonstrates a lower level of knowledge regarding the statement than the addressee (Kendrick 2018: 10). This function seems to show the uncertainty from the speaker. As for Fang & Hengeveld (2020), they refute considering the *ba* particle as a mere epistemic marker of uncertainty, given that its occurrence is not limited only to declaratives and questions but also other sentence types (Fang & Hengeveld 2020: 289–290). They argue that this particle is a pragmatic marker, namely a mitigator, meaning to reduce the possible unwelcome perlocutionary effects on the addressee (Fang & Hengeveld 2020: 285). Also, note that the mitigating impact becomes specified in different situations, such as attaining more politeness and leaving more space for negotiation (Fang & Hengeveld 2020: 303–304).

The functions of the *ba* particle seem to vary from one study to another. However, after this overview, I wish to argue that its core function is to show that the speaker is still

determining the statement. Therefore, this uncertainty further paints the sentences with different discursive functions, such as suggestions, solicitation of agreement, and politeness. In addition, regardless of the sentence types or illocutionary forces, given the uncertainty indicated by the *ba* particle, this particle makes the sentence appear as questioning to some extent. In a declarative sentence, the *ba* particle mitigates the assertiveness of the statement, making it easier to be disagreed by the addressee (Fang & Hengeveld 2020: 301–302). In an imperative sentence, the *ba* particle softens the tone and has the directive speech act open for more space for negotiation, which becomes more suggestive (Fang & Hengeveld 2020: 300). Accordingly, the speaker seeks a response about the suggestion from the addressee and implicitly hopes the order to be carried out. Hence, both declarative and imperative sentences with *ba* particle are facing a similar risk as questioning, which is, being confirmed or rejected.

In polar questions, the *ba* particle not only encodes the uncertainty of the speaker, but also mitigates the interrogative force and makes the question less inquisitive (Fang & Hengeveld 2020: 300). On the scale of interrogativity (Halliday & McDonald 2004: 336; Li 2007: 121), polar questions with the *ba* particle have slightly stronger interrogativity compared polar questions using intonation, but it appears to be weaker compared to polar questions with the *ma* particle. Accordingly, the *ba* particle is situated in the middle of the scale. Although the degree of uncertainty becomes higher when the *ba* particle is used, its mitigator function helps to tone down the interrogative force. Finally, it has been noted that both *ma* and *ba* particles occur in polar questions to seek confirmation (Simpson 2014: 160). However, they are mutually exclusive (Han 1995: 113–114). Not only do they differ in the cline of interrogativity, but they also vary discursively in the degree of predictability (Li 2006: 35). That is, when asking a *ba* question, the inquirer is highly certain about the answer from the addressee (Li 2006: 35), while a *ma* question shows that the speaker might be completely unaware of the answer. Given the higher predictability and less inquisitive force of the *ba* particle, this type of question is often correlated with tag questions, the answer of which is maximally predictable (Li & Thompson 1981: 309–310; Chu 2017: 286; Yip & Rimmington 2021: 171; Pan 2021: 2).

Similar to *ma* and *ba* particles, the *a* particle is also used in polar questions. However, its functions remain controversial, given its occurrence in a broader range of sentence types (Fang & Hengeveld 2022: 894–895).

2.5.1.3 A particle

Among the Mandarin particles discussed in this section, only this particle contains a single vowel. This particularity makes it much easier to fuse with the vowel that ends in the preceding syllable (Chao 1968: 803; Zhu 1999: 238). Such a fusion creates variants of the ***a* particle** mentioned in the literature, such as 呀 *ya* (= the word ending with the vowel-*a/-e/-i/-o/-ü* + 啊 *a*), 哇 *wa* (= the word ending with the vowel *-u* + 啊 *a*), 哪 *na* (= the word ending with the vowel *-n* + 啊 *a*). In addition, a fusion with the whole preceding phoneme is also attested, such as 啦 *la* (= 了 *le* + 啊 *a*) (Chao 1968: 796, 803; Zhu 1999: 238, 240; Lu 1999: 46; Li 2007: 144; Fang & Hengeveld 2022: 894). Given that these variants are due to the fusion with the *a* particle, they should share the same characteristics and function with a single *a* particle.

Phonetically, unlike the *ma* particle, having two tones leads to the difference in sentence type distribution and even two different characters (cf. Section 2.5.1.1), the *a* particle can be pronounced with either high or low pitch, but with the same character and no distinction among sentence types (Wu 2004: 128; Li 2006: 41, 43–50; Chu 2009: 283). However, Wu (2004: 178–179) suggests that notably low pitch occurs mostly in interrogatives, while flat or rising pitch occurs in non-interrogatives situations.

The *a* particle has the widest distribution in sentence types in comparison to other particles (Zhu 1999: 238; Li 2006: 37, 52–53; Fang & Hengeveld 2022: 894–895), as shown in the following examples in (2.130).

(2.130)

- a. 我 並 沒 有 做 錯 啊。(Declarative)
wǒ bìng méi yǒu zuò cuò a.
 1SG on-the-contrary NEG have do wrong PTCL
 ‘I didn’t do it wrong.’ (Li 2006: 52)
- b. 你 不 去 啊? (Polar question)
nǐ bú qù a?
 2SG NEG go PTCL
 ‘You are not going?’ (Li 2006: 52)
- c. 哪 本 書 啊? (*Wh*-question)
nǎ běn shū a?
 which CL book PTCL
 ‘Which book is it then?’ (Fang & Hengeveld 2022: 895)

- d. 你 去 不 去 啊? (Disjunctive question)
 nǐ qù bú qù a?
 2SG go NEG go PTCL
 ‘Do you want to go or not?’ (Zhu 1999: 238, my translation)
- e. 大家 記住 啊! (Imperative)
 dàjiā jìzhù a!
 everyone learn-by-heart PTCL
 ‘Everyone, please learn it by heart.’ (Fang & Hengeveld 2022: 895)
- f. 你 這 小子 好 聰明 啊! (Exclamative)
 nǐ zhè xiǎozi hǎo cōngmíng a!
 2SG DEM little.kid very clever PTCL
 ‘How clever you are as a such little kid!’ (Li 2007: 149, my translation)

These examples above show that the *a* particle is generous in combining with different sentence types. However, Wu (2004: 129) notes that the *a* particle primarily helps construct confirmation questions. Such a claim is opposed to most studies arguing that the *a* particle is not an indicator for questions (Li & Thompson 1981: 314; Li 2006: 38; Luo 2013: 6). Moreover, Li (2007: 149) argues that the *a* particle is least likely to occur in questions in comparison with other particles. Also, the author finds that the *a* particle is the most strongly associated with exclamative sentences. Accordingly, this particle is labelled as an assessment particle (Li 2007: 127). In my opinion, the point made by Li (2007) should be that the function of the *a* particle is frequently related to exclamative use, rather than that the *a* particle occurs more regularly in such a sentence type. Given a wide range of distribution across sentence types, most discussions center on the functions of this particle (Li 2006: 37).

Ten functions are listed in Chao (1968: 804–806): soften questions, vocative use, commands, impatience, warning, pause for hearer, enumeration, etc. However, some of these functions might contradict each other. For example, this particle softens questions, but on the other hand, it encodes commands and warnings (Wu 2004: 127–128). In addition, other studies also argue that some functions, such as vocative, command, impatience, and warning, are not directly related to the particle itself but to other constituents in the sentence (Li & Thompson 1981: 314–316; Fang & Hengeveld 2022: 895–896). Nevertheless, these studies agree that the *a* particle reduces the forcefulness of the statement and softens the tone, which makes sentences less blunt. In other words, the

a particle plays the role of mitigator to pragmatically reduce the potential face-threatening effect (Fang & Hengeveld 2022: 897).

However, Kubler & Ho (1984: 39) and Chu (2002: 4, as cited in Wu 2004:128) suggest that the *a* particle is used to intensify the force contained in the utterance, such as statement, exclamation, and interrogation. Moreover, Chu (2009: 284–285) suggests that all other functions of the *a* particle are precisely derived from the core, namely expressing the involvement of the speaker, namely, the speaker's personal concern (Chu 2002: 4, as cited in Wu 2004:128). This function conveys the speaker's opinion and implies a closing conversation (Li 2006: 47–48). This speaker-oriented function is also reflected in the pitch of the *a* particle, which is often lower. This contrasts with the listener-oriented function of *a* particle, which invites the addressee to respond and continue discussing. The pitch of the particle in such a situation is high (Chu 2009: 284, 2017: 48).

Within the interaction in discourse, using the *a* particle increases the degree of contextual relevance, especially from the speaker's perspective (Li 2006: 42; Chu 2009: 283). Speakers using the *a* particle try to draw attention from the hearer and to make themselves more actively engage in the conversation (Li 2006: 50–51). In my opinion, such a function is exactly the core function of *a* particle, which is to show the speaker's personal concern. Not only does the speaker try to engage in a conversation, but adding this particle also reduces the face-threatening effect when calling attention from the hearer.

To conclude, the *a* particle may have two main functions derived from its core function. It reduces the forcefulness of a statement from the speaker's position and increases the degree of contextual relevance when speakers try to get involved in a conversation. Likewise, the contextual function seems essential to the *ne* particle, which will be discussed in the next section.

2.5.1.4 *Ne* particle

As discussed in Section 2.3.2, the ***ne* particle** can only be used in question types other than polar questions. That is, **only *wh*-questions and disjunctive questions** (with a 還是 *háishì* 'or' connective and an A-not-A construction) allow the occurrence of the *ne* particle, as shown in examples (2.131).

As two *ne* particles are distinguished based on sentence types (i.e., 呢 in questions and exclamation vs. 啊 in declarative sentences), Chao (1968: 801–803) argues for several functions of the *ne* particle. Two functions are specific to questions, while three functions are restricted to declarative sentences (i.e., continued state, assertion of equalling degree, and interest in additional information). The ‘mild warning’ function is attested in exclamative sentences. Following Chao (1968), other scholars, such as Han (1995: 116–118) and Zhu (1999: 235–237), suggest two main functions for this particle. One is to mark the ‘continued state’ in declarative sentences, like a progressive marker (Han 1995: 116). This function is often accompanied by the durative aspectual marker 著 *zhe* (Zhu 1999). However, such a function may not be due to the particle, but to the aspectual marker (Li & Thompson 1981: 302–303). The other function of the particle is found in questions. Both studies point out that the use of the *ne* particle is similar to that of the *ma* particle, as they often co-occur with uncertainty elements (Han 1995: 116).

Recently, Fang & Hengeveld (2022: 885–890) have reviewed the functions of the *ne* particle proposed in the literature. They also argue for two different functions. One is a marker of contradiction occurring in declarative, exclamative, and directive sentences, and is labelled *ne*₁. They also mention that the *ne*₁ can be used in questions (2022: 887–888), as in (2.133). I think this example is an exception and may not be accepted by the speaker of Taiwan Mandarin. Moreover, *ma* and *ne* co-occurrence is not allowed (Han 1995: 116; Zhu 1999: 237).

(2.133)

咱們	先	走	吧，	沒	看見	爸爸	正在	忙	著
<i>zánmen</i>	<i>xiān</i>	<i>zǒu</i>	<i>ba</i>	<i>méi</i>	<i>kànjiàn</i>	<i>bàba</i>	<i>zhèngzài</i>	<i>máng</i>	<i>zhe</i>
1PL	<i>first</i>	<i>go</i>	PTCL	NEG	<i>see</i>	<i>father</i>	<i>being</i>	<i>busy</i>	PROG

呢 嗎？

ne ma?

PTCL PTCL

‘Let’s go first. Don’t you see Dad is busy right now?’

(Fang & Hengeveld 2022: 887–888)

The other is *ne*₂ is used as a reinforcer in questions to enhance the interrogative illocutionary in any type of question, except polar questions.

Unlike previous studies that distinguish functions based on sentence types, Chu (2009: 295; 2017: 50) argues for two main functions of the *ne* particle, which apply to

any sentence type. These two main functions are deduced from various functions of the *ne* particle proposed in the literature (Chu 2009: 289–293). The first function is to require the hearer to look for the contrast in the previous context by comparing it to what is said in the current utterance. The second function is to ask the hearer to carry on the discourse. The two functions of the *ne* particle are related to a higher level of discourse purpose, that is, to increase the level of relevance of the previous context (Chu 2009: 298).

The reason that studies still distinguish the *ne* particle used in questions from other sentence types is that the two *ne* particles evolved from two different characters and the two characters were distributed according to sentence type (Qi 2002: 38). The character 爾 *ěr*, 那 *nà*, and 甞 *nǐ* were found in questions, while the character 哩 *lǐ* was found in declarative and exclamative sentences. However, after extensive mixing of characters in use, the correspondence between character and sentence type no longer existed (Qi 2002: 40). Therefore, only one character for this particle (i.e., 呢 *ne*) is preserved in modern Chinese. This unification for the character of the *ne* particle could lead to its single function, as argued in certain studies.

Hence, Li & Thompson (1981: 300–307) argue that in any sentence type, the *ne* particle is used to respond to the hearer's expectation (1981: 307). Discursively, based on the hearer's belief or expectation, the *ne* particle not only tries to connect to the previous utterance but also to call the hearer's particular attention to the information conveyed by this sentence (1981: 300, 304–305). Likewise, in the case of questioning, the *ne* particle is to express the meaning of 'in connection with your claim and expectation, let me find out...' (Li & Thompson 1981: 306) (see also Chu 1998; 2009 for the similar point of view, suggesting that the *ne* particle is used to solicit participation or highlight the relevance of background information). Similarly, Li (2006: 16, 21) considers that the *ne* has only one function as an evaluative marker in both declarative sentences and questions. When using the *ne* particle, the speaker evaluates the information that is potentially unusual and particularly important. In other words, hearers must also pay attention to this information.

No consensus has been reached on the function of the *ne* particle. Still, scholars are trying, very often from different linguistic perspectives, to either suggest multiple functions or to pinpoint a core function of the *ne* particle. The *ne* particle is complicated not only in terms of its functions but also its status as a question particle.

I label the particles discussed in this dissertation, namely *ma*, *ba*, *a*, *ne*, as question particles, regardless of their contribution of interrogative force. However, the status of the *ne* particle is still controversial in the literature (Luo 2013: 39; Pan 2021; Fang & Hengeveld 2022: 885). Given its frequent co-occurrence with *wh*-questions and disjunctive questions, the *ne* particle is labelled *question particle* (Huang, Y.-H. Audrey Li & Li 2009: 237), *wh-particle* (Cheng 1991: 21; Wei 2022: 266). However, the *ne* particle is syntactically optional in questions as shown in (2.131) (see Pan & Paul 2016; Hsu 2020 for further discussion on the optionality of the *ne* particle), other studies name it *sentence-final particle* without the connection to the question (Li 2006: 13; Pan & Paul 2016; Paul & Yan 2022).

Despite the disagreement, it is still reasonable to label the *ne* particle as a question particle, as do Li & Thompson (1981: 305). They label *ne* as a question particle only because it is used in disjunctive, *wh*-questions, and truncated questions. In addition, the terminology of the *ne* particle may be unrelated to its contribution to the force of the sentence, as Cheng (1991: 25) points out in a footnote. This *ne* particle, nonetheless, is in fact closely related to the force of the question.

It is true that, unlike the *ma* particle, the *ne* particle does not function as an element turning a declarative sentence into a question (Li & Thompson 1981: 305). However, the *ne* particle often pertains to questioning and can be strongly associated with interrogative (Halliday & McDonald 2004: 336, 341). It has been further argued that the core meaning of the *ne* particle is an interrogative mood (Qi 2002: 43). The core meaning becomes obvious, especially in a so-called truncated question or fragment question, as shown in the part highlighted in bold in examples (2.134).

- (2.134)a. (我 還 沒 看 完), 你 呢?
 (wǒ hái méi kàn wán), **nǐ ne?**
 1SG still NEG read finish 2SG PTCL
 ‘(I still haven’t finished reading it,) what about you?’
 (Halliday & McDonald 2004: 336)
- b. A:... 告訴 我, 你 媽媽 呢?
 gàosù wǒ, **nǐ mama ne?**
 tell 1SG 2SG.POSS mother PTCL
 ‘Tell me where is your mother?’

B: 在 樓上。
 zài lóushàng.
 at upstairs
 ‘Upstairs’

(Fang & Hengeveld 2022: 889)

The truncated question is formed with a structure of a constituent (e.g., nominal phrases, verbal phrases, temporal phrases, and locational phrases) attached by the *ne* particle (Wei 2022: 267), as shown in examples (2.135). This type of question can also be a conditional clause, with an overt conjunction 如果 *rúguǒ* ‘if’ in (2.135a) or without in (2.135b).

- (2.135) a. 如果 你 是 政治 人物 呢？
 rúguǒ nǐ shì zhèngzhì rénwù ne?
 if 2SG COP politics figure PTCL
 ‘What if you are a political figure?’ (Fang & Hengeveld 2022: 889)
- b. 有 人 不 同意 呢？
 yǒu rén bù tóngyì ne?
 have people NEG agree PTCL
 ‘And if there are people who do not agree?’ (Alleton 1981: 100)

In truncated questions, the question word can be unambiguously re-established from the context (Fang & Hengeveld 2022: 889), such as *what about* in (2.134a), *where* in (2.134b), *what will happen* in (2.135a), *what is to be done* in (2.135b). Given the retrievability of *wh*-words and the acceptability with the *ne* particle, they are often considered questions reduced from *wh*-questions (Chu 2006: 16; Her, Che & Bodomo 2022: 14–16). Moreover, such a truncated structure of conditional clause is highly conventionalised, given that the *ne* particle is absolutely necessary (Fang & Hengeveld 2022: 890). In addition to being a question particle, this *ne* particle is used as a topic marker in truncated questions (Chu 2006: 16; Li 2006: 21; Wei 2013: 162; 2022: 274).

To conclude this sub-section, the functions of the *ne* particle and its status are still an issue of debate. After reviewing the literature, first, I favour the distinction of the *ne* particle between interrogative and non-interrogative sentences. In other words, there are two *ne* particles. Although they are both written with the same character 呢 *ne* in modern Chinese, they may differ in pronunciation. The *ne* particle in questions has a higher pitch. In comparison, the other *ne* particle in non-interrogative scenarios has a relatively lower pitch (Li 2006: 46). Also, I think they function differently in that they are two distinguished particles. In questions, following Chu (2009; 2017), the *ne* particle is used to show relevance with the previous context and gently require the hearer to continue the

dialogue. In a non-interrogative situation, following Li (2006), it is an evaluative marker to highlight the claim that seems unnatural or essential to the speaker. Concerning its status, I do not agree with the argument that the *ne* particle is not a question particle only because it is optional in questions, not to mention that this particle is required in truncated questions. When it comes to the optionality of particles, I think the four particles discussed above can all be optional to some extent. For instance, the *ma* particle, often referred to as a typical question particle, can also be optional in polar questions. However, the four of them can contribute additional information to questions.

The following subsection aims to summarise, after a brief description of the four question particles, the most important points about each particle and their comparison in terms of sentential distribution and pragmatic functions in questions.

2.5.2 Summary of four particles in Mandarin

Four particles in Mandarin, *ma*, *ba*, *a*, and *ne*, are all allowed in both interrogative and non-interrogative settings. Despite the difficulty in determining their actual functions, Table 18 below summarises their distribution in sentence types and their possible functions in questions based on the review offered in Section 2.5.1.

Table 18 Summary of four particles *ma*, *ba*, *a*, *ne*

	<i>ma</i>	<i>ba</i>	<i>a</i>	<i>ne</i>
Declarative	✓	✓	✓	✓
Exclamative	✓	✓	✓	✓
Polar question	✓	✓	✓	✗
<i>Wh</i> -question	✗	✗	✓	✓
Disjunctive	✗	✗	✓	✓
Imperative	✓	✓	✓	✓
Function in questions	To reinforce the speech act (i.e., questioning) to obtain the confirmation from the addressee	To show uncertainty of the speaker less directly and to solicit agreement from the hearer	To reduce the forcefulness and to increase the contextual relevance	To show the significance with the previous context and gently require hearers to continue the dialogue

Given the focus of the dissertation, I want to draw attention to the distribution of the four particles in questions. According to the two-way distinction of question types presented in Section 2.3, the *ma* particle and the *ne* particle are mutually exclusive, as shown in Table 18. As for the *ba* particle, it has the same distribution as the *ma* particle (Li 2006: 30), while the *a* particle is allowed in all sentence types.

The *ma*, *ba*, and *a* particles are all allowed in polar questions, but the functions that they attribute to the polar question are different (Li 2006: 30). This could be due to the varied degree of interrogativity, face-threatening, and predictability among the three particles, as shown in Table 19.

Table 19 Comparison of *ma*, *ba*, and *a* in polar questions

	<i>ma</i>	<i>ba</i>	<i>a</i>
Interrogativity		>	=
Face-threatening		>	=
Predictability		<	=
Confirmation		=	>

As mentioned in Section 2.5.1.1, **the *ma* particle encodes the strongest interrogativity** (Halliday & McDonald 2004: 336) in that the inquirer has a relatively low degree of knowledge concerning the truth. In other words, the answer is maximally unpredictable to the speaker (Li 2006: 35). Thus, polar questions with the *ma* particle would face a higher face-threatening risk.

In contrast, the *ba* particle and the *a* particle both function similarly as a tag question (Li 2006: 34–35; Fang & Hengeveld 2022: 896–897). Using either of them in polar questions, the confirmation requested is minimised (Levinson 2012: 23) so as to mitigate the face-threatening risk toward the addressee (Fang & Hengeveld 2022: 897). In terms of the predictability to answers, Li (2006: 35) argues that the answer is maximally predictable when the *ba* particle is used in polar questions. Although the *a* particle may have no effect on predictability but performs an emotive or expressive function (Li 2006: 55), I would argue that **the *a* particle has the same level of predictability as the *ba* particle because both particles require no answer from the addressee** (Han 1995: 115). The only difference between the two particles lies in that the *a* particle has a prediction opposed to what has been stated in the question, given its expression of astonishment (Paul 2014: 92), while the *ba* particle has a prediction identical to what is said in the question, also meaning to wait for a positive confirmation (Paul & Yan 2022: 191). Therefore, the *ba* and *a* particles have predictability to the answer, but the polarity of their predictions is different. In addition, they act differently regarding the degree of confirmation. **The *ba* particle is more confirmation-inviting than the *a* particle** (Fang & Hengeveld 2022: 896–897). In my opinion, this is may be one of the reasons why the *ba* particle is similar to the *ma* particle in terms of syntactic distribution. Both *ma* and *ba* are only allowed in polar questions (i.e., confirmation-seeking questions) not

information-seeking ones, while the *a* particle can appear in both question types. Accordingly, instead of genuinely seeking confirmation in polar questions, the *a* particle generally checks whether having correctly heard what has been said (Chao 1968: 804; Zhu 1999: 238; Han 1995: 115; Wu 2004: 138).

The *a* and *ne* particles both occur in *wh*-questions but function differently in their discourse ‘deviance’ (Li 2006: 39, 50). Both particles convey different meanings in a discourse that lacks of relevance and consistency in the context. The *a* particle establishes the link with the previous context and calls the hearer’s attention to what is being asked (Li 2006: 50), meaning *Could anyone of you tell me...?* (Li 2006: 55–56). While the *ne* particle reinforces the relevance with the previous context and highlights the importance of asking such a question, for example, *How to...? This is really a question* (Li 2006: 56).

In addition to comparative works on these particles, scholars have paid attention to the sequence of co-occurrence of sentence-final particles in Mandarin (Zhu 1999: 234–235; Li 2006: 61–63; Pan 2021: 8–10; Paul & Yan 2022: 180–181; Fang & Hengeveld 2022: 899–908). These works contributed to discussing the hierarchical order of the sentence-final particles under the generative framework. Different versions of the cartography have been presented, such as in Li (2006: 64–65), Huang (2007: 185), Paul (2014: 83) based on the three distributional classes from Zhu (1999: 234–237); in Simpson (2014), Pan (2018; 2022), Fang & Hengeveld (2022: 898, 908) based on the framework of Functional Discourse Grammar (FDG), and among others. Given the focus of the current study, without going into details on this topic, I only mention it to show that the extensive work on particles in Mandarin is still developing despite a long history of particle study in this language.

To conclude, this sub-section presented and explored the syntactic distribution and their pragmatic functions of the four main question particles (i.e., 嗎 *ma*, 吧 *ba*, 啊 *a*, and 呢 *ne*) in Mandarin. However, I have noted that much of the literature discussed in this section is based on the spoken Mandarin on a small or unclear scale, and while some of discussion of written Mandarin does come from corpora, none of them are specific to Taiwan Mandarin. Therefore, I hope to show the use of question particles through the corpus of novels in Taiwan Mandarin.

In addition, my corpus study of the question particle will be mainly limited to a discussion of syntactic distribution. However, by discussing the pragmatic functions of individual question particles in Mandarin, this helps to account for some syntactic patterns observed in the corpus. Moreover, the purpose of understanding their pragmatic functions is also to establish the *tertium comparationis* with the French particle. It can then lead to a more meaningful discussion and in-depth comparisons between the question particles in Mandarin and French.

2.5.3 Question particles in French

Although the term *particle* appears more frequently in discussions of Asian languages, it is also used when studying French questions, including *est-ce que*, *ça*, and *donc*. In contrast to Mandarin, French particles are less diverse and are used differently. However, when the particles in both languages are used in questions, they are formally invariable and they are already or almost losing their original meaning. In addition, question particles in Mandarin and French do have certain syntactic features (e.g., question type distribution) and pragmatic functions (e.g., to maintain the contextual coherence) in common. Therefore, it is reasonable to label *est-ce que*, *ça*, and *donc* as *question particles*, although this is not the case in every study.

The following sub-sections are dedicated to the discussion of French particle *est-ce que*, *ça*, and *donc* in terms of their use in questions. The examples used in the literature and their quantitative analyses are based on different sources. Some of the literature study spoken French (Coveney 1996; Hansen 2001: 479; Quillard 2001: 58–59; Druetta 2002: 69; Druetta 2011: 26; Dekhissi 2013: 74, 81; Adli 2015: 179–180; Lefevre 2020: 24), while other studies focus on written French (Lefevre 2006; Guryev 2017: 123–133; Lefevre 2018; Reinhardt 2019; Rossi-Gensane et al. 2021; Smirnova & Abeillé 2021; Guryev & Delafontaine 2022: 55–56; Benzitoun 2022: 7).^{12 13}

¹² Although these studies use spoken French, the register of the spoken data may be different. It includes the conversations ranging from informal contexts (e.g., friends, family) to formal settings (e.g., radio debates, university lectures). In addition to spontaneous spoken data, dialogues from the movies are also considered in Dekhissi (2013). Certain corpora are clearly documented in the studies, such as the Corpus of spoken French G.A.R.S d'Aix-en-Provence used in Druetta (2002; 2011), Spoken French sub-corpus of a multilingual database Sgs used in Adli (2015), and a small spoken corpus ESLO2 used in Lefevre (2020).

¹³ The source of the written French is relatively homogeneous. The corpus *Frantext* is extensively used to study the written French, as in Rossi-Gensane et al. (2021), Lefevre (2006; 2018), Smirnova & Abeillé (2021), and Benzitoun (2022). A smaller scale corpus of ten XXIst century detective novels was also built

One of the possible marking strategies for French questions includes the use of particles (cf. Section 2.2.2). The most well-known particle used in French is *est-ce que* ‘is it that’. This particle contains a verb *est* (i.e., the third person singular of the verb *être* ‘be’), a demonstrative pronoun *ce*, and a complementiser *que* (Dryer 2013a). This particle is derived from the grammaticalization of the inversion of a cleft structure (De Boer 1926: 314; Druetta 2011: 28; see Munaro & Pollock 2008: 563 for different opinions). However, certain constraints on its usage indicate that its grammaticalization has not yet been completed (Druetta 2018: 25). For instance, this *est-ce que* particle can only be followed by clauses with finite verbs, but not infinitive structures, as in (2.136) (Druetta 2018: 26). In addition, other constituents can still be inserted between the *est-ce* and *que* (Druetta 2011: 28), as in (2.136).

- Moreover, Paul (2014: 81–82) argues against the fact that *est-ce que* is a particle because such a structure can appear in a negative form (i.e., *n'est-ce pas que*). Despite the controversy on the status and the constraint of using the *est-ce que* particle, this latter is largely considered as a question particle in French questions (Wagner & Pinchon 1962: 537; Harris 1998: 237; Hansen 2001: 471; Coveney 2011: 120; Morin 2017: 21; Druetta 2018: 25; Larrivée & Guryev 2021: 11–13). Therefore, I think it is reasonable to regard the *est-ce que* as a question particle, but also to acknowledge that it is moving toward a full-fledged particle, which distinguishes it from genuine particles.

and used in Reinhardt (2019: 249). In contrast to formal written French, the text messages are also studied using *Le Corpus Suisse de SMS* in Guryev (2017) and Guryev & Delafontaine (2022).

questions and disjunctive questions (Abeillé & Godard 2021: 1402), as shown in the following examples (2.137) repeated from section 2.2.2.

- (2.137) a. *Est-ce qu' elle va au cours ?* (Polar question)
 PTCL 3SG go PREP class
 'Is she going to class?' (Larrivée & Guryev 2021: 12)
- b. *Où est-ce qu' elle va ?* (Wh-question)
 where PTCL 3SG go
 'Where does she go?' (Larrivée & Guryev 2021: 12)
- c. *Est-ce que je continue ou est-ce que j' arrête ?* (Disjunctive question)
 PTCL 1SG continue or PTCL 1SG stop
 'Do I go on or do I stop?' (Coveney 2011: 113)

Although the *est-ce que* particle is categorised by Dryer (2013a) as a question particle that occurs at the beginning of the sentence, it does not appear in the sentence-initial position when it is used in a *wh*-question, as shown in (2.137b). However, some tendency of co-occurrence with question words and some constraints are observed. For instance, the *est-ce que* particle is rarely used with the question word *où* 'where' (Rossi-Gensane et al. 2021: 170). In addition, it is not allowed to co-occur with *quoi* 'what' and *quel* 'which' used as an adjective, as in (2.138).

- (2.138) a. * *Quoi est-ce que tu veux?*
 what PTCL 2SG want
 Intended for 'What do you want?' (Druetta 2018: 36, my translation)
- b. * *Quel est-ce qu' est ton nom ?*
 which PTCL COP POSS.2SG name
 Intended for 'What is your name?' (Druetta 2018: 36, my translation)

Given its occurrence in questions, the *est-ce que* particle is often treated as a marker for questions in Modern French. In other words, this marker in French works similarly to the *ma* particle in Mandarin, which turns affirmative into interrogative (Hansen 2001: 471; Druetta 2011: 29). This may seem logical and plausible in the case of polar questions. However, when it comes to *wh*-questions, the *est-ce que* particle must have functions other than altering the modality, given that *wh*-words can already mark the interrogative force. Diachronically, the *est-ce que* particle is first attested in *wh*-questions and later found in polar questions (Foulet 1921; Rouquier 2002: 98; Rouquier 2003: 339; Marchello-Nizia et al. 2020: 1228). Prior to grammaticalization, the particle was considered a reinforcement marker and was primarily used to emphasise the emotional value in utterances, such as surprise and admiration (Togeby 1974: 126; Jensen 1990:

232). In addition, the abundant usage of the *est-ce que* particle has been attested with the co-occurrence of the question word *que* in Ancient French (Rouquier 2002: 101), even in Modern French (Terry 1970: 481–482; Druetta 2002: 71–72; Larrivée & Guryev 2021: 15). Such frequent co-occurrences result from the deficit properties of the question word *que* in phonology and syntax (Terry 1970: 482; Rouquier 2002: 102; Munaro & Pollock 2008; Druetta 2021: 31; Druetta 2018: 37). As the time progressed, the *est-ce que* particle started to expand into other question words, such as *pourquoi* ‘why’, *comment* ‘how’, *où* ‘where’ (Foulet 1921: 257; Marchello-Nizia et al. 2020: 1229). Moreover, it is assumed that the *est-ce que* particle used in the polar question originated from the structure *qu’est-ce que* (Foulet 1921: 265; Togeby 1974: 60). Consequently, it seems unlikely that the *est-ce que* particle is used to change the modality of a sentence given its first attestation in *wh*-questions and later usage in polar questions. Hence, I argue that the *est-ce que* particle in polar questions is not only a modality-altering particle as it seems at first sight, but also has the function of emphasizing the personal emotion. Otherwise, it would be difficult to account for the marking choices among polar questions because the inversion is also available in French when altering the modality in addition to the particle (Larrivée & Guryev 2021: 12; Abeillé & Godard 2021: 1403).

Regarding the choices to form a question in French, the use of the *est-ce que* particle seems to be favoured in the everyday spoken language (Druetta 2002: 67, 2011: 29). When using the *est-ce que* particle, not only the interrogative modality can be easily and immediately caught by the addressees, but a declarative word order following the particle is also less cognitive demanding for speakers (Guryev & Delafontaine 2022: 64). However, the frequency of the *est-ce que* particle varies from one study to another. As observed Larrivée & Guryev (2021: 15–16) and Guryev & Delafontaine (2022: 61), the frequency documented from previous studies on the spoken French can range from 2 % to 20 % in polar questions. In the same vein, the occurrence of the *est-ce que* particle in *wh*-questions can go from minor as attested in a corpus of movie (Dekhissi 2013: 113) to nearly half of the usage as observed in a spoken corpus of interviews (Coveney 1996: 118). In writing, the *est-ce que* particle is relatively rare in both polar and *wh*-questions, as shown in a corpus of text messages (Guryev 2017: 197; Guryev & Delafontaine 2022: 58) and a corpus of contemporary French novels (Reinhardt 2019: 249–260). In addition, a few small-scale corpus studies based on *Frantext* corpus focus on a specific *wh*-word,

such as *où* ‘where’ and *combien* ‘how much’. Their results show that the co-occurrence of these specific question words with the *est-ce que* particle is close to zero (Rossi-Gensane et al. 2021: 170; Benzitoun 2022: 8).

In addition to the *est-ce que* particle, French has other strategies to mark polar and *wh*-questions, such as intonation and inversion (cf. Section 2.2.2). **The choice between these strategies may be conditioned by two linguistic perspectives: sociolinguistic (including socio-stylistic and socio-demographic factors) and cognitive-pragmatics (including communicative and linguistic aspects)** (Quillard 2001; Coveney 2011: 112; Rossi-Gensane et al. 2021; Guryev & Delafontaine 2022: 63).

From a **sociolinguistic perspective**, the three marking strategies for French questions (i.e., inversion, declarative form, and the *est-ce que* particle) are often compared in terms of their **formality** (Wagner & Pinchon 1962: 539; Coveney 1989: 143–145), as shown in Figure 10.

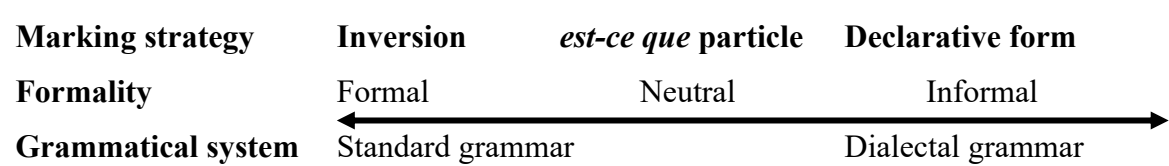


Figure 10 Marking strategies within the degree of formality and grammatical system

Figure 10 shows that these three marking strategies are situated differently on the socio-stylistic spectrum of formality. The inversion and the declarative form are found at the two ends of the extreme (i.e., formal vs. informal), while the *est-ce que* particle is situated in the middle. Note that the use of particles is somehow less elegant or considered a worry of contamination from spoken language in writing (Wagner & Pinchon 1962: 540; Grevisse & Goosse 2007: 498–499; Coveney 2011: 124). Similarly, on the scale of grammatical system, it shows the contrast between two extremes: standard school grammar, which imposes the standard and formal use of the language (e.g., inversion and the *est-ce que* particle) and spontaneous dialectal grammar, which covers more informal use (e.g., declarative form) (Rowlett 2007: 9; Zribi-Hertz 2011: 240; Guryev & Delafontaine 2022: 53). Nevertheless, such a sociolinguistic distinction based on formality may be too simplistic (Hansen 2001: 465). Therefore, studies argue that cognitive-pragmatics may equally account for the marking strategy choices in French questions.

Regarding **cognitive-pragmatics**, studies mainly focus on the interaction between the speaker and the hearer in different contexts. Such an interaction is believed to strongly influence on the marking choices (Hansen 2001: 515). Hence, the choice of morpho-syntactic marking for questions may be related to certain aspects of interaction, including the **interrogative intensity** (Terry 1970), as shown in (2.139) and the **level of *potentiel interactif* ‘interactive potential**, as shown in (2.140) (Guryev & Delafontaine 2022).

(2.139) **Interrogative intensity**

The est-ce que particle > *Inversion* > *Declarative order*

The *est-ce que* particle has a stronger intensity. Speakers choosing the particle in the question demonstrate more interest in knowing the answer from the addressee (see Coveney 1989: 293 for different opinions), while using inversion means the speaker is less curious about the answer. Compared to inversion, given that using declarative order has a higher degree of predictability toward the answers of the addressee (Haan & Van Heuven 2003: 67), it must have the weakest interrogative intensity. Questions with declarative form are used in general to elicit confirmation (Abeillé, Crabbé & Marandin 2012), while the other two marking strategies are more interrogative-oriented and serve as actual questions (Guryev & Delafontaine 2022: 70–71).

(2.140) **Interactive potential**

Declarative order > *Inversion* = *The est-ce que particle*

The interactive potential is defined as the ability of variants to make interaction easier in an ordinary speech context, such as a private, dialogue-like, and spontaneous interaction between interlocutors that are close to each other (Guryev & Delafontaine 2022: 54). On the scale, the declarative order has higher interactive potentiality in that it is preferred in informal conversations. Therefore, it is a variant of strong interactive potential. By contrast, inversion and the *est-ce que* particle have a lower or moderate degree of interactive potential. In other words, they are less preferable in such a context. Accordingly, they are both weak variants (Guryev & Delafontaine 2022: 63, 65).

In addition, the interactive potential links to the factor of ‘face-work’ (i.e., politeness). This factor is also relevant in terms of the choice because an act of questioning is full of potential face-threatening, especially from the speaker’s perspective. Hence, the speaker often prefers using questions with declarative order for private

conversations with socially equal or close interlocutors, whereas inversion and the *est-ce que* particle are used in more distant, unequal situations (Hansen 2001: 477). Therefore, questions with the *est-ce que* particle give a sense of expressing politeness in spoken language (Coveney 1989: 344; Coveney 2011: 141), not to mention that inversion is often considered as formal language use, as shown previously on the scale of formality.

In the communicative framework, other factors are also taken into account, such as the information conveyed and the role of the speaker and addressee (Hansen 2001: 474–478; Abeillé, Crabbé & Marandin 2012: 2). It should also be noted that the nature of the subjects (e.g., first, second or third person; singular or plural) and the tenses of the verb both have an impact on the choice of marking strategies in questions (Wagner & Pinchon 1962: 541; Terry 1970: 482; Hansen 2001; Grevisse & Goosse 2007; Coveney 2011; Adli 2015; Guryev 2017; Guryev & Delafontaine 2022: 73–75). Moreover, media can play a role, such as subtitling (Coveney 2011: 135) or text message (Guryev & Delafontaine 2022: 76).

2.5.3.2 Question particle ‘*ça*’ and ‘*donc*’

In addition to the *est-ce que* particle, there are two other potential question particles in French: a pronoun *ça* ‘that’ (*cela* in written form) and a conjunction *donc* ‘then’. Both are explored and discussed in corpus-based studies (Lefevre 1999: 167–168; Lefevre 2006: 82–83; Lefevre 2018; Reinhardt 2019; Lefevre 2020; Smirnova & Abeillé 2021) and slightly touched upon in syntactic studies of the generative framework (Obenauer 1976: 114; Cheng & Rooryck 2000). Recently, Smirnova & Abeillé (2021) have specifically investigated both particles based on a contemporary corpus of written French (*Frantext* database) and labelled them as two discourse particles in French questions.

Compared to the *est-ce que* particle, **the use of *ça* particle is relatively limited**. On the one hand, it is **only found in *wh*-questions**; on the other hand, it **always follows question words** (Obenauer 1976; Cheng & Rooryck 2000; Smirnova & Abeillé 2021: 240), as in (2.141).

- (2.141)a. *Qui ça ?* (who)
 b. *Quand ça ?* (when)
 c. *Où ça ?* (where)
 d. *Pourquoi ça ?* (why)

- e. *Comment ça ?* (how) (Lefevre 2018: 183)

However, the question words *que* ‘what’, determiners *quel* ‘which’, and *lequel* ‘which one’ are not allowed in such a structure. It should also be noted that the co-occurrence of the *quoi* ‘what’ and the *ça* particle has a low frequency (Lefevre 2018: 190–191; Smirnova & Abeillé 2021: 244, 246). In addition, given that *wh-ça* structure lacks of verbs, it is generally categorised as a **verbless question**, which is common in the spoken language (Druetta 2011: 30; 2018: 33–35), or to a lesser extent in the discourse of written language (Lefevre 2018: 184).

There are two types of verbless *wh*-questions. One is a question word followed by a noun (*wh*-N) as in examples (2.142).

- (2.142)a. *A quand ton mariage ?*
 PREP when 2SG.POSS marriage
 ‘When is your marriage?’ (Lefevre 2018: 183, my translation)
- b. *Dans quel dessein ?*
 PREP what purpose
 ‘In what purpose?’ (Grevisse & Goosse 2007: 490, my translation)

The other is a question word followed by the *ça* particle (*wh-ça* structure). In addition to the bare form as in (2.141), it can be followed by a fragment (*wh-ça*-fragment structure) as in (2.143).

- (2.143) *Qui ça les autres ?*
 who PTCL ART others
 ‘Who are the others?’ (Druetta 2018: 34, my translation)

Both verbless questions (i.e., *wh*-N and *wh-ça*) are not only attested in spoken French, but they also appear in the dialogues of written French. Based on the results from the studies using *Frantext* (Lefevre 2018: 183–184; Smirnova & Abeillé 2021: 252–253), the *wh-ça* is used more frequently than the *wh*-N in the dialogue of texts. This might be because the *wh-ça* structure can easily occur in a verbal question, being placed in-situ in (2.144) or fronted in (2.145).

- (2.144) *Tu as vu qui ça (cet après-midi) ?*
 2SG have see.PTCP who PTCL this afternoon
 ‘Who did you meet this afternoon?’ (Cheng & Rooryck 2000, my translation)

(2.145) *Comment ça je la vois plus ?*
 how PTCL 1SG her see no.more
 ‘How is it possible that I cannot see her anymore?’

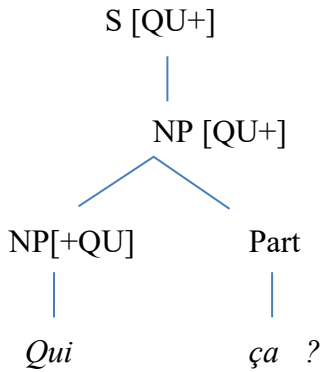
(Druetta 2018: 34, my translation)

I will mainly discuss the verbless questions with the *ça* particle, as my interest of study is the *ça* particle.

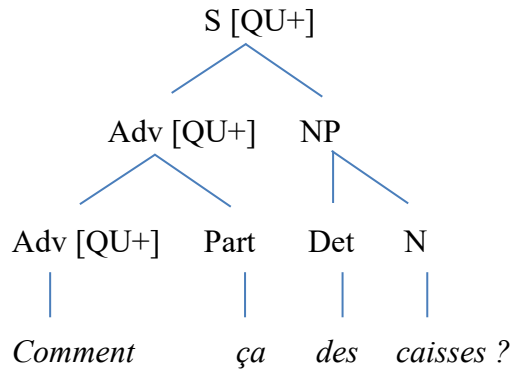
The syntactic structure of *wh-ça* is perceived formally as a combination of a *wh*-word functioning as a predicate followed by a pronominal subject *ça* (Lefeuve 2018: 188; Lefeuve 2020: 29) or a relation between a rheme (i.e., *wh*-word) and a theme (i.e., the *ça* particle) (Lefeuve 1999: 277; Druetta 2018: 34). However, Smirnovaa & Abeillé (2021: 244) argue that the *ça* particle is never a subject. The *wh-ça* and *wh-ça*-fragment are short questions derived from an ellipsis (also see Coveney 2011: 141). The *wh-ça* question, as in (2.141), is *wh*-sluicing. The *wh-ça*-fragment question, as in (2.143) is *wh*-stripping. Following the Direct interpretation approach (Culicover & Jackendoff 2005: 235; Kim & Kim 2023), Smirnovaa & Abeillé (2021) suggest that each *wh-ça* projects its syntactic structure, as shown in (2.146): unary in (2.146a) or binary in (2.146b).

(2.146)

a.



b.



(Smirnova & Abeillé 2021: 253)

Without going into the details of the syntactic derivation, I will pay more attention to the function of the *ça* particle in such an elliptical question. Although studies may disagree on the syntactic derivation of such an elliptical structure, there may be consensus on the discourse function brought by the *ça* particle and this type of question.

Given its nature as a demonstrative pronoun, the *ça* particle has two specific features. One is that it does not require a well-defined antecedent, and the other is that it behaves indifferently to the categorical status of such an antecedent (Riegel, Pellat &

Rioul 2009: 377). In other words, this *ça* particle can refer to any possible candidate of the antecedent in previous discourse as long as it is relatively relevant and salient to what it is saying in the following (Cadiot 1988: 191). The two particularities of the *ça* particle assist in establishing and identifying the anaphoric relation with one or more constituents or propositions in the context, such as the response to A in example (2.147).

(2.147) A: *Quelqu'un me l' a dit.*
 someone 1SG.OBJ it have.3SG say.PTCP
 'Someone told me that.'

B: *Qui ça ?*
 who PTCL
 'Who?'

(Smirnova & Abeillé 2021: 250)

This anaphoric relation can not only link to the closest element but also to more than one referent with similar semantic value (Lefevre 2018: 199–200). In addition, the information of the contextual referent can be implicit, explicit, or even have no correlation with the *wh-ça* structure (Smirnova & Abeillé 2021: 249). When the information may be ambiguous, it is necessary to add a fragment to specify the referent, as in (2.148).

(2.148) A: *Ils ont arrêté Paul.*
 3PL have arrest.PTCP Paul
 'They have arrested Paul.'

B: *Qui ça, ils ?*
 who PTCL 3PL
 'Who they?'

(Smirnova & Abeillé 2021: 252)

In this example, the pronoun *ils* 'they' added in B helps to identify the referent. Hence, the referent of *ça* is the subject *ils* 'they' in the statement of A but not the object *Paul*. Accordingly, on the one hand, the *ça* particle facilitates the link between co-references; on the other, the added fragment can focalise the referent in the cases where the identification work cannot easily be done (e.g., when multiple candidates are likely to be the referent in the previous context).

In conversations, the *wh-ça* question often appears as a reactive utterance instead of a starter (Coveney 2011: 142; Abeillé & Godard 2021: 1116). Its reactive role has two functions: information-seeking and clarification-seeking (Smirnova & Abeillé 2021: 250). The first function is to seek the exact information related to the *wh*-word, as in B in (2.149).

(2.149)A: *J' ai dit que Thérèse était ailleurs.*
 1SG have say.PTCP COMP Thérèse COP.IMPARF elsewhere
 'I said that Thérèse was somewhere else.'

B: *Où ça ?*
 where PTCL
 'Where?'

(Smirnova & Abeillé 2021: 251)

In the context, the question *où ça* 'where' shows that the speaker is looking for the details of a place not explicitly mentioned by the interlocutor A. Accordingly, the type of *wh*-word in the *wh-ça* structure determines the content of the response that the inquirer is seeking, such as *pourquoi ça* 'why' for a justification, *comment ça* 'how come' for an explanation (Lefeuve 2018: 200).

The information-seeking function is also found when the question is formed with the *wh-ça*-fragment, as shown in B in (2.150).

(2.150)A: *Avec ces personnes-là, j' ai été obligé*
 with DEM.PL people-there 1SG have be.PTCP require.PTCP
d' être confronté de me battre
 PREP COP confront.PTCP PREP 1SG.REFL fight
physiquement [...]
 physically
 'With these people, I was obliged to be confronted to fight physically [...]'

B: *Avec euh Où ça, à l' école primaire ?*
 with euh where PTCL PREP ART school primary
 'With, euh... Where, in elementary school?'

A: *Au collège.*
 PREP junior.high.school
 'In junior high school.'

(Druetta 2018: 35)

B in (2.150) is seeking information on the place related to the statement that A has mentioned. This relation is connected by the *ça* particle in that the particle correlates with the previous context. In addition, the fragment added here *à l'école primaire* 'in elementary school' is not a reprise but one of the anticipated responses (Druetta 2018: 34). However, *wh-ça*-fragment structures is more likely to be used as a reprise question (Smirnova & Abeillé 2021: 252), which is the second function of *wh-ça* questions.

The second function is asking for a clarification of the content that appeared in the prior utterance (Noh 1998: 604). Accordingly, this type of *wh-ça* is also known as a reprise question or echo question (Druetta 2018: 34; Smirnova & Abeillé 2021: 261). The effect of reprise can also be attributed to the function that the *ça* particle brings in

discourse, namely the co-reference to the previous statement (Druetta 2018: 34; Lefeuve 2018: 197). The referent of the reprise is often a definite noun phrase (Smirnova & Abeillé 2021: 251), as *Jean-Paul Sartre* in (2.151).

(2.151)A: *Jean-Paul Sartre s' y exerçait bien à*
 Jean-Paul Sartre REFL there practice.IMPARF well PREP
la boxe.
 ART boxing
 'Jean-Paul Sartre was practicing boxing.'

B: *Qui ça ?*
 who PTCL
 'Who?'

(Smirnova & Abeillé 2021: 251)

The effect of reprise becomes clearer when the fragment is added, as in B in (2.152).

(2.152)A: *Avant, au moins, il y avait les mariages [...].*
 before at.least 3SG there have.IMPARF ART marriages
 'Before, at least, there were marriages.'

B: *Avant? Quand ça, avant?*
 before when PTCL before
 'Before? When before?'

(Smirnova & Abeillé 2021: 252)

Generally, the second function is more dominant in *wh-ça* and *wh-ça*-fragment structures (Smirnova & Abeillé 2021: 265). This has to do with the pragmatic function of the *ça* particle, namely, to correlate with the preceding context. In addition to the two functions, the question of *wh-ça* structure encodes an emotion of surprise from the speaker as a reaction to the previous utterance (Abeillé & Godard 2021: 1115).

The *donc* 'then' particle has the same syntactic distribution as the *ça* particle. Both particles must follow the *wh*-words, and they are not compatible with *que* 'what', *determiner quel* 'which' and *lequel* 'which one' (Obenauer 1976: 114; Abeillé & Godard 2021: 1115). However, compared to the *ça* particle, the *donc* particle can easily co-occur with the question word *combien* 'how much' and *quoi* 'what' (Smirnova & Abeillé 2021: 246), as in (2.153).

(2.153)a. *Qui donc ?* (who)

b. *Quoi donc ?* (what)

c. *Pourquoi donc ?* (why)

d. *Où donc ?* (where)

e. *Combien donc ?* (how much)

(Smirnova & Abeillé 2021: 246)

Unlike the *ça* particle, *wh-donc* appears mainly in a verbal construction and is often fronted, as in (2.154).

- (2.154) a. *Qui* *donc* *avait* *parlé ?*
 who PTCL have.IMPARF talk.PTCP
 ‘Who had talked?’ (Smirnova & Abeillé 2021: 246)
- b. *Quand* *donc* *serais-je* *grand ?*
 when PTCL COP.COND-1SG big
 ‘When will I grow up?’ (Smirnova & Abeillé 2021: 246)

Regarding verbless constructions, *wh-donc* appears more frequently by itself as in (2.153), instead of adding another fragment as in B in (2.155).

- (2.155) A: *Tu* *as* *mis* *des* *gens* *là-haut ?*
 2SG have put.PTCP ART people up-there
 ‘Did you put people up there?’
- B: *Où* *donc,* *là-haut ?*
 where PTCL up-there
 ‘Where up-there?’ (Smirnova & Abeillé 2021: 259)

Concerning the syntactic structure of *wh-donc*, it is an adverb added onto the *wh*-word (Lefevre 2006: 82–83; Abeillé & Godard 2021: 1115). It is also regarded as a discourse particle, as in the case of *ça*. Accordingly, Smirnova & Abeillé (2021: 260) argue that the derivation of the *wh-donc* is in an elliptical form based on the Direct interpretation approach (see (2.156)).

- (2.156) a. b.
- S [QU+]

↓

NP [QU+]

↙ ↘

NP[+QU] Part

↓ ↓

Qui *donc* ?

S [QU+]

↙ ↘

Adv [QU+] NP

↙ ↘

Adv [QU+] Part

↓ ↓

Où *donc* *là-haut ?*

(Smirnova & Abeillé 2021: 260)

Despite the same derivation structure as *wh-ça*, the *wh-donc* is often a non-genuine question, mainly when *wh-donc* is fronted in a verbal structure, as in (2.154). Such questions tend to have a rhetorical and self-addressed purpose (Smirnova & Abeillé 2021:

256). However, an elliptical *wh-donc* is often used as a reprise question (Smirnova & Abeillé 2021: 258–259), as in (2.157).

(2.157)A: *Comment elle s' appelle ?*
 how 3SG REFL call
 'How is she called?'

B: *Qui donc ?*
 who PTCL
 'Who?'

A: *La vache.*
 ART cow
 'The cow.'

(Smirnova & Abeillé 2021: 259)

The pragmatic function of *wh-donc* in both verbal or elliptical structures shows that the *donc* particle marks the continuation of the discourse instead of correlating with the preceding context (Smirnova & Abeillé 2021: 261).

2.5.4 Summary of particles in French

The term *particle* seems rarely used in French literature when referring to the three items discussed above (i.e., *est-ce que*, *ça*, and *donc*). However, I think they are legitimate to be labelled as such, given that **they are primarily invariable and gradually or already losing their original semantic meaning when being used in questions**. Despite the occurrence in questions, the three particles differ in terms of syntax, semantics, and pragmatics.

Regarding their **syntactic distribution**, they are either free or strongly restricted to one question type. The *est-ce que* particle is allowed in any question, while the *ça* and *donc* are both limited to *wh*-questions. In addition, the *est-ce que* is required to occur in verbal constructions, where the verb is tensed, whereas the *ça* and *donc* particles appear in both verbal and verbless constructions.

When it comes to **semantics**, the *est-ce que* particle is no longer an inversion of a cleft after the grammaticalization, but a relatively fixed structure assumed to mark interrogative force. As for the *ça* and *donc* particles, they still preserve their original meaning, but these meanings may have little contribution to the *wh-ça* and the *wh-donc* questions. Derived from a demonstrative pronoun, the *ça* particle helps to link the anaphoric relation with the previous context. Such an effect can ensure contextual

coherence (Lefevre 2018). As an adverb marking a conclusion or a result (Hansen 1997: 165), the *donc* particle used in question indicates the continuation of the discourse.

Pragmatically, the difference is first found at the register level. The *est-ce que* is mostly neutral, while the *ça* and *donc* particles are informal or at least frequently found in spoken language (Lefevre 2006: 91; Grevisse & Goosse 2007: 490). Second, regarding interrogative intensity, the *est-ce que* particle seems to be the strongest among them. The *ça* particle has stronger interrogative intensity when used to seek information, whereas it becomes weaker when used for clarification. Likewise, the *donc* particle may be the weakest on the level of intensity because it often occurs in rhetorical questions and self-addressed situations.

Overall, the topic of question particles in French often remains under-researched as most previous studies have focused on all the structural variations of questions, *wh*-questions in particular. The topic on question particles has received less attention, possibly also because the grammaticalization of particles is still in progress. I think *ça* and *donc* may be in the middle of their grammaticalization since their original meanings still have an impact when used as particles. However, *est-ce que* may be almost at the end of the process, but not yet fully grammaticalized. Regardless of their progress in grammaticalization, studies have shown that they are particles-to-be (De Boer 1926: 314; Druetta 2011: 28, 2018: 25). Although the scholars do not yet agree using the term *particle*, they have strived to study both oral and written sources to gain a detailed understanding of French particles. Comparing their quantitative results for the *est-ce que* particle in particular, this particle is not distributed similarly in spoken and written language, and even in different spoken contexts (e.g., formal, informal). While the nature of the data may play a role in the quantity of the particles, the studies show a relatively consistent qualitative analysis of the question particles, including their pragmatic functions and the interrogative intensity. Table 20 summarises the three question particles in French based on their syntactic distribution, interrogative intensity and functions.

Table 20 Summary of the three question particles in French

	<i>est-ce que</i>	<i>ça</i>	<i>donc</i>
Polar question	✓	✗	✗
Wh-question	✓	✓	✓
Disjunctive question	✓	✗	✗
Register	Neutral	Informal	
Interrogative Intensity	Strong	Strong / Weak	Weak
Function in questions	Reinforce the interrogative intensity and show politeness	Identify the referent in precedent context to enhance contextual coherence	Mark the continuation in the discourse

2.5.5 A tentative conclusion on question particles from a cross-linguistic perspective

After reviewing the question particles in Mandarin and French, **Mandarin has a wide range of particles, whereas French does not**. Such a difference may be attributed to the constraint concerning phonetic variations in tonal languages (Fang & Hengeveld 2022: 886). Due to the lack of phonetic variations in pitches and contours, a set of multiple particles can help tonal languages, such as Mandarin, to express certain discourse functions that are usually achieved by virtue of intonation in non-tonal languages, such as French (Hancil, Post & Haselow 2015: 18; Fang & Hengeveld 2022: 886). Although only a small number of question particles are recognised in French, they have drawn attention from scholars, especially in the case of *est-ce que*. In addition, the *ça* and *donc* particles have also gained interest in corpus-based research. The fact that they are not fully grammaticalized as particles accounts for the lack of discussion and the controversy that they can be labelled as *particles* (Hancil, Post & Haselow 2015: 5–6). In brief, the variety of particles and the way in which linguists recognise their morphological categories are very different in Mandarin and French.

Nevertheless, **they do share common features when used in questions**. To the best of my knowledge, the studies of particles (Bailey 2013; Hancil, Post & Haselow 2015; Del Gobbo, Munaro & Poletto 2015), including question particles (Dryer 2013a; Luo & Wu 2017), have been conducted on the scale of typology, but a cross-linguistic study in Mandarin and French has not yet been carried out. Therefore, based on the

insights gained from the literature review, in this summary section, I aim to offer an overview of the contrasts and similarities of question particles in terms of syntactic distribution and discourse functions.

Syntactically, Mandarin makes a clear distinction between the use of the *ma*, *ba*, and *ne* particles depending on the question type. The *ma* and *ba* are restricted to confirmation-seeking questions, whereas the *ne* is only used in information-seeking questions (*wh*- and disjunctive questions). Conversely, the *a* particle applies to any question type in Mandarin. Likewise, the use of the *est-ce que* particle in French can be used across question types. However, the other two particles, *ça* and *donc*, are restricted to *wh*-questions. The summary of the syntactic distribution of the particles in the two languages is shown in Table 21.

Table 21 Syntactic distribution of question particles in Mandarin and French

	Mandarin				French		
	<i>ma</i>	<i>ba</i>	<i>a</i>	<i>ne</i>	<i>est-ce que</i>	<i>ça</i>	<i>donc</i>
Polar	✓	✓	✓	✗	✓	✗	✗
Wh	✗	✗	✓	✓	✓	✓	✓
Disjunctive	✗	✗	✓	✓	✓	✗	✗

Cross-linguistically, the *ma* and *ba* particles are distributed exactly opposite to the *ça* and *donc*. The former is only used in polar questions, while the latter, as the *ne* particle in Mandarin, is limited in information-seeking questions. As for the *a* and *est-ce que* particles, they have the most flexibility to appear in any type of question. However, noted that all these question particles are syntactically optional.

In terms of their **functions** in questions, given that all particles can be optional in both languages, their fundamental function is to reinforce the interrogative force. However, the literature indicates that they differ from each other in terms of interrogative intensity. The *ma* and the *est-ce que* particles both have the strongest intensity because such questions show stronger curiosity as to elicit responses from the addressees. The intensity of other particles is weaker because they are mainly used for clarification and seeking agreement from the interlocutor, not for questioning. In discourse, particles, such as *a* and *ne* in Mandarin and *ça* in French, contribute to contextual coherence. They maintain coherence between questions and their previous statement. In addition to linking

the context, the *donc* particle serves as a marker of continuation. Similarly, the *ne* particle is used to ask the hearer in a less harsh way to carry on the conversation, given the politeness encoded in the *ne* particle (Alleton 1981: 98; Luo & Wu 2017: 88).

Thanks to the previous studies in Mandarin and French respectively, which have argued the syntactic distribution and pragmatic functions of question particles based on the spoken or written corpus, I can make a preliminary argument and draw a theoretical conclusion on the syntactic and functional comparisons between the question particles used in Mandarin and French. However, no cross-linguistic corpus-based study has been carried out on this aspect. Therefore, on the basis of this initial theoretical comparison, I now aim to take a closer look at the similarities and differences in the use of question particles from both syntactic and pragmatic perspectives, using a bidirectional parallel corpus of contemporary novels in Mandarin and French.

2.6 Summary of Chapter 2

This chapter provided an overview of three key topics in Mandarin and French (i.e., taxonomy of question, *wh*-in-situ, and question particles), which will be further explored in the bidirectional parallel corpus analysis.

The first part of this chapter (Sections 2.1 to 2.3) focused on question types and their morphosyntactic structures from both typological and monolingual perspectives. While a three-way distinction of question types is commonly found in the literature, a dichotomic view of question types from Her, Che & Bodomo (2022) offers a more plausible and efficient framework. According to this view, all languages distinguish between two question types at the highest level: confirmation-seeking and information-seeking questions. However, languages differ at the morphosyntactic level, as evidenced by the experimental results from Mandarin and French. **Mandarin is able to distinguish these two question types through the use of adverbs and question form, whereas French does not display such morphosyntactic differentiation.** More specifically, polar question in French can only be distinguished prosodically.

The second part (Section 2.4) focused on the position of question words, with particular attention to *wh*-in-situ, a feature shared by both Mandarin and French. In both languages, the licensing of *wh*-in-situ involves similar syntactic mechanisms, such as LF (Logical Form) movement and binding relations. **Despite these structural and**

theoretical parallels, the nature of *wh*-in-situ differs between two languages. In Mandarin, *wh*-in-situ is a default syntactic structure, whereas in French it is an optional construction. Furthermore, the two languages impose different syntactic constraints on *wh*-in-situ. In Mandarin, both *wh*-arguments and *wh*-adjuncts are generally not sensitive to local intervenors. In contrast, both types of *wh*-elements in French are subject to locality effects. Additionally, while only a subset of *wh*-adjuncts in Mandarin is affected by strong island constraints, such constraints do not appear to apply to *wh*-adjuncts in French.

The final part of the chapter (Section 2.5) examined the syntactic distribution and pragmatic functions for four particles in Mandarin and three in French. **Syntactically, the particles in Mandarin are distributed according to question type, whereas the particle *est-ce que* in French can occur with all types of questions.** From a pragmatic perspective, the **particles in both languages share several common functions**, such as reinforcing interrogative force, mitigating the forcefulness, and enhancing contextual relevance and discourse coherence.

The chapter has provided a theoretical foundation for the three core topics investigated in this dissertation. The literature review here offered several insights that will help the analysis of authentic data in Chapter 4 and 5. Moreover, the interaction between theoretical accounts and corpus-based evidence enables a more comprehensive understanding of the contrastive properties of questions in Mandarin and French. After reviewing this theoretical groundwork, Chapter 3 introduces the bidirectional parallel corpus employed in this study and details its structure and design.

Chapter 3 Data and Methodology: The design, compilation and analysis of a bidirectional parallel corpus

This chapter focuses on the corpus used in the dissertation, namely the comparable and parallel corpora of novels in Taiwan Mandarin and French. First, I will start with an overview of corpus-based cross-linguistic studies and understand how corpus is being used to investigate different topics, particularly with questions. Then, I will describe the compilation of the corpus used in the present cross-linguistic study of questions and explain the rationale for selecting the novels. Finally, I will present the methods used for parallel alignment of source texts and their translation, extraction of target sentences (i.e., questions) and the details of the coding process.

3.1 Corpus-based contrastive studies: tools, languages and topics

Following the discussion regarding the theoretical framework and methodology in Chapter 1, in contradistinction to comparative syntax focusing mostly on theory, contrastive studies are often associated with empirical bases and applied linguistics (Hasselgård 2020: 184), with the initial aim of drawing attention to interlinguistic similarities and differences in the context of second language teaching and learning in particular (McArthur et al. 1992: 261; König 2012: 3; Johansson 2007: 1). However, due to ineffective results in second language teaching and learning after 1970s, it was not until the late 1990s that the contrastive studies experienced a revival of interest and began to attract attention once again (Hasselgård 2020: 185; Ramón García 2021: 395). Such a change is due to the development of the large databases (i.e., corpora). Given that these corpora provide access to authentic data and empirical adequacy in the investigation of languages (Gómez González & Doval-Suárez 2005: 26), contrastive studies largely benefit from these corpora (Granger 2010; Izquierdo & Sanz-Villar 2023: 2). Comparable and parallel corpora, in particular, prove very useful in highlighting differences and similarities between languages (Anthony McEnery & Xiao 2007: 18; Granger 2010). The usefulness of these two kinds of corpora derives from the nature of their composition.

A **comparable corpus** is constructed by collecting texts from different languages according to a shared sampling frame, ensuring the comparability. However, it is acknowledged that achieving full comparability of a comparable corpus might be rarely possible across languages (Aijmer & Lewis 2017: 4). The sampling frame is considered

comparable in that texts are selected based on the same criteria: genres, formality, and time of publication, etc. (Tognini-Bonelli 2001: 59–62; Johansson 2007: 9; Aijmer 2008: 276; Aijmer & Lewis 2017: 3). For example, genres refer to “situationally defined text categories”, including fiction, broadcast, and article, which reflect different communicative situations, purposes, and function of the text (Biber 1994: 245). The dimension of formality concerns linguistic variation across registers, for instance, scientific writing vs. private letters, and parliamentary debate vs. coffee-table conversation (Rissanen 2007: 57). Furthermore, the time span of publication depends on the research focus. A short range may be used for precisely defined research questions on contemporary language use, whereas a longer period is more suitable for investigations of language change (Claridge 2008: 243). These sampling criteria help ensure that a comparable corpus is balanced and representative in terms of the language varieties and genres it contained (Anthony McEnery & Xiao 2007: 21). Despite these important considerations, constructing a comparable corpus seems easier to assemble since some existing corpora can be used already (Aijmer 2008: 277).

In addition to comparable corpus, another major type of corpus used in contrastive studies is the **parallel corpus**. This type of corpus comprises bilingual or multilingual source texts and their translations (Johansson 2007: 9; Aijmer 2008: 276; Granger & Lefer 2020b: 167), and is often referred to as a *translation corpus*. Depending on text availability and the research objectives, parallel corpora can be further distinguished by monodirectional (e.g., English original to Mandarin translation) or bidirectional (e.g., English original to Mandarin translation and Mandarin original to English translation) (Lefer 2020: 258–259). Although a parallel corpus does not require the same stringent comparability criteria as a comparable corpus, several challenges arise during its compilation. These include issues related to text availability, metadata collection, alignment, and multilingual linguistic annotations (Lefer 2020: 267–271). The issue of text availability often results from asymmetry and imbalance between translation directions: fewer texts may be translated from one direction to another, or certain text types may exist only in one of the two directions. Metadata collection is also essential, as it involves gathering information about the text (e.g., authors, translators, publisher, genre, publication date) and translator’s profiles (e.g., professional vs. volunteer). Such information helps researchers understand the contextual characteristics of the parallel

texts. Alignment, which enables the identification of translation equivalents at different linguistic levels (e.g., lexical and syntactic) is another crucial issue. However, it is not a straightforward task, since translators may split or merge sentences, making sentence-level correspondence difficult to establish. Finally, the multilingual linguistic annotation facilitates more sophisticated studies using multilingual corpus. Nonetheless, challenges persist regarding the choice between language-neutral and language-specific annotation schemes (Johansson 2007: 306).

Despite comparability requirements for comparable corpora and the challenges inherent in compiling parallel corpora, each type contributes uniquely to contrastive studies due to the nature of its textual composition. **Comparable corpora**, which consist of naturally occurring language data, **allow researchers to identify cross-linguistic similarities and differences independently of translation influence**. In contrast, **parallel corpora enable the examination of mutual correspondences and divergence between languages** through the systematic comparison of original texts and their translations. Nevertheless, certain drawbacks have been encountered when using only one of the corpora for cross-linguistic studies (see the discussion in Granger & Lefer 2020b: 169–170). For a comparable corpus, it can show the distinctiveness from the two independent linguistic systems, respectively, but it is hard to achieve a perfect comparability of corpus and cannot easily identify the elements to be compared. By contrast, using a parallel corpus, it is easier to identify the *prima facie* equivalents (Tognini-Bonelli 2001: 134) in the two languages, but the translation languages can be inevitably influenced by the source texts and by translators (Aijmer & Lewis 2017: 3; Granger & Lefer 2020b: 170). Therefore, Johansson (1998; 2007: 33–34) suggests combining parallel and comparable corpora within a single bidirectional parallel corpus model. This integrated method (also called *Johansson procedure*, De Baets, Vandevoorde & De Sutter 2020: 87) has received widespread recognition and is still one of the main trends in contrastive studies (Granger & Lefer 2020a: 18). By integrating comparable and parallel corpora, contrastive studies can draw on the strengths of each corpus type. Comparable corpora reflect the general language use of each language, allowing researchers to ‘safely’ identify the similarities and differences between the languages under comparison (Johansson 2007: 10). Meanwhile, parallel corpora provide a reliable *tertium comparationis* (i.e., translation equivalence), which refers to items in two

languages conveying more or less the same meaning (Hasselgård 2010: 6; Granger & Lefer 2020b: 169) Therefore, this method offers a balanced foundation for conducting contrastive studies.

When applying the integrated method, there are two possible approaches (Johansson 2007: 33; Granger & Lefer 2020b: 171). One approach is to start with comparable and then parallel, or vice versa. Only the type of research and the type of linguistic items investigated can determine the choice of approach (Granger & Lefer 2020b: 171; Hasselgård 2020: 190). For instance, contrastive studies often start with a comparable corpus analysis and then proceed with a parallel corpus analysis. Firstly, a comparable corpus serves the purpose of offering a full linguistic description and contrasting the patterns found in each language. Subsequently, a parallel corpus highlights the cross-linguistic correspondences of lexical items, lexical-syntactic patterns or grammatical structures investigated (Lefer 2020: 262). Indeed, by integrating advantages of both comparable and parallel corpora, this approach facilitates the growth and expansion of corpus-based contrastive studies (Hasselgård 2020: 193), which, in turn, can be seen as enabling more systematic cross-linguistic comparisons.

As the field of corpus-based contrastive studies has developed over the past 30 years, studies have attempted to depict its progress over time from various perspectives, including the topics discussed and the languages compared. Given the focus of the dissertation, the following discussion will be mainly based on the investigations carried out by Hasselgård (2020). This study provides a comprehensive overview of the evolution in the field thanks to a detailed survey of the publications in the journal *Languages in Contrast* published by John Benjamins. She first confirms the increasing number of corpus-based studies since the 1990s (see also Defrancq 2015: 1). One possible explanation for this trend is the growing interest in the use of multilingual corpora (Johansson 2012: 4–5). As for the predominant topics in the field, the discussion on discourse, lexis, and syntax are frequently addressed in corpus-based contrastive studies (Hasselgård 2020: 195). Regarding the languages being compared, the majority of corpus-based studies focus on English paired with either the Germanic or Romance languages. However, a rise is also attested when comparing other languages to English. The findings suggest that the contrastive studies have grown quite adept at incorporating corpora into their research. Nevertheless, the diversity of language pairs examined

remains somewhat restricted, largely within a ‘comfort zone’ where English is the dominant language. Moreover, questions remain an understudied topic in corpus-based contrastive research, as noted by Curry & Chambers (2017: 327) and further evidenced with the limited number of studies reviewed by Axelsson (2020: 238–239). Due to the focus of the study, in the following paragraphs, I review the few corpus-based studies devoted to questions.

Considered as the first corpus-based investigation of questions, Wikberg (1996) examines question types and structures (e.g., declarative form, tags, cleft sentences) on the basis of the English-Norwegian Parallel Corpus (ENCP). He highlights the overuse of tags in Norwegian translation from English due to the lack of equivalent use of tags in Norwegian. In addition, questions in declarative form show considerable formal variation when translated into another language, which includes inversion and declarative sentences that convey suggestions or statements (Wikberg 1996: 24). Following this study, there are studies of questions focusing on the comparison of different language pairs and on question-related topics, as summarised in Table 22.

Table 22 Selected corpus-based studies on questions

Study		
Wikberg (1996)	Language	English-Norwegian
	Topic	Question types and structures (pragmatic and discourse levels)
	Corpus	English-Norwegian Parallel Corpus (ENPC)
	Corpus type	Written
	Main findings	- Overuse of tags in Norwegian translation - Variation of questions in declarative form in translation
Malá (2007)	Language	English and Czech
	Topic	Negative polar questions
	Corpus	Parallel corpus of Czech and English

	Corpus type	Written
	Main findings	- Czech negative questions are mostly translated into English with an inversion - The Czech negative questions are reflected in English in the sentence types prototypically associated with a discourse function other than a question, such as statement, exclamation, and directive
Malá (2010)	Language	English and Czech
	Topic	English declarative polar questions
	Corpus	Parallel Czech - English corpus
	Corpus type	Written
	Main findings	- Both in English and in Czech, the contextual boundness of declarative is often introduced by additive or contrastive conjunctions and conjuncts - The unmarked English declarative questions are translated with bias markers in Czech counterparts - Czech declarative questions containing bias markers can be translated with the constructions other than declarative questions in English

Axelsson (2009)	Language	English and Swedish
	Topic	Tag questions
	Corpus	English-Swedish Parallel Corpus (ESPC)
	Corpus type	Written
	Main findings	- The most typical Swedish translation is a declarative tagged with invariant - The most typical Swedish source is a declarative question with the modal particle, which is multifunctional
Gómez González (2014)	Language	English, Spanish and Portuguese
	Topic	Canonical tag questions
	Corpus	International Corpus of English-Great Britain (ICE-GB) and Reference Corpora for Spoken Romance Languages (C-ORAL-ROM) for Peninsular Spanish and for European Portuguese
	Corpus type	Spoken
	Main findings	- Canonical tag questions are more common in Spanish and Portuguese - The main type of canonical tag consists of a declarative form followed by a tag across three languages - Canonical tag questions in English show less flexibility in terms of functions (e.g.,

		informational, regulatory) than their Spanish and Portuguese counterparts
Buysse (2017)	Language	English and Dutch
	Topic	Question tags
	Corpus	English-Dutch parallel corpus
	Corpus type	Written
	Main findings	- Most English constant polarity tags have a correspondent in Dutch, although there is a lack of specific shared feature between two languages - English canonical question tags are translated into Dutch with the interjection and modal particle - The modal particle in Dutch has two different functions according to its syntactic position
Curry & Chambers (2017)	Language	English and French
	Topic	Functions of questions
	Corpus	Multilingual comparable corpus of research articles
	Corpus type	Written
	Main findings	- French uses more questions than English - French uses the functions, 'framing the discourse' and

		‘organizing the text’, more than English - French has more varied forms of questions than English
Axelsson (2020)	Language	English-Swedish
	Topic	All question types
	Corpus	English- Swedish Parallel Corpus
	Corpus type	Written
	Main findings	- The question type is more often changed in translations from English into Swedish, except for polar questions - Translation patterns specific to one of the languages, such as modal particles in Swedish and <i>wh</i> -fragments such as those with <i>what about</i> in English - Declarative questions are different phenomena between English and Swedish - Both underuse and overuse of several question types can be attributed to source-language influence

Most studies are based on corpora compiled from written data (e.g., novels) to investigate question-related phenomena, with particular focus on tag questions and declarative questions. They primarily employ parallel corpora and their findings successfully demonstrate how questions or specific questions (e.g., tags and declarative questions) are translated across languages, both structurally and functionally. Nevertheless, corpus-based contrastive studies of questions still have considerable room for further development, both in thematic scope and the range of language pairs examined.

Even though corpus-based contrastive studies have been developing for more than thirty years and an increasing number of multilingual corpora have become available to researchers, the range of languages and the selection of topics explored remain relatively narrow, suggesting potential areas for future research. Similarly, corpus-based studies of questions mainly focus on two specific forms (i.e., declarative questions and tags), as noted by Axelsson (2020: 239). One possible reason for the limited number of corpus-based studies on questions may be the relative typological proximity of the languages most often selected, typically European languages. While these languages yield valuable insights into both structures and function of questions in translations, their structural similarities in questions (e.g., a declarative form and inversion) may reduce the likelihood of identifying salient cross-linguistic contrasts, which are often of particular interest in contrastive studies (Salkie, Aijmer & Barlow 1998: vi; Ramón García 2021: 394). In other words, the inclusion of typologically diverse languages may help reveal broader patterns and unexplored contrasts.

Therefore, in view of the relevant issues raised in the field, such as topic and linguistic diversity, the focus of this dissertation centres on the topic of questions. The comparison is based on the language pair, Taiwan Mandarin and French. On the one hand, this topic aims to fill the gap that has been neglected for a long time. On the other hand, the choice of this language pair for comparison directly addresses the underrepresentation of Mandarin in corpus-based studies (Izquierdo & Sanz-Villar 2023: 4), thereby filling a notable gap in the field. Furthermore, the typological differences between Taiwan Mandarin and French offer an opportunity to examine how question forms are realised across structurally diverse languages, thus contributing to a deeper understanding of cross-linguistic variation.

3.2 Corpus design and compilation

Comparable and parallel corpora of Chinese (Xiao & Hu 2015: 38) mostly represent Mandarin Chinese from Mainland China rather than Taiwan Mandarin. It has long been attested in Chinese linguistics that the Mandarin Chinese spoken in these two geographical areas differs not only phonologically and lexically, but also syntactically (Cheng 1985; Kubler 1985: 157; Xiao & Hu 2015: 38; Sandel, Yueh & Lu 2017: 123). Accordingly, given the focus of the present research on Taiwan Mandarin and French, I

decided to compile my own **bidirectional parallel corpus of novels** from both languages. Following the design of the bidirectional parallel corpora proposed in contrastive studies, Figure 11 presents graphical representation of the corpus used in this dissertation.

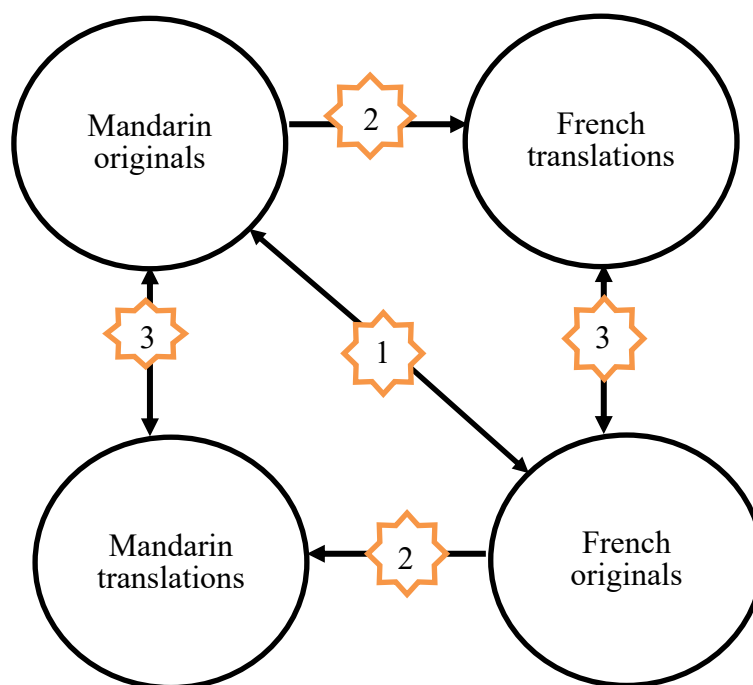


Figure 11 Design of bidirectional parallel corpus of Taiwan Mandarin and French novels

This bidirectional parallel corpus consists of four sub-corpora: original novels published in Taiwan Mandarin (Mandarin originals) and in French (French originals), along with their translations into the other language, that is, Mandarin translations and French translations. This bidirectional parallel corpus can be explored through **three analytical approaches**. First, the **bilingual comparable approach** (indicated as link 1) compares the two sets of original texts in Mandarin and French. Second, the **bidirectional parallel approach** (indicated as link 2) operates in two directions: from Mandarin originals to French translations and from French originals to Mandarin translations. Finally, the **monolingual comparable approach** (indicated as link 3) also functions bidirectionally. This approach is typically used in translation studies to compare original and translated language (original vs. translated Mandarin and original vs. translated French).

The self-built bidirectional parallel corpus contains novels, that is written language and written representation of fictional dialogues. The main reason for this is that spoken

or audio-visual data is not easily accessible for the language pair Taiwan Mandarin and French. In addition, questions occur less frequently in spoken language in general (Zimmermann & Kaiser 2019: 133). The specific criteria used to select in the bidirectional parallel corpus are delineated in the following subsections. I will detail and discuss the selection criteria applied for the novels included in the corpus, starting with Taiwan Mandarin novels.

3.2.1 Selection of Taiwan Mandarin novels

As part of Sinophone literature, Taiwan literature has evolved through distinct historical periods, influenced by its unique location at the crossroads of the west Pacific Ocean (Wang 2007: vii–x; Lin & Li 2023: 1–2). In addition, the historical specificity and diversity of Taiwan, reflected in its colonial past, multilingual society, and complex cultural identity, have made Taiwan literature more distinctive and differentiated from other ‘Chinese literatures’ (Pino & Rabut 2024a: 123, 131). From the Japanese colonization, the post-war Nationalist control to the island’s democratization, the field of Taiwan literature has grown into a place full of opportunities for openness, change, and growth; a place where different ethnic groups and communities interact (Pino & Rabut 2024b: 14).

The concept of Taiwan literature emerged in the early 1980s (Tang 1999: 379, 382). In the past, Taiwan literature occupied a peripheral position within Sinophone literature (Pino 2006: 164), and especially in the context of World literature, to which its access and recognition have long been limited due to the dominance of works from European and North American (Lin & Li 2023: 5). However, Taiwan literature has endeavoured to create meaningful relations with other cultures and literatures through translation into other languages and adaptation into various forms, such as a theatrical production (Chiu 2018: 13–17). This conforms with the definition qualifying as World literature, i.e., a mode of circulation, reading, and interaction with other literary works from different countries and regions (Damrosch 2003: 4–5; Chen 2016: 177–178; Chiu 2018: 13–15). To achieve this goal, the translation of Taiwan literature into other languages has therefore become essential (Lin & Li 2023: 12). In addition to English, three European languages (i.e., German, French, Italian) are the most common target languages for translation of Taiwan literature (Pino 2006: 155–156; Pino & Rabut 2024b: 15).

Regarding translations for the Francophone public, the first French translation of Taiwan literature was published in 1979, that is a short story, "Sayonara, Zaijian" by Huang Chun-ming (Pino 2006: 158; Lin-Rosolato 2011: 18; Gaffric 2011: 147). However, it was not until after 2000 that the translation of Taiwan literature became widely recognised (Pino & Rabut 2024a: 122). Moreover, it was a time when the rate and number of publications largely surpassed those during the previous 21 years (1979–2000) (Pino 2006: 163–164). Such a phenomenon can be seen as a consequence of *Translation policy*. This includes factors determining which text-types are selected for translation during a given period (Tourey 2012: 82). Hence, increase in the translation of Taiwan literature can be attributed to the efforts of editors, publishers, translators, literary critics, scholars, foundations, medias, and government (Lin-Rosolato 2011: 20–21; Gaffric 2011: 152; Pino & Rabut 2024b: 14). Together, they play a decisive role in determining which works are translated and published in French. In addition, several crucial factors, such as geopolitical issues (e.g., Taiwan's special relationship with Mainland China), historical perspectives (e.g., traces of colonization by different authorities in Taiwan), and ideology (e.g., Sinocentric vs. anti-Sinocentric perspectives), also shape the selection of literary works. These factors may limit the diversity of works selected (Lin & Li 2023: 6; Pino & Rabut 2024a: 125). This, in turn, can possibly affect the choice of works to be included in a corpus aimed at linguistic analysis, especially when a corpus with good cross-linguistic comparability is needed.

Consequently, despite a steady trend of Taiwan Mandarin works being published in French over the past ten years (Rivet 2024: 177), the various factors mentioned above still play a role in the translation of Taiwan literature. This has resulted in a relatively limited range of works available for my bidirectional parallel corpus. In addition, due to the synchronic perspective taken here, I decided to exclude works published prior to 2010, despite their recent translation into French. After carefully considering various options, I selected a total of six novels, mostly in the fantasy and detective genres, all originally published in Taiwan Mandarin after 2010, as shown in Table 23.

Table 23 Selection of novels in Taiwan Mandarin and their translations in French

Source	Translated	Genre
Title (Taiwan Mandarin) (Author) (Year published)	Title (French) (Translator) (Year published)	
複眼人 <i>Fùyǎnrén</i> (吳明益) (2011)	<i>L'homme aux yeux à facettes</i> (Gwennaël Gaffric) (2014)	Fiction, fantasy, environment
私家偵探 <i>Sījiā zhēntàn</i> (紀蔚然) (2011)	<i>Rue du dragon couché</i> (Emmanuelle Péchenart) (2019)	Fiction, detective
天橋上的魔術師 <i>Tiānqiáoshàng de móshùshī</i> (吳明益) (2012)	<i>Le Magicien sur la passerelle</i> (Gwennaël Gaffric) (2017)	Fiction, fantasy
泡沫戰爭 <i>Pàomò zhànzhēng</i> (高翊峰) (2014)	<i>La guerre des bulles</i> (Gwennaël Gaffric) (2017)	Fiction, fantasy
炒飯狙擊手 <i>Chǎofàn jūjīshǒu</i> (張國立) (2019)	<i>Le sniper, son wok et son fusil</i> (Alexis Brossollet) (2021)	Fiction, detective, noir fiction
鬼地方 <i>Guǐdìfāng</i> (陳思宏) (2019)	<i>Ghost Town</i> (Emmanuelle Péchenart) (2023)	Fiction, fantasy

Two main criteria were used to ensure the comparability of the bilingual corpus: publication time (only novels originally published after 2010 were selected) and genre (limited to fantasy and detective fiction). In addition, the inclusion of the six novels was also influenced by **domestic reception** and **recognition by French publishers**.

First, **the degree of domestic reception** is reflected in the reviews, search frequency, and recognition through winning or being nominated for prizes. A survey was

conducted using two book review platforms, *aNobii*,¹⁴ one of the most common review platforms in the Sinophone area (Lin & Chiu 2010: 273), and *Openbook*,¹⁵ founded by the Taiwan Reading Association (Chiu 2023: 68). Table 24 and Table 25 show the reception of the selected novels based on the reader ratings and the number of relevant entries retrieved through keyword searches for works and their authors.

Table 24 Reception survey of selected Taiwan Mandarin novels

	<i>aNobii</i>	<i>Openbook</i>
複眼人 <i>Fùyǎnrén</i>	80 reviews 4.5/5 (565 ratings)	48 entries
私家偵探 <i>Sījiā zhēntàn</i>	68 reviews 4/5 (394 ratings)	69 entries
天橋上的魔術師 <i>Tiānqiáoshàng de móshùshī</i>	108 reviews 4.5/5 (817 ratings)	179 entries
泡沫戰爭 <i>Pàomò zhànzhēng</i>	0 review 3/5 (5 ratings)	5 entries
炒飯狙擊手 <i>Chǎofàn jūjīshǒu</i>	1 review 4/5 (15 ratings)	38 entries
鬼地方 <i>Guǐdìfāng</i>	1 review 4.5/5 (33 ratings)	108 entries

According to the survey of novels, the novels are rated on average four out of five points, with an average of 74.5 searches per novel. Notably, two novels have more than one hundred relevant entries: *Tiānqiáoshàng de móshùshī* and *Guǐdìfāng*. The popularity of the novel *Tiānqiáoshàng de móshùshī* is attributed to its television adaptation. In contrast, the great visibility of the novel *Guǐdìfāng* can be attributed to the queer issue it addresses, which is relevant to the legalization of same-sex marriage in Taiwan in 2019. Although the other four novels do not have higher number of entries, they are generally well-rated, except for the novel *Pàomò zhànzhēng*. According to Fangming Chen, a scholar of Taiwan literature, the relatively low rating of this novel may be attributed to its title and content being too ‘children’s fairy-tale’. Nevertheless, its thematic focus coincides with a significant social event in Taiwan (i.e., the Sunflower Student Movement in 2014). Through its use of social satire and anti-utopian narrative techniques, the novel *Pàomò*

¹⁴ <https://www.anobii.com/zh-Hant> (Date of last visit: January 2, 2025)

¹⁵ <https://www.openbook.org.tw/> (Date of last visit: January 2, 2025)

zhànzhēng offers a profound reflection on contemporary Taiwan society and its political movements. Despite receiving limited attention in terms of ratings, the novel holds considerably great value and merits inclusion in the corpus as a representative work of modern Taiwan literature.

Table 25 Reception survey of the authors from Taiwan

Author	<i>Openbook</i>
吳明益 Wu Ming-yi	623 entries
陳思宏 Kevin Chen	122 entries
張國立 Chang Kuo-li	81 entries
高翊峰 Kao Yi-feng	40 entries
紀蔚然 Chi Wei-jan	10 entries

Regarding the authors, Wu Ming-yi is undoubtedly the most sought-after author, given the popularity of his works in Taiwan literature. He is followed by Kevin Chen, who has publicly come out as queer author and whose works reflect gender issues in contemporary Taiwanese society. The relatively high number of entries ($n = 81$) for Chang Kuo-li is likely due to the diversity of his works, which span poems, screenplays, novels, and travel literature. In contrast, the lower number of entries for Kao Yi-feng may be attributed to his smaller body of published novels. However, all three of his novels have been shortlisted for major literary awards, such as *Taiwan Literature Golden Laurel Award* (台灣長篇小說金典獎 *táiwān chángpiān xiǎoshuō jīngdiǎn jiǎng*) and *Taipei International Book Exhibition Book Prize* (台北國際書展大獎 *táiběi guójì shūzhǎn dàjiǎng*). Finally, the author Chi Wei-jan has only ten entries, possibly because his works are mainly for theatre rather than fiction. It is worth noting that the novel *Sījiā zhēntàn* is his debut work in fiction. However, including an author like him enrich the corpus by adding diversity in authorship. His background as a playwright may contribute to distinctive narrative structures and dialogue that reflect more colloquial language use in fiction.

Although the reception of these works and authors in domestic literary communities varies from one novel to another and from one author to another, the fact that their novels have been translated into other languages indicates that these works are circulating spatially and culturally in the space of World literature (Chiu 2018: 17). In other words, these five Taiwanese authors and their works have expanded their horizon to the whole

world, as the genres and topics of these novels illustrate the uniqueness of Taiwan literature.

Accordingly, genre appears to play a decisive role in determining whether these six novels are selected for translation into French, as **French publishers** tend to specialise in or prefer particular literary genre. *Sījiā zhēntàn* (Publisher: Calmann-Lévy/ Collection: Calmann-Lévy Noir) and *Chǎofàn jūjīshǒu* (Publisher: Gallimard/ Serie noire) are published in a collection of noir series in France. *Pàomò zhànzhēng* is published in *Mirobole*, which aims to showcase literary works in the fantasy genre. In addition to genre, the issue of homosexuality in *Guǐdīfāng* (Publisher: Seuil) may be linked to one of the dominant themes in modern Taiwan literature, namely queer issue (Lin & Li 2023: 6). As for *Fùyǎnrén* (Publisher: You Feng) and *Tiānqiáoshàng de móshùshī* (Publisher: Asiathèque/ Collection: Taiwan Fiction), their selection may result from the decision of the publisher, possibly influenced by an orientalist perspective or by a collection's viewpoint that provides a vision of Taiwan from its cultural vitality and social richness (Pino & Rabut 2024a: 136).

Finally, note that translators also play an important factor in introducing Taiwan literature to Francophone readers. Among the six selected novels, there are only three different translators. This limited number of translators can be partly attributed to the linguistic diversity in Taiwan, which is also reflected in its literary works. Such a multilingualism increases the complexity of translation, as most French translators are trained primarily in Mandarin Chinese (Gaffric 2011: 141, 143–144) and may not be familiar with other languages present in Taiwan literature.

Overall, despite the growing number of French translations of Taiwan literature, the selection of the six Taiwan Mandarin novels examined above is based on a relatively limited amount of works. This likely reflects the choices made by several actors, such as the financial support from the government, publishers, and translators. In addition, the domestic reception of the works and the authors also play a significant role. Both works and the authors may be received differently based on the frequency of entries, but their thematic concerns and the recognition they have received from literary awards in Taiwan suggest that they can be considered representative of the main characteristics of Taiwan literature during this period.

3.2.2 Selection of French novels

French literature undoubtedly has a more extensive history than Taiwan literature, given that the earliest text of French literature dates to the end of 9th century (Bergez et al. 1995: 11). With such a long history, French literature is rich in genres and themes. Due to its great diversity and the status of French as a *lingua franca* in Europe during the 17th and 18th centuries (Wright 2016: 134), French literature has long circulated throughout the world and has been translations into various languages.

During the Japanese colonization in Asia, Taiwan began to engage with foreign literatures, particularly Japanese literature, through translation (Chen 2016: 189). As for French literary works, although some have been gradually translated and imported into Taiwan's literary market, it was not until the 1990s that they were directly translated from French without going through a pivot language, such as English or Japanese (Editor group 2016: 142). By now, Taiwan is the fourth biggest market in Asia for French literature, preceded by Japan, China, and South Korea.¹⁶ According to an academic survey of foreign literature translated into Taiwan Mandarin and government reports on publication and trends (National Central Library 2022; 2023), French literature normally occupies the first three largest languages translated in Taiwan, after English (Tsay & Dong 2009: 40, 41, 43). Hence, the high visibility of French literature in Taiwan can be attributed to the institutional support from different organizations. For example, *le programme franco-taïwanais d'aide à la publication Hupinching* provides funding for the work of translating French into Taiwan Mandarin;¹⁷ the *Bureau International de l'Édition française* participates in different book fairs and promotes French publication worldwide; the *Association Taïwanaise des Traducteurs de Français* organises workshops for translators in Taiwan (Editor group 2016: 145–146).

Despite the great success and reception of French literature in Taiwan's market, compared to Taiwan literature in the Francophone world, the Mandarin translation of French literature in Taiwan appears to be criticized for its lack of diversity and consistency. This is mainly due to the publishing policy oriented by marketing (Tsay & Dong 2009: 48; Editor group 2016: 146), which is largely shaped by the preference of the readers in Taiwan. According to an online report, Taiwanese readers are considerably

¹⁶ Le marché du livre français traduit à Taïwan : <https://voila.tw/livres-fr/livres-francais-traduits-a-taiwan/>

¹⁷ Le domaine du livre à Taiwan : <https://france-taipei.org/Le-domaine-du-livre-a-Taiwan>

open to heterogeneous subjects, and the two main publications from French literature in Taiwan are novels and youth literature.¹⁸ Although it means that the selection of the French novels for the bidirectional parallel corpus may be less restricted than in the other translation direction, maintaining the comparability of the corpus necessitates applying consistent selection criteria. Therefore, I applied the same selection criteria as for the novels from Taiwan Mandarin based on the time of publication (i.e., originally published after 2010) and genre (i.e., fantasy and detective) for the selection of French novels, as shown in Table 26.

¹⁸ Le marché du livre français traduit à Taïwan : <https://voila.tw/livres-fr/livres-francais-traduits-a-taiwan/>

Table 26 Selection of novels in French and their translation in Taiwan Mandarin

Source	Translated	Genre
Title (French) (Author) (Published year)	Title (Taiwan Mandarin) (Translator) (Published year)	
<i>Le voleur d'ombres</i> (Marc Levy) (2010)	偷影子的人 <i>Tōu yǐngzi de rén</i> (Duan Yun-ling) (2011)	Fiction, fantasy
<i>Gataca</i> (Franck Thilliez) (2011)	左撇子的殺意 <i>Zuǒpiēzi de shāyì</i> (Chang Chiao-wen) (2012)	Fiction, detective
<i>Un avion sans elle</i> (Michel Bussi) (2013)	沒有她的飛機 <i>Méiyǒu tā de fēijī</i> (Liang Sophie) (2014)	Fiction, detective
<i>La vie des elfes</i> (Muriel Barbery) (2015)	精靈少女 <i>Jīnglíng shàonǚ</i> (Wu Qing-zhu/ Duan Yung-ling/ Nian Geng-jia) (2015)	Fiction, fantasy
<i>Trois jours et une vie</i> (Pierre Lemaitre) (2016)	三天一生 <i>Sāntiān yīshēng</i> (Miao Yung-hua) (2018)	Fiction, detective, noir fiction
<i>L'Anomalie</i> (Hervé Le Tellier) (2020)	異常 <i>Yícháng</i> (Chen Yung-Wei) (2024)	Fiction, fantasy

Although market trends and the reception from Taiwanese readers play a crucial role in the selection of French literary works for publication in Taiwan, publishers in Taiwan also closely follow literary trends in France. Accordingly, half of the French novels translated and published in Taiwan after 2007 are works by the best-selling authors in France (Liao 2019: 74–75), such as Marcy Levy, Michel Bussi, Muriel Barbery, and Pierre Lemaitre (Bessard-Banquy, Ducas & Gefen 2021: 488). In addition, literary works

that have won literary awards, such as *Le prix Goncourt*, are usually translated into Taiwan Mandarin, such as *L'Anomalie*. In fact, **the reception and recognition of works and authors in France** serve as important criteria for selecting which titles to translate and publish in Taiwan. Therefore, to better understand how the popularity and critical reception of French works in France affect their chance of being translated into Taiwan Mandarin, I conducted a survey on two of the largest French social networks for book readers, *Babelio* and *Booknode* (Einman 2018: 192; Deley 2019: 109, 111).^{19, 20} Table 27 shows the number of reviews and ratings of the six authors.

Table 27 Reception survey of the authors from France

Author	<i>Babelio</i>	<i>Booknode</i>	Total
<i>Franck Thilliez</i>	13,044 reviews	4,698 reviews	17,742 reviews
	4.17/5	8.01/10	12.18/15
	(102,785 ratings)	(9,568 ratings)	(112,353 ratings)
<i>Marc Levy</i>	7,015 reviews	7,421 reviews	14,436 reviews
	3.69/5	7.71/10	11.4/15
	(86,623 ratings)	(12,891 ratings)	(99,514 ratings)
<i>Michel Bussi</i>	9,708 reviews	2,954 reviews	12,662 reviews
	3.91/5	7.61/10	11.52/15
	(67,591 ratings)	(5,764 ratings)	(73,355 ratings)
<i>Pierre Lemaitre</i>	8,317 reviews	1,363 reviews	9,680 reviews
	4.16/5	7.8/10	11.96/15
	(67,896 ratings)	(2,580 ratings)	(70,476 ratings)
<i>Hervé Le Tellier</i>	1,858 reviews	168 reviews	2,026 reviews
	3.66/5	7.32/10	10.98/15
	(14,326 ratings)	(430 ratings)	(14,756 ratings)
<i>Muriel Barbery</i>	1,565 reviews	433 reviews	1,998 reviews
	3.78/5	7.47/10	11.25/15
	(18,564 ratings)	(810 ratings)	(18,997 ratings)

¹⁹ <https://www.babelio.com/> (Date of last visit: January 15, 2025)

²⁰ <https://booknode.com/> (Date of last visit: January 15, 2025)

Based on the total number of reviews and the ratings from the two websites, the authors considered best-sellers differ greatly in their level of online visibility. The gap can reach over ten thousand reviews when comparing *Marc Levy* ($n = 14,436$) and *Muriel Barbery* ($n = 1,998$). However, their average ratings still confirm their popularity as best-selling authors, with the exception of Hervé Le Tellier, whose rating is likely due to the visibility gained after receiving *Le Prix Goncourt*. Secondly, regarding their work included in the corpus, they also receive various degrees of recognition based on the reviews and the ratings from the readers, as shown in Table 28.

Table 28 Reception survey of selected French novels

Novel	<i>Babelio</i>	<i>Booknode</i>	Total
<i>L'Anomalie</i>	1,441 reviews	152 reviews	1,593 reviews
	3.65/5	4/5	7.65/10
	(12,396 ratings)	(394 ratings)	(12,790 ratings)
<i>Un avion sans elle</i>	986 reviews	411 reviews	1,397 reviews
	4.01/5	4/5	8.01/10
	(10,876 ratings)	(745 ratings)	(11,621 ratings)
<i>Trois jours et une vie</i>	720 reviews	118 reviews	838 reviews
	3.82/5	3.5/5	7.32/10
	(4,604 ratings)	(214 ratings)	(4,818 ratings)
<i>Le voleur d'ombres</i>	343 reviews	438 reviews	781 reviews
	3.72/5	4/5	7.72/10
	(3,126 ratings)	(719 ratings)	(3,845 ratings)
<i>Gataca</i>	359 reviews	155 reviews	514 reviews
	4.27/5	4.5/5	8.77/10
	(4,051 ratings)	(354 ratings)	(4,405 ratings)
<i>La vie des elfes</i>	106 reviews	15 reviews	121 reviews
	2.73/5	3/5	5.73/10
	(364 ratings)	(30 ratings)	(394 ratings)

From this table, only the two novels, *L'Anomalie* and *Un avion sans elle*, have received over a thousand reviews. The former won *Le Prix Goncourt* in 2020, while the latter received *Les Prix des Maisons de la presse – Roman* in 2012. Although the other works may not be as widely known to the public, they are rated on average above 7 points out

of 10, except the novel *La vie des elfes*. Based on some of the reviews, readers approached this book with expectations built on author's prior popularity and success.

Finally, the pivotal **role of Taiwanese publishers** in the promotion and diversification of French literature is also undeniable. This contribution stems from their strategic positioning and keen marketing intelligence. Taiwanese publishers not only closely track the sales and recognition of French literature in France (Liao 2019: 98), but also paying attention to the evolution of the tastes of Taiwanese readers. In addition, the publication efforts are partially facilitated by institute support, with financial resources and aid provided by both government and local associations. This extended backing has, to a certain extent, contributed to the diversification of French literature publications of in Taiwan, where the local market continues to dictate publishing policies.

3.2.3 Summary on the selection of novels for the bilingual corpus

In the previous two sub-sections, I have explained the rationale for selecting of the novels in both languages. The selection reflected the **publication and translation strategies** that determine which works are considered marketable or representative. These strategies are shaped factors, such as government support, publishers, translators, and domestic recognition. Although these factors play an essential role in the publishing industries of both Taiwan and France, the two countries hold slightly different perspectives on importing literature from each other. **For Taiwanese publishers, French novels translated into Taiwan Mandarin tend to be well-known domestically**, namely, they are often bestsellers or works by bestselling authors. In contrast, **French publishers welcome Taiwanese novels that have gained recognition within Taiwan while also foregrounding the uniqueness of Taiwanese culture and society**. Such differing publication and translation strategies, where Taiwanese publishers prioritise commercial success, while French publishers focus on cultural representation, have indeed highlighted the particularity of each other's literature for local readers. Nevertheless, they have also, to some extent, limited the diversity of literary choices, and consequently, the range of materials available for this study. Practical constraints, such as institutional resources for obtaining texts, further shaped the selection. To maintain a certain level of comparability for the self-built corpus, given the limited availability of novels translated from Taiwan Mandarin into French, I focused on works from the fantasy and detective

genres. Moreover, since this study adopts a synchronic approach to questions, only works originally published after 2010 were included. Following Krzeszowski (1990: 26) for ensuring the comparability of corpora in cross-linguistic research, the **two selection criteria (i.e., genres and publication time) serve to guarantee the comparability of the present corpus**. These considerations, combined with the need to maintain comparability across genres and publication periods, guided the choice of the twelve novels included in the corpus. Other factors, such as authorship and translators, may also influence the study's outcome and will be taken into account during analysis.

Following the two selection criteria, this self-built bidirectional parallel corpus consists of twelve original novels, six in Taiwan Mandarin and six in French. Given that this dissertation is methodologically positioned as a corpus-based study, the corpus needs to be available in electronic format to undergo automated analysis. The next section presents some of the technical aspects of the dissertation, namely, the preparation of the data, including corpus digitalization and alignment, and data extraction and coding.

3.3 Data preparation: digitalization, alignment, extraction

Several steps were taken to prepare the data. First, I **digitalised** the twelve original novels and their translations using the OCR function embedded in Google Document. Second, I **aligned** the original texts with their translations using the software editor InterText (Vondříčka 2014). Then, all original and aligned texts were uploaded to Sketch Engine (Kilgariff et al. 2014). Third, given that these texts were not tagged for parts of speech and a question mark helps directly identify questions in writing (Kearsley 1976: 358), Flöck & Geluykens (2015) suggest eliciting questions by **identifying sentences containing question marks**. However, given that question words (e.g., *who*, *what*, *where*) and word order change (i.e., subject-verb inversion) can already be used as indicators of questionhood, questions are not always marked with a question mark. Moreover, a question mark does not necessarily cover all the sentences and fragments that function as questions (Axelsson 2020: 241). In other words, the results of extraction targeting question marks may not be comprehensive. However, this method is still commonly practiced, as seen in Wikberg (1996), Curry & Chambers (2017: 335), and Axelsson (2020). In addition, since the question mark used in Modern Mandarin was imported from western writing style on the second half to the end of 19th century (Yuan, Guan & Yue

2002: 281–283; Lin & Lin 2008: 107–108), the prescriptive use of question marks in Mandarin and French is relatively similar. In both languages, the use of a question mark does not always indicate a sentence functioning as a genuine question; it may also appear in rhetorical questions (National Languages Committee 2008: 10; Abeillé & Godard 2021: 2200–2201). Accordingly, I used the advanced Concordance function (search on the character: ?) in Sketch Engine to automatically extract all the sentences containing question marks. The results are shown in Table 29 for original languages and Table 30 for translated languages.

Table 29 Question marks in the original Mandarin and French

	Mandarin	French
Raw frequencies	3,921	4,224
Relative frequencies per 100,000 tokens	762	652
Corpus size (tokens)	514,540	647,180

Table 30 Question marks in translated Mandarin and French

	Mandarin (translation)	French (translation)
Raw frequencies	4,433	4,188
Relative frequencies per 100,000 tokens	737	600
Corpus size (tokens)	601,139	697,504

Finally, after extracting all sentences with question marks from the four sub-corpora (original Taiwan Mandarin, original French, translated Taiwan Mandarin, and translated French), I proceeded to **data coding** for three dimensions: question type, morphosyntactic structure, and register.

3.4 Data coding: question type, morphosyntactic structure, register

With the questions extracted from the corpus, I coded them for their **question types, morphosyntactic structures, and registers** in fiction. Before delving into the data coding scheme I used, it is important to stress that some automatically retrieved examples were excluded from the final dataset used in this study. One category is questions in other languages, such as Taiwan Southern Min in (3.1), and English in (3.2) and (3.3).

- (3.1) 寡濟 錢?
 guā tsē tsínn?
 how.much money
 ‘How much is it?’ (天橋上的魔術師 *Tiānqiáoshàng de móshùshī*)
- (3.2) *One night only?* (炒飯狙擊手 *Chǎofàn jūjīshǒu*)
- (3.3) *Beautiful men in uniforms?* (*L’Anomalie*)

The dataset also excludes interjections that occur with question marks, since such cases are difficult to categorise, such as 啊 *a*, 喂 *wéi* ‘hello (phone greeting)’ in Mandarin; *hein*, *pardon* ‘sorry’, *allô* ‘hello (phone greeting)’, *oui* ‘yes (phone answering)’ in French. The next three sub-sections present the coding scheme for the three dimensions: *question type*, *morpho-syntactic structure*, and *register*.

3.4.1 Question type

The first dimension ***question types*** follows the conventional three question types distinction.

Following Stivers & Enfield (2010: 2621–2622), the coding scheme for *question type* is based on the logical semantic structure and pragmatic intention:

Polar question	Question that makes relevant affirmation/confirmation or disconfirmation e.g., <i>Are you a student?</i>
<i>Wh</i> -question	Question that seeks the identity of one element, such as person, place, time e.g., <i>Who stole my newspaper?</i> (Stivers & Enfield 2010: 2621)
Disjunctive question	Question that includes the proposal of a restricted set of alternative answers in their formulation e.g., <i>Do you want corn or flour tortillas?</i> (Stivers & Enfield 2010: 2622)

In addition, as shown in Sections 2.2.1 and 2.2.2, certain unique morpho-syntactic structures or constituents are also efficient to determine the type of question in Mandarin and French. For instance, in Mandarin, the 嗎 *ma* particle only appears in polar questions;

the A-not-A structure and VP-Neg are exclusive to disjunctive questions. In both languages, a connective *or* is used in disjunctive questions, and question words are mainly reserved for *wh*-questions. The coding for morphosyntactic structures is presented in the next section.

3.4.2 Morphosyntactic structure

The second dimension, *morphosyntactic structures*, involves morphological and syntactic marking strategies mentioned in Section 2.1.3. Within a single language, each question type may have its own marking strategies (e.g., question words, disjunctive constructions) and sometimes one marking strategy can be shared among different question types (e.g., particles). Across languages, they can employ language-specific strategies, such as A-not-A structure in Mandarin and inversion in French. By contrast, both languages can also share the same strategy as observed in polar questions with declarative word order. These polar questions are not formed with any conventional morpho-syntactic marking strategy, such as inversion or particles. When expressed in spoken language, these polar questions rely on intonation instead of conventional morphosyntactic marking strategies. Even though it is not easy to ascertain whether these polar questions are truly marked with intonation in the texts, unless through certain narrative description (Leech & Short 2007: 258), I have taken these cases into account and coded them as *No inversion (declarative)* in both languages. This code means that these polar questions have the same word order as declarative sentences (or without inversion, from a French-language perspective). In the following paragraphs, I will present the coding criteria for morphosyntactic structures in both languages in two tables. Each table is followed by a brief explanation of the corresponding coding criteria and by an illustrative example retrieved from the corpus.

The coding for morphosyntactic structures in **Mandarin** questions is shown in Table 31.

Table 31 Coding for morphosyntactic structures in Mandarin questions

Polar	<i>Wh</i>	Disjunctive
- No inversion (declarative) - Particle	- No particle - Particle	- <i>Or</i> - A-not-A - VP-Neg - A-not-A + Particle - Or + Particle

Polar questions in Mandarin were coded according to two criteria. The first, *No inversion (declarative)*, designates questions that retain declarative word order and lack any morphosyntactic marking, as shown in (3.4).

- (3.4) 還 不 退休？
hái bú tuìxiū?
 yet NEG retire
 ‘Still not retired?’ (炒飯狙擊手 *Chǎofàn jūjīshǒu*)

The second criterion is *Particle*, meaning that the polar question is followed by one of the particles (e.g., *ma*, *ba*, *a*), as shown in (3.5) with a particle *ma*.

- (3.5) 你 先生 開車 嗎？
nǐ xiānshēng kāichē ma?
 2SG husband drive PTCL
 ‘Does your husband drive?’ (私家偵探 *Sījiā zhēntàn*)

***Wh*-questions** in Mandarin were also coded for two criteria. The first criterion is *No particle*, meaning the *wh*-question is not followed by any particle (e.g., *ne*, *a*), as shown in (3.6).

- (3.6) 你 怎麼 知道？
nǐ zěnmē zhīdào?
 2SG how know
 ‘How did you know?’ (泡沫戰爭 *Pàomò zhànzhēng*)

The second criterion, *Particle*, is for the *wh*-questions followed by either *a* or *ne* particle, as shown in (3.7) with the *ne* particle.

- (3.7) 她 種 什麼 呢？
tā zhòng shénme ne?
 3SG plant what PTCL
 ‘What did she plant?’ (複眼人 *Fùyǎnrén*)

Disjunctive questions have five criteria. The first one *Or* shows that disjunctive questions use a connective *or* (e.g., 還是 *háishì*), as shown in (3.8).

- (3.8) 眼睛 大 還是 小?
yǎnjīng dà háishì xiǎo?
 eye big or small
 ‘Eyes, big or small?’ (鬼地方 *Guǐdìfāng*)

Note that certain cases without an explicit connective were also coded as *Or*. For instance, the two options are juxtaposed without a connective, as seen in (3.9) with 男的 *nánde* ‘male’ and 女的 *nǚde* ‘female’, and in (3.10) 是 *shì* ‘be’ and 否 *fǒ* ‘not’.

- (3.9) 駕駛 是 男的 女的?
jiàshǐ shì nánde nǚde?
 driver COP mal female
 ‘Is the driver a man or a woman?’ (私家偵探 *Sījiā zhēntàn*)

- (3.10) 是 否 該 跟 他 說 什麼?
shì fǒ gāi gēn tā shuō shénme?
 COP NEG should PREP 3SG say something
 ‘Should I say something to him?’ (鬼地方 *Guǐdìfāng*)

The second criterion is a distinctive structure for disjunctive questions in Mandarin. That is an *A-not-A* structure, such as 是不是 *shìbúshì* in (3.11).

- (3.11) 這 究竟 是-不-是 魔術?
zhè jiùjìng shì-bú-shì móshù?
 DEM after.all COP-NEG-COP magic
 ‘Is this magic or not?’ (天橋上的魔術師 *Tiānqiáoshàng de móshùshī*)

The third criterion is another distinctive structure, namely *VP-Neg*. This structure shows an ellipsis of the verb from its negative part, but it retains the negation morpheme, such as 沒 *méi* in (3.12).

- (3.12) 到底 查到 沒?
dàodǐ chádào méi?
 after.all find.out NEG
 ‘Have you found out yet?’ (炒飯狙擊手 *Chǎofàn jūjīshǒu*)

In addition to the three criteria, I coded disjunctive questions followed by particles (e.g., 呢 *ne*, 啊 *a*), as *Or+particle* in (3.13) and *A-not-A+particle* in (3.14).

(3.13) 那 算是 我 的 心 還是 他 的 心 呢？
 nà suànshì wǒ de xīn háishì tā de xīn ne?
 DEM consider 1SG DE heart or 3SG DE heart PTCL
 ‘Is that my heart or his?’ (天橋上的魔術師 *Tiānqiáoshàng de móshùshī*)

(3.14) 有沒有 明確 的 時間點 呢？
 yǒu-méi-yǒu míngquè de shíjiāndiǎn ne?
 have-NEG-have specific DE time.point PTCL
 ‘Is there a specific point in time?’ (鬼地方 *Guǐdìfāng*)

When coding morphosyntactic markings, other pieces of information were also coded, such as the question word in *wh*-questions (e.g., *who*, *what*, *where*, *when*, *how*) in both languages and the particle used in Mandarin (e.g., *ma*, *ba*, *a*, *ne*, *a*). Although the analysis of morphosyntactic structures primarily focuses on questions with typical (i.e., complete sentential) forms, coding fragment questions (cf. Sections 2.5.1.4) ensures that additional data are available for potential future studies. Accordingly, fragment questions were also coded for their morpho-syntactic structures (e.g., verb phrase, noun phrase, prepositional phrase, insubordination phrase, and tags) and for the use of particles.

Regarding **French**, the coding is similar to that of Mandarin, except for inversion and in-situ, as shown in Table 32.

Table 32 Coding for morphosyntactic structures in French questions

Polar	<i>Wh</i>	Disjunctive
- No inversion (declarative) - Inversion - Particle <i>est-ce que</i>	- No inversion (<i>wh</i>-subject/ object/ adverbial) - Inversion (<i>wh</i>-object/ adverbial) - In-situ - Particle <i>est-ce que</i> - Particle <i>ça</i> - Particle <i>donc</i>	- No inversion + <i>ou</i> - Inversion + <i>ou</i> - Particle <i>est-ce que</i> + <i>ou</i>

Polar questions in French can be expressed through three morphosyntactic strategies. The first, *No inversion (declarative)*, is shown in (3.15).

- (3.15) *Tu veux un café ?*
 2SG want ART coffee
 ‘Would you like a coffee?’ (L’Anomalie)

The second criterion is *Inversion* as in (3.16). The different types of inversion (cf. Section 2.2.2) are not coded in this study.

- (3.16) *Peut-on se rencontrer ?*
 can-3SG REFL meet
 ‘Can we meet?’ (Un avion sans elle)

Finally, the third criterion is *Particle est-ce que* as in (3.17).

- (3.17) *Est-ce que vous la reconnaissez ?*
 PTCL 2PL 3SG.FMN recognise
 ‘Do you recognise her?’ (Gataca)

As for the coding of **wh-questions**, the situation is more complex given that various structures are attested in the literature (cf. Section 2.2.2). I have selected the four most frequent ones and two extra particles for the present analysis. In addition, the function of question words is also coded, such as *wh*-subject, *wh*-object, and *wh*-adverbial.

The first criterion is *No inversion*, meaning that the question words are placed initially and followed by a declarative order, as in (3.18) - (3.20).

- (3.18) *Qui t’ a dit ça ?* (*wh*-subject)
 who 2SG.OBJ have tell.PTCP that
 ‘Who told you that?’ (Trois jours et une vie)
- (3.19) *À quoi tu joues ?* (*wh*-object)
 PREP what 2SG play
 ‘What are you playing at?’ (Le voleur d’ombres)
- (3.20) *Comment c’ est possible ?* (*wh*-adverbial)
 how DEM COP possible
 ‘How is this possible?’ (Un avion sans elle)

The second criterion is *Inversion*, meaning the question word is followed by an inversion, regardless of the type of inversion, as shown in (3.21) and (3.22).

- (3.21) *Que veux-tu faire plus tard ?* (*wh*-object)
 what want-2SG do much late
 ‘What do you want to do later?’ (L’Anomalie)
- (3.22) *Quand cela est-il arrivé ?* (*wh*-adverbial)
 when DEM COP-3SG happen.PTCP
 ‘When did this happen?’ (Gataca)

The third criterion *In-situ* represents a phenomenon where a question word is found in the same position as the answer. Namely, the question words are found at the end of the sentence, as *quelle année* ‘which year’ shown in (3.23).

- (3.23) *Vous êtes en quelle année ?*
 2SG COP PREP which year
 ‘What year are you?’ (Trois jours et une vie)

The fourth criterion is the *est-ce que* particle, which is only used with a fronted question word, as shown in (3.24). This criterion also includes the form *est-ce qui* (Coveney 2011: 117; Reinhardt 2019: 260), as shown in (3.25).

- (3.24) *Qu’ est-ce que tu crois ?*
 what PTCL 2SG believe
 ‘What do you think?’ (Un avion sans elle)

- (3.25) *Qu’ est-ce qui n’ est pas pareil ?*
 what PTCL NEG COP NEG same
 ‘What’s not the same?’ (La vie des elfes)

The last two criteria are related to the discourse particle, *ça* ‘that’ as in (3.26) and *donc* ‘then’ as in (3.27).

- (3.26) *Comment ça, un policier?*
 how PTCL ART police
 ‘What do you mean, a policeman?’ (Gataca)

- (3.27) *Pourquoi donc?*
 why PTCL
 ‘Why is that?’ (L’Anomalie)

French **disjunctive questions** follow the same morphosyntactic criteria as polar questions. However, they usually include the connective *ou* ‘or’ to join two nominal or sentential options. Accordingly, the first criterion is *No inversion + ou*, as shown in (3.28).

- (3.28) *Tu ne crois pas ou tu en es sûr ?*
 2SG NEG believe NEG or 2SG it COP sure
 ‘You don’t believe it or you’re sure about it?’ (Le voleur d’ombres)

The second criterion is *Inversion+ ou*, as shown in (3.29).

- (3.29) *Faisait-elle partie du rêve ou de la réalité ?*
 do.IMPART-3SG part ART dream or PREP ART reality
 ‘Was it part of a dream or reality?’ (Trois jours et une vie)

Disjunctive questions using the third criterion *Particle est-ce que + ou* are not attested in original French in the corpus. However, such usage is suggested in the literature, as illustrated in example (3.30) (see Section 2.2.2).

- (3.30) *Est-ce que je continue ou est-ce que j'arrête ?*
 PTCL 1SG continue or PTCL 1SG stop
 'Do I go on or do I stop?' (Coveney 2011: 113)

Like in the coding of *morpho-syntactic structures* for Mandarin questions, *wh*-words and particles (i.e., *est-ce que*, *ça*, and *donc*) were also coded for French questions. In addition, fragment questions in French (cf. Section 2.5.3.2) were also coded according to their morphosyntactic structures as noun phrase, verb phrase, prepositional phrase, insubordination, tags (e.g., *hein*, *c'est ça*, *n'est-ce pas*; Abeillé, Crabbé & Marandin 2012: 3). Since the present study focuses on the morphosyntactic structures of fully formed questions, fragment questions were mainly coded for particle analysis, specifically the *ça* and *donc* particles.

3.4.3 Register

The third coding dimension is **register** (narrative vs. dialogue), following the approach used by Ebeling & Ebeling (2020) to distinguish different language style in fiction. Linguistic studies have noted variations in language use across different registers, such as fiction, news, and academic writing (Biber 1995: 18–23; Biber et al. 1999: 4, 29–35). However, the situation may seem more complicated within a specific register, such as fiction (Biber & Conrad 2019: 132). This register, on the one hand, is detached from the real world by constructing its fictional world. On the other hand, the world knowledge is somehow required in such a fictional context. Consequently, the language produced in fiction should not merely be an aesthetic choice but it can truly show the function of real language use. To such an extent, a distinction between narration and speech seems reasonable (Egbert & Mahlberg 2020: 74). Moreover, it is reasonable to assume that dialogues reflect, to some extent, the use of question in spoken language such as spontaneous conversations, as pointed out Guillot-Barbance, Pincemin & Lavrentiev (2018: 54). They argue that these spoken scenes in fiction are explicitly presented as a reproduction of spoken words and represent the passages of direct speech in the texts, either fully or partially. However, I also recognise that dialogues in fiction may not completely correspond to the daily spontaneous conversations produced by speakers,

since they are elaborated in advance for fictional characters by authors (Lefevre 2020: 23). Nonetheless, I think that the application of the register distinction within fictions can at least compensate for the lack of spoken data in the present corpus and provide valuable insights for the contrastive study of questions.

Accordingly, following previous research on the topic (Egbert & Mahlberg 2020: 74; Ebeling & Ebeling 2020: 289–290), I distinguished **between narrative and dialogue in fiction based on punctuation**. In Mandarin, questions are marked with a quotation mark in the dialogue, as shown in (3.31). As for French, questions are presented with several different punctuation marks in dialogues (Abeillé & Godard 2021: 2205, 2206, 2208, 2209): a long dash, as in (3.32); a quotation mark, as in (3.33); or a colon, as in (3.34).

- (3.31) 「 這 怎麼 用 ? 」
zhè zěnmē yòng?
DEM how use
‘How does this work?’ (複眼人 *Fùyǎnrén*)
- (3.32) — *Que veux-tu jouer ?*
what want-2SG play
‘What do you want to play?’ (*La vie des elfes*)
- (3.33) « *Comment est-ce possible ?* »
how COP-DEM possible
‘How is it possible?’ (*L’Anomalie*)
- (3.34) CW: *Connaissez-vous cette chanson ?*
know-2SG DEM song
‘CW: Do you know this song?’ (*L’Anomalie*)

The above examples are coded as **dialogue** for register, because they represent **direct and free direct speech**, whose words and syntactic structures are reported exactly as uttered by the original speakers in the fiction (Semino & Short 2004: 12). In addition, in the (free) direct speech, the words considered as the quoted verbatim are presented within a quotation mark, such as inverted commas (Ebeling & Ebeling 2020: 289–290), as shown in (3.35) for direct speech and (3.36) for free direct speech.

- (3.35) “*The bastards,*” *she cursed.* (Ebeling & Ebeling 2020: 289)
- (3.36) “*The filthy stinking fascist bastards.*” (Ebeling & Ebeling 2020: 289)

The only difference between both speech presentation is the presence of the reporting clause, such as *she cursed* in (3.35) for direct speech. By contrast, free direct speech

contains no reporting clause and it presents the words of character or original speaker with no apparent “interference” from the narrator or reporter (Semino & Short 2004: 10–11). However, as pointed out by Leech & Short (2007: 258), certain extreme cases of free direct speech might not contain any punctuation. For the sake of coding consistency, any questions not marked by the above punctuations used in Mandarin and French (i.e., long dash, quotation mark, and colon) and clearly not part of a dialogue are coded as narrative. Note that free indirect speech can be ambiguous, since it combines deictic and other features associated with direct and indirect speech (Semino & Short 2004: 13). Despite its ambiguity, such a speech form is coded as *narrative* in this study. There are several reasons for this. First, free indirect speech is typically found in the context of narrative reports (Leech & Short 2007: 261). Second, although it is the closest indirect speech form to direct speech, readers can feel the narrator intervening between them and the character (Semino & Short 2004: 14). Also, it is hard to identify such a hybrid speech form in the text. Moreover, no standardised linguistic features clearly define the context in which free indirect speech can occur (Leech & Short 2007: 264). Finally, an empirical study shows that it is the least frequently occurring form of speech presentation in fiction (Semino & Short 2004: 13, 82). Thus, I coded free indirect speech as a form of narration.

In addition to the two categories *dialogue* and *narrative*, another category *other* is used to cover cases that are not part of *dialogue* or *narrative*, such as sentences in a text message, the iteration of a text, and the headline of a news report.

3.5 Data coding: interrater reliability

The coding for the three dimensions outlined above in Mandarin and French was checked with interrater reliability (Spooren & Degand 2010; Lange 2011). This is to ensure the robustness of the coding scheme used in the current dissertation. The process of interrater reliability assessment in both languages involved another coder (one for each language) and included the following steps: (1) a comprehensive briefing on the coding scheme, including its categories and examples, (2) a trial with 20 sentences randomly extracted from the dataset, and (3) a discussion after the trial. After resolving and discussing any problems encountered during the trial, the raters in both languages were given a sample set of data. A Mandarin native rater from Taiwan, who is a PhD student in linguistics in Belgium, coded 20% of the sentences ($n = 790$) randomly sampled from

the original Mandarin dataset; a Belgian-French native rater, who is a postdoctoral researcher in linguistics, coded 20% of the sentences ($n = 850$) randomly sampled from the original French dataset. The following sub-section is divided into three parts. Each part presents the results of the interrater reliability by reporting the agreement rate and the kappa statistic score. The kappa score, used as a benchmark, indicates the strength of the agreement, as shown in Table 33 (Landis & Koch 1977: 165; Artstein & Poesio 2008: 576).

Table 33 Kappa statistic score and the correspondent strength of agreement (Landis & Koch 1977)

<u>Kappa statistic</u>	<u>Strength of agreement</u>
< 0.00	Poor
0.00 – 0.20	Slight
0.21 – 0.40	Fair
0.41 – 0.60	Moderate
0.61 – 0.80	Substantial
0.80 – 1.00	Almost perfect

3.5.1 Question type

First, the results of the coding for *question types* are shown in Table 34 for Mandarin and Table 35 for French.

Table 34 Results of interrater reliability for *question type* in Mandarin

		Coder 1			
	Question type	Polar	Wh	Disjunctive	Excluded Total
Coder 2	Polar	337	11	6	3 357
	Wh	7	317	0	2 326
	Disjunctive	8	2	87	0 97
	Excluded	0	1	0	9 10
	Total	352	331	93	14 790

Out of 790 questions, 14 questions were removed from the calculation of interrater reliability according to the coding criteria, as they were interjections and questions in other languages. Thus, the results of the two coders have 96% of agreement on question types in **Mandarin** ($k = 0.93$). This kappa indicates an almost **perfect agreement**.

Table 35 Results of interrater reliability for *question type* in French

		Coder 1				
Coder 2	Question type	Polar	Wh	Disjunctive	Excluded	Total
	Polar	453	4	1	1	459
	Wh	17	353	0	0	370
	Disjunctive	5	0	11	0	16
	Excluded	0	0	0	5	5
Total		475	357	12	6	850

Among the 850 cases, six of them were excluded based on the coding criteria. Hence, they were not included in the calculation of interrater reliability. The results of the two coders reach 97% agreement on question types in **French** ($k=0.93$). This kappa indicates an almost **perfect agreement**, as was the case for Mandarin above.

3.5.2 Morphosyntactic structure

The coding for *morphosyntactic structures* was checked with interrater reliability and the procedure was the same as outlined above for *question type*. However, only the cases with agreement on the coding of *question type* are considered here. The results are shown and separated by languages and three question types. In addition, given that the focus is on questions with typical forms, fragment questions are categorised as ‘*others*’ when calculating interrater reliability for structures.

As for Mandarin, the results on morphosyntactic structures are presented in Table 36 for polar questions, in Table 37 for the *wh*-questions, and in Table 38 for disjunctive questions.

Table 36 Results of interrater reliability for *morphosyntactic structures* in Mandarin polar questions

		Coder 1			
Coder 2	Question structures	No inversion (declarative)	Particle	Others	Total
	No inversion (declarative)	115	2	6	123
	Particle	1	141	10	152
	Others	8	2	52	62
	Total	124	145	68	337

The coding of the morphosyntactic structures for **Mandarin polar questions** reaches 91% agreement ($k=0.86$). This kappa indicates an almost **perfect agreement**.

Table 37 Results of interrater reliability for *morphosyntactic structures* in Mandarin *wh*-questions

		Coder 1			
	Question structures	No particle	Particle	Others	Total
Coder 2	No particle	212	0	4	216
	Particle	3	25	3	31
	Others	11	1	58	70
	Total	226	26	65	317

The coding of the morphosyntactic structures for **Mandarin *wh*-questions** reaches 93% agreement ($k=0.84$). This kappa indicates an almost **perfect agreement**.

Table 38 Results of interrater reliability for *morphosyntactic structures* in Mandarin disjunctive questions

		Coder 1				
	Question structures	Or	A-not-A	VP-Neg	A-not-A + particle	Others Total
Coder 2	Or	7	0	0	0	1 8
	A-not-A	1	72	0	0	0 73
	VP-neg	0	0	2	0	0 2
	A-not-A + particle	0	0	0	2	0 2
	Others	0	1	0	0	1 2
	Total	8	73	2	2	2 87

The coding of the morphosyntactic structures for **Mandarin disjunctive questions** has 97% agreement ($k= 0.87$). This kappa indicates an almost **perfect agreement**.

As for French, the results are shown and separated by question types, in Table 39 for polar questions, in Table 40 for *wh*-questions, and in Table 41 for disjunctive questions.

Table 39 Results of interrater reliability for the *morphosyntactic structures* in French polar questions

		Coder 1				
	Question structures	Inversion	No inversion (declarative)	Particle <i>est-ce que</i>	Others	Total
Coder 2	Inversion	75	1	0	0	76
	No inversion (declarative)	1	211	0	5	217
	Particle <i>est-ce que</i>	1	0	7	0	8
	Others	0	1	0	151	152
	Total	77	213	7	156	453

The coding of the morphosyntactic structure for **French polar questions** has 98% agreement ($k=0.96$). This kappa indicates an almost **perfect agreement**.

Table 40 Results of interrater reliability for the *morphosyntactic structures* in French *wh*-questions

		Coder 1					
	Question structures	Inversion	No inversion	Particle <i>est-ce que</i>	<i>Wh</i> -in-situ	Others	Total
Coder 2	Inversion	146	0	0	0	0	146
	No inversion	4	44	0	2	0	50
	Particle <i>est-ce que</i>	1	0	44	0	0	45
	<i>Wh</i> -in-situ	0	1	0	19	0	20
	Others	0	2	0	0	90	92
	Total	151	47	44	21	90	353

The coding of the morpho-syntactic structure for **French *wh*-questions** has 97% agreement ($k=0.96$). This kappa indicates an almost **perfect agreement**.

Table 41 Results of interrater reliability for the *morphosyntactic structures* in French disjunctive questions

		Coder 1			
Question structures		Inversion + <i>ou</i>	No inversion + <i>ou</i>	Others	Total
Coder 2	Inversion + <i>ou</i>	6	0	0	6
	No inversion + <i>ou</i>	0	2	0	2
	Others	0	0	3	3
	Total	6	2	3	11

The coding of the morphosyntactic structure for **French disjunctive questions** reached agreement between the two coders. However, the number of disjunctive questions is relatively lower in this reliability test compared to the other two question types to the extent that the calculation of Cohen Kappa value may not be reliable. Therefore, I only present the results here with raw frequencies for the morphosyntactic structures of disjunctive questions.

3.5.3 Register

In addition to question types and morphosyntactic structures, questions are also coded for *registers* (i.e., narrative and dialogue) and interrater reliability was checked. Similarly, the cases that were excluded from the dataset according to the coding criteria (14 in Mandarin; 6 in French) were not considered in the calculation of reliability. The results are shown in Table 42 for Mandarin and in Table 43 for French.

Table 42 Results of interrater reliability for *register* in Mandarin

		Coder 1			
	Register	Dialogue	Narrative	Other	Total
Coder 2	Dialogue	527	12	2	541
	Narration	32	198	4	234
	Other	0	0	0	0
	Total	559	210	6	775

Note that one case was initially excluded by coder 2, thus only 775 cases were considered in the calculation of reliability. Hence, the coding for register in **Mandarin** questions has 94% agreement ($k=0.84$). This kappa indicates an almost **perfect agreement**.

Table 43 Results of interrater reliability for *register* in French

		Coder 1			
	Register	Dialogue	Narration	Other	Total
Coder 2	Dialogue	593	18	0	611
	Narration	5	212	2	219
	Other	0	11	3	14
	Total	598	241	5	844

The coding of the register for **French** questions has 96% agreement ($k=0.89$). This kappa indicates an almost **perfect agreement**.

3.5.4 Overall results of interrater reliability

Overall, the coding scheme for **the three dimensions reaches a high level of agreement** between the two coders, as shown in Table 44 for Mandarin and Table 45 for French.

Table 44 Overall results of interrater reliability in Mandarin

Dimension of coding	Agreement	Kappa score
Question type	96%	0.93
Morphosyntactic structure		
Polar question	91%	0.86
<i>Wh</i> -question	93%	0.84
Disjunctive question	97%	0.87
Register	94%	0.84

Table 45 Overall results of interrater reliability in French

Dimension of coding	Agreement	Kappa score
Question type	97%	0.93
Morphosyntactic structure		
Polar question	98%	0.96
<i>Wh</i> -question	97%	0.96
Disjunctive question	100%	NA
Register	96%	0.89

As a result, the coding scheme in Mandarin and French achieves perfect agreement, indicating that **the coding scheme is robust and the results of coder 1** (i.e., the author of this doctoral research) **are confirmed to be reliable**. Regarding disagreements between the two coders, the final decision was made after a discussion with the other raters in Mandarin and in French, respectively. If necessary, a third person (a Taiwan Mandarin native PhD student and a Belgium-French native professor in linguistics) was asked for confirmation. Thus, after the discussion (and, when necessary, the consultation with a third person), the remaining cases without full agreement are retained in the dataset

for the following comparable and parallel analyses. In the following sub-section of this chapter, I will briefly present the statistical tests that used in the analyses.

3.6 Data analysis: statistical tests

The statistical analyses were carried out in the open-source statistical software R (R Core Team 2024), including both descriptive and inferential statistical analyses. More specifically, two generalised linear models (GLMs) were applied for inferential statistical analyses of both comparable and parallel data. One was **Poisson regression**, implemented via the *glmer* function (Bates et al. 2015). It is used to analyse count data (e.g., frequency of question marks) while accounting for random variability across sources (e.g., authors and works), as this model is appropriate for modelling count-based outcomes (Winter 2020: 218). The other inferential model was a **Bayesian multinomial mixed-effects model** using the *brm* function (Bürkner 2021). This model is particularly suited for analyses involving more than two categorical outcome categories (e.g., question types). In the current study, the only dependent variable was the number of occurrences of the topic under investigation. Fixed effects were included, such as languages, literary genres (fantasy vs. detective), translation status (i.e., source vs. target), while the model also accounted for random intercepts for grouping factors, such as authors, translators, and works. The R-code for these models is provided in the Appendix. In addition to the two main generalised linear models, chi-square tests of independence were also included in the detailed discussion of each topic (i.e., taxonomy of questions, *wh*-in-situ, and particles). These tests were used in Chapter 4 for comparable results across languages and in Chapter 5 to examine both parallel patterns between translation pairs and monolingual comparison between non-translated and translated languages. This **small-scale inferential statistical analysis** was used to examine whether the observed frequencies differ significantly from expected frequencies. It served as a tool to assess associations between categorical variables more closely.

3.7 Summary of Chapter 3

This chapter provided an overview of the role of corpora in cross-linguistic studies of questions. The overview highlighted the need to broaden the scope of corpus-based research on questions, as most studies are restricted to certain types of question structures, such as declarative questions and tag questions. Only recently has Axelsson (2020)

conducted a more comprehensive corpus-based contrastive study by examining the entire question systems of English and Swedish. Pursuing a similar objective, the present study aims to provide a thorough investigation of the question systems of Taiwan Mandarin and French by means of a bidirectional parallel corpus. The design of this corpus was presented in Section 3.2, together with the rationale for novel selection in the two languages. The corpus material was uploaded on Sketch Engine, which was subsequently used to automatically extract all sentences containing question marks. These instances were then coded according to the annotation scheme introduced in Section 3.4, with respect to question type, morphosyntactic structure, and register in fiction (narrative vs. dialogue). The coding scheme was validated through interrater reliability checks. The results reported in Section 3.5 demonstrated the robustness of the coding scheme developed in the study. Finally, the last section of the chapter outlined the statistical analyses employed, ranging from small-scale tests (e.g., chi-square tests of independence) to regression modelling (i.e., Poisson and Bayesian regression models).

This chapter has provided the methodological foundation upon which the study is based. On the basis of the bidirectional parallel corpus compiled for the study, the analysis proceeds from three complementary perspectives: bilingual comparable, bidirectional parallel, and monolingual comparable. The next chapter begins with a contrastive comparison of the two languages under scrutiny (i.e., Taiwan Mandarin vs. French), adopting a bilingual comparable approach.

Chapter 4 Comparable analysis of questions in Taiwan Mandarin and French

This chapter first provides a broad overview of distribution of questions, followed by a detailed contrastive analysis of questions in Mandarin and French, based on bilingual comparable data from fiction. All observations therefore refer to the use of these sentences in novels, rather than to general linguistic tendencies. It is divided into six subsections. Section 4.1 examines the frequency of question marks across authors, novels, and genres. Section 4.2 presents the distribution of questions in terms of their question types, morphosyntactic structures, and registers. Sections 4.3 to 4.5 discuss three topics in greater detail based on the overall results of the comparable analysis: the taxonomy of questions and question particles, while the *wh*-in-situ phenomenon is analysed only for French, as it constitutes the default structure in Mandarin. Section 4.6 summarises the key findings of each topic and presents the hypotheses and expected findings for the parallel analyses presented in Chapter 5.

4.1 Overall frequency of questions in comparable corpus data

The bidirectional parallel corpus used in this study can be used as a comparable corpus that contains six original novels in Taiwan Mandarin and six others in French. All instances containing question marks were extracted with the advanced Concordance function in Sketch Engine (Kilgarrieff et al. 2014), searching for the character “?”. The results are shown in Table 46.

Table 46 Question marks extracted from the bidirectional parallel corpus

	Taiwan Mandarin (original)	French (original)
Raw frequencies	3,921	4,224
Relative frequencies per 100,000 tokens	762	653
Corpus size (tokens)	514,540	647,180

A chi-square test of independence revealed **a significant difference in the frequency of question marks in Taiwan Mandarin and French** ($\chi^2(1) = 11.272, p < .001$), indicating that **questions occur more frequently in Mandarin than in French novels**. The higher frequency in Taiwan Mandarin may reflect language-specific tendencies to use questions to perform certain indirect speech acts (e.g., requests, suggestions, or permissions). Alternatively, it may be related to non-linguistic factors, such as size of each novel, the

author's style, or the distribution of genres (detective vs. fantasy) in the corpus. While the present study cannot directly address the first possibility, it examines whether genre-related variation plays a role in shaping the observed distribution of question marks. Table 47 presents the frequency of question marks together with metadata for each novel in Taiwan Mandarin and French, including information about the author, work, genre, and text size.

Table 47 Metadata and frequency of question marks for the novels in Taiwan Mandarin and French (raw and relative frequencies)

Novels in Taiwan Mandarin				
Author	Work	Genre	Text size (number of tokens)	Raw frequency of question mark (per 10,000 tokens)
張國立 Chang Kuo-li	炒飯狙擊手 <i>Chǎo Fàn Jū Jí Shǒu</i>	Detective	70,395	868 (123)
紀蔚然 Chi Wei-jan	私家偵探 <i>Sī Jiā Zhēn Tàn</i>	Detective	113,488	1,117 (98)
陳思宏 Kevin Chen	鬼地方 <i>Guǐ Dì Fāng</i>	Fantasy	99,742	685 (69)
高翊峰 Kao Yi-feng	泡沫戰爭 <i>Pào Mò Zhàn Zhēng</i>	Fantasy	73,973	519 (70)
吳明益 Wu Ming-yi	複眼人 <i>Fù Yǎn Rén</i>	Fantasy	98,816	483 (49)
吳明益 Wu Ming-yi	天橋上的魔術師 <i>Tiān Qiáo Shàng De Mó Shù Shī</i>	Fantasy	56,281	249 (44)
Novels in French				
Franck Thilliez	<i>Gataca</i>	Detective	173,688	1,450 (84)
Michel Bussi	<i>Un avion sans elle</i>	Detective	153,598	1,204 (78)
Hervé Le Tellier	<i>L'Anomalie</i>	Fantasy	99,591	657 (66)
Marc Levy	<i>Le voleur d'ombres</i>	Fantasy	70,879	373 (53)
Pierre Lemaitre	<i>Trois jours et une vie</i>	Detective	67,852	304 (45)
Muriel Barbery	<i>La vie des elfes</i>	Fantasy	81,271	236 (29)

Building on these descriptive results, a Poisson regression was applied using the *glmer* function (Bates et al. 2015) to analyse the number of question marks (i.e., dependent variable) across genres, works, and authors in each language. The fixed effect is genre (i.e., detective vs. fantasy). Given that one author essentially corresponds to one work, only “work” is treated as a random effect in the statistical analysis. In addition, the text size (number of tokens) of each novel was included as an offset to adjust for differences in frequency of question marks. The model outputs are summarised in Table 48 and Table 49 for Taiwan Mandarin and French, respectively.

Table 48 Poisson regression output for frequency of question marks in Taiwan Mandarin

Predictor	Estimate	Standard Error	z value	p value
(Intercept)	-4.5094	0.1226	-36.773	< .0001***
Genre Fantasy	-0.6575	0.1513	-4.346	< .0001***

Formula= `glmer (Count ~ Genre+ (1 | Work), family = poisson, offset = log(Total), control = glmerControl(optimizer = "bobyqa"), data = .)`

Table 49 Poisson regression output for frequency of question marks in French

Predictor	Estimate	Standard Error	z value	p value
(Intercept)	-5.0115	0.1801	-27.828	<.0001 ***
Genre Fantasy	-0.3560	0.2555	-1.394	0.163

Formula= `glmer (Count ~ Genre+ (1 | Work), family = poisson, offset = log(Total), control = glmerControl(optimizer = "bobyqa"), data = .)`

Based on the statistical results, **the occurrence of question marks differs significantly across genres in Taiwan Mandarin** ($z = -4.346$, $p < .0001$), **but not in French** ($z = -1.394$, $p = 0.163$). In other words, when text size is taken into account, **the frequency of question marks is significantly lower in fantasy novels than in detective novels in Taiwan Mandarin**. Regarding the effect of individual works, **the estimated variance of the random effect is small** (0.029 in Taiwan Mandarin and 0.096 in French), indicating that **differences between works or authors have little influence on the frequency of question marks**. Thus, the occurrence of question marks appears to be unaffected by the selection specific novels or authors, although the random effect variance is somewhat larger in French than in Taiwan Mandarin.

Since the relative occurrence of question marks is higher in Taiwan Mandarin than in French, and given that genre has been shown to influence the distribution of question marks in Taiwan Mandarin, I further examine cross-linguistic differences by analysing the occurrences of question marks within each genre (i.e., detective vs. fantasy) separately. Specifically, this analysis aims to determine whether the higher frequency observed in Taiwan Mandarin at the initial stage persists when genre is controlled. The results for **detective novels** are shown first in Table 50.

Table 50 Metadata and frequency of question marks for the detective novels in Taiwan Mandarin and French (raw and relative frequencies)

Language	Author	Work	Text size (number of tokens)	Raw frequency of question mark (per 10,000 tokens)
Taiwan Mandarin	張國立	炒飯狙擊手	70,395	868
	Chang Kuo-li	<i>Chǎo Fàn Jū Jí Shǒu</i>		(123)
	紀蔚然	私家偵探	113,488	1,117
	Chi Wei-jan	<i>Sī Jiā Zhēn Tàn</i>		(98)
French	Franck Thilliez	<i>Gataca</i>	173,688	1,450
	Michel Bussi	<i>Un avion sans elle</i>	153,598	(84)
	Pierre Lemaitre	<i>Trois jours et une vie</i>		1,204
			67,852	(78)
				304
				(45)

A Negative binomial regression model was employed instead of a Poisson model in order to address the issue of overdispersion (Winter 2020: 227). This model shows an adequate fit (dispersion = 1.47), as dispersion is smaller than 1, confirming that the estimates are reliable, as shown in Table 51.

Table 51 Negative binomial regression output for frequency of question marks in detective novels in Taiwan Mandarin and French

Predictor	Estimate	Standard Error	z value	p value
(Intercept)	-4.502	0.186	-24.21	<.001
Language French	-0.472	0.24	-1.964	0.144

Formula= glm(QuestionMarks ~ Language, family = negative.binomial(theta = theta_hat_Detective), offset = offset, data = Q_Mark_Detective)

The results indicate although the frequency of question marks is a slightly lower in French detective novels than Taiwan Mandarin, **the difference is not statistically significant** ($\beta = -0.472$, $z = -1.964$, $p = .144$). In other words, the **overall pattern in which question marks are more frequent in Taiwan Mandarin novels than in French does not hold within the detective genre**. For the **fantasy novels**, the same statistical model was applied to examine the effect of languages. Table 52 presents the information about the fantasy novels in both languages.

Table 52 Metadata and frequency of question marks for the fantasy novels in Taiwan Mandarin and French (raw and relative frequencies)

Language	Author	Work	Text size (number of tokens)	Raw frequency of question mark (per 10,000 tokens)
Taiwan Mandarin	陳思宏 Kevin Chen	鬼地方 <i>Guǐ Dì Fāng</i>	99,742	685 (69)
	高翊峰 Kao Yi-feng	泡沫戰爭 <i>Pào Mò Zhàn Zhēng</i>	73,973	519 (70)
	吳明益 Wu Ming-yi	複眼人 <i>Fù Yǎn Rén</i>	98,816	483 (49)
	吳明益 Wu Ming-yi	天橋上的魔術師 <i>Tiān Qiáo Shàng De Mó Shù Shī</i>	56,281	249 (44)
French	Hervé Le Tellier	<i>L'Anomalie</i>	99,591	657 (66)
	Marc Levy	<i>Le voleur d'ombres</i>	70,879	373 (53)
	Muriel Barbery	<i>La vie des elfes</i>	81,271	236 (29)

The outcome of the Negative binomial regression model indicates a perfect model fit (dispersion = 1.277), suggesting that the parameter estimates are robust, as shown in Table 53.

Table 53 Negative binomial regression output for frequency of question marks in fantasy novels in Taiwan Mandarin and French

Predictor	Estimate	Standard Error	z value	p value
(Intercept)	-5.15	0.149	-34.47	<.001
Language [French]	-0.164	0.228	-0.719	0.504

Formula = glm(QuestionMarks ~ Language, family = negative.binomial(theta = theta_hat_Fantasy), offset = offset, data = Q_Mark_Fantasy)

The pattern is similar to that observed for the detective novels. **The frequency of question marks does not differ significantly between the two languages** ($\beta = -0.164$, $z = -0.719$, $p = 0.504$). Taken together, the results from the two Negative binomial models suggest that **there is no significant cross-linguistic difference in the frequency of question marks within either the detective or the fantasy genre.**

Overall, the frequency of questions marks appears higher in Taiwan Mandarin than in French, with relative frequencies of 762 and 653 per 100,000 tokens, respectively. This difference may initially suggest **a language-specific preference for using questions.** However, it may also be influenced by **non-linguistic factors**, such as the size of each novel, author's style, or the genre of the novel. To examine these possibilities, several regression models were applied to assess the effect of non-linguistic factors on the distribution of question marks.

First, two Poisson regression models were conducted separately for Taiwan Mandarin and French, with work and genre included as predictors. The results show that **the effect of work variety is minimal.** In **Taiwan Mandarin**, however, **a significant difference is found between the genres** ($z = -4.346$, $p < .0001$), while in **French** genre has **no significant effect.** Next, two negative binomial regression models were applied within each genre to test cross-linguistic differences in question mark frequency. The results indicate that **within both detective and fantasy novels, there is no significant difference between Taiwan Mandarin and French.** This suggests that the overall higher frequency of question marks observed in Taiwan Mandarin does not persist once genre is controlled and is therefore unlikely to reflect a language-specific preference. Rather, the initial cross-linguistic difference appears to be associated with differences in corpus composition, as well as language-internal, genre-related variation in Taiwan Mandarin, although the present analyses do not establish a direct causal link between

these factors and the observed cross-linguistic pattern. Given the importance of genre in affecting the distribution of questions, it should be taken into account when analysing patterns of language use in cross-linguistic corpus studies.

4.2 The distribution of questions in registers, question types, and morphosyntactic structures

Following an overview and analysis of the distribution of question marks from comparable data, this subsection presents the results of the coding for question types, morphosyntactic structures, and registers in Taiwan Mandarin and French novels. After closer examination of questions from these perspectives, contrastive insights are provided in Section 4.2.4. The study excluded non-qualifying cases, such as interjections and questions expressed in other languages (see Section 3.4 on data coding). Therefore, a total of 82 cases in Mandarin and 28 cases in French were removed from the subsequent analysis. The discussion begins with the results of the register coding, which offer a more holistic view of how questions are distributed between the two registers (i.e., dialogue and narrative). This analysis serves as the foundation for the examination of question types and morphosyntactic structures in the following sections.

4.2.1 Register

Table 54 and Figure 12 present the overall distribution of questions across registers, followed by the discussion of the difference between the **two main registers** (dialogue vs. narrative).

Table 54 Question mark distributions in the two registers

	Taiwan Mandarin		French	
	Raw Frequency	%	Raw Frequency	%
Dialogue	2,765	72	2,959	70.5
Narrative	1,037	27	1,201	28.6
Other	37	1	36	0.9
Total	3,839	100%	4,196	100%

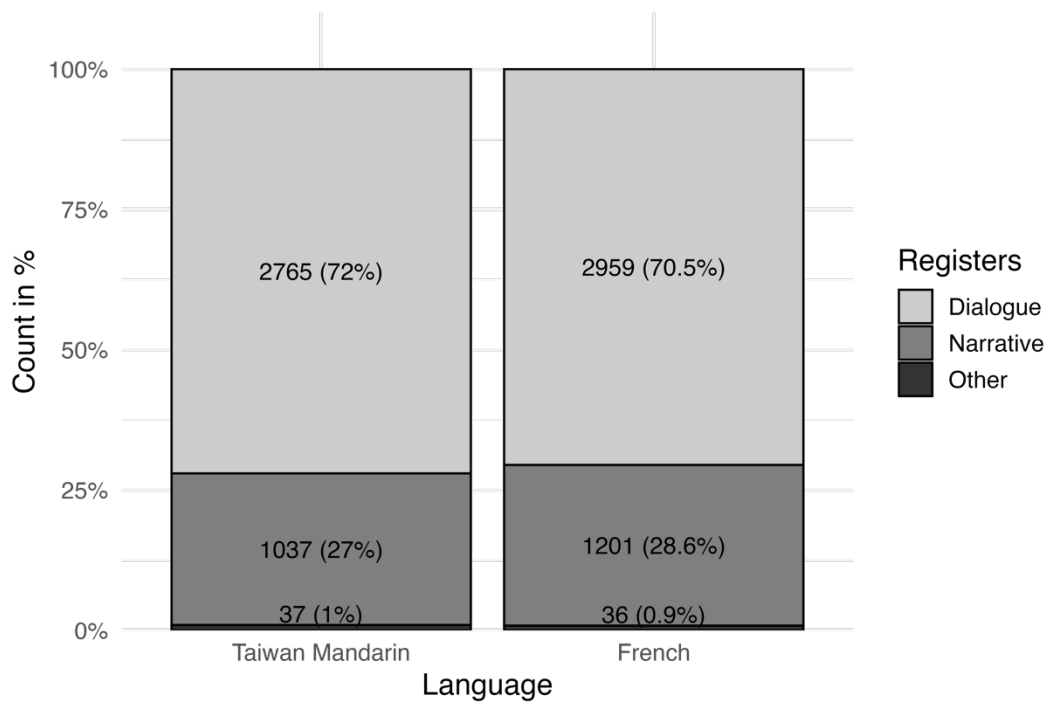


Figure 12 Question mark distributions in the two registers

The distribution of questions across registers is broadly similar in Taiwan Mandarin and French. **In both languages, questions in dialogue account for over 70% of the data, whereas fewer than 30% of the questions are found in narrative.** Within each language, the difference in frequency between registers is **statistically significant** in both Taiwan Mandarin ($\chi^2(2) = 2976, p < .0001$) and French ($\chi^2(2) = 3096.2, p < .0001$). The results of adjusted residual in Table 55 show that the question marks have significantly **higher occurrence in dialogue** than expected (adjusted residuals > 2), while the occurrence of question marks is significantly **lower in narrative** than expected (adjusted residuals < -2). No significant difference is found in the category *Other*.

Table 55 Adjusted residuals for question mark distributions across registers

	Taiwan Mandarin	French
Dialogue	13.8	14.1
Narrative	-10.2	-11
Other	-0.6	-0.7

Given the significant differences between the two registers, the results presented in the next two subsections (4.2.2 on question types and 4.2.3 on morphosyntactic structures) are also differentiated by register.

4.2.2 Question type

The questions extracted from the bidirectional parallel corpus were classified into *polar*, *wh*-, and *disjunctive* questions, following the traditional three-way distinction. Results are presented in Table 56 and Figure 13.

Table 56 Question types in Taiwan Mandarin and French

	Taiwan Mandarin		French	
	Raw Frequency	%	Raw Frequency	%
Polar questions	1,756	45.7	2,328	55.5
<i>Wh</i> -questions	1,617	42.1	1,793	42.7
Disjunctive questions	466	12.1	75	1.8
	3,839	100%	4,196	100%

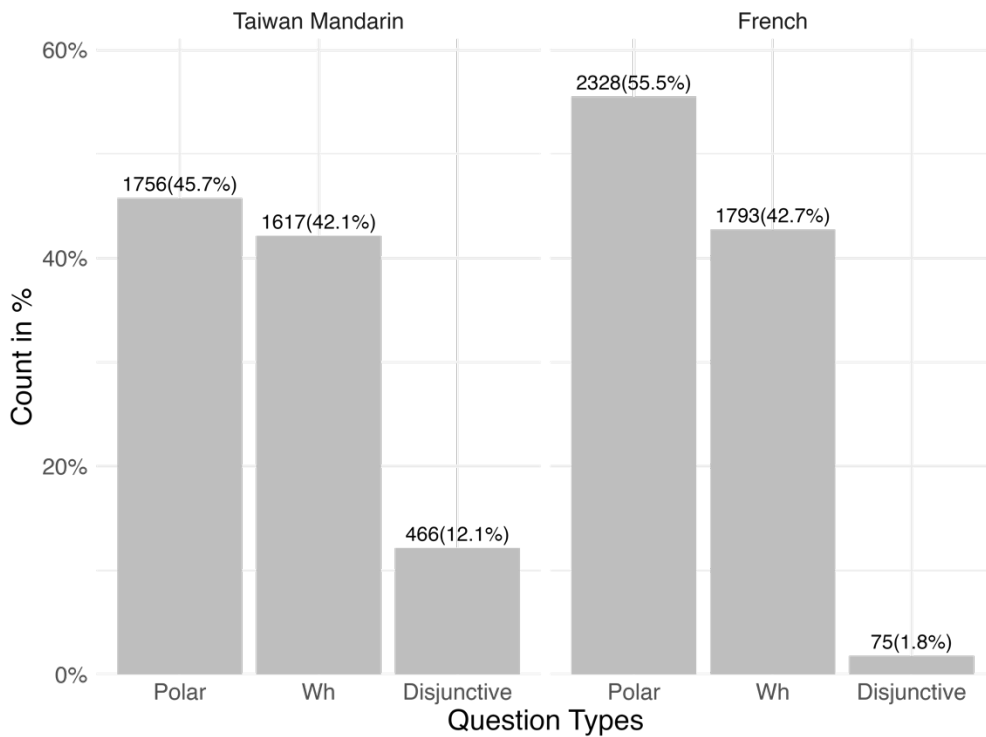


Figure 13 Question types in Taiwan Mandarin in French

Polar questions are the most frequent question type in both languages, followed by *wh*-questions and disjunctive questions. Notably, **a substantial difference between the two languages emerges in the distribution of polar and disjunctive questions**, with a discrepancy of nearly 10%, whereas the proportion of *wh*-questions remains largely

similar. In addition to the overall distribution of question types, I also examine their distribution across different registers, as presented in Table 57 and Figure 14.

Table 57 Question types in Taiwan Mandarin and French by registers
(raw and relative frequencies)

Taiwan Mandarin				
	Dialogue	Narrative	Other	Total
Polar questions	1,341 (76%)	397 (23%)	18 (1%)	1,756 (100%)
<i>Wh</i> -questions	1,100 (68%)	501 (31%)	16 (1%)	1,617 (100%)
Disjunctive questions	324 (70%)	139 (30%)	3 (0.64%)	466 (100%)
	2,765	1,037	37	3,839
French				
Polar questions	1,815 (78%)	497 (21%)	16 (0.7%)	2,328 (100%)
<i>Wh</i> -questions	1,118 (62%)	657 (37%)	18 (1%)	1,793 (100%)
Disjunctive questions	26 (35%)	47 (63%)	2 (2.7%)	75 (100%)
	2,959	1,201	36	4,196

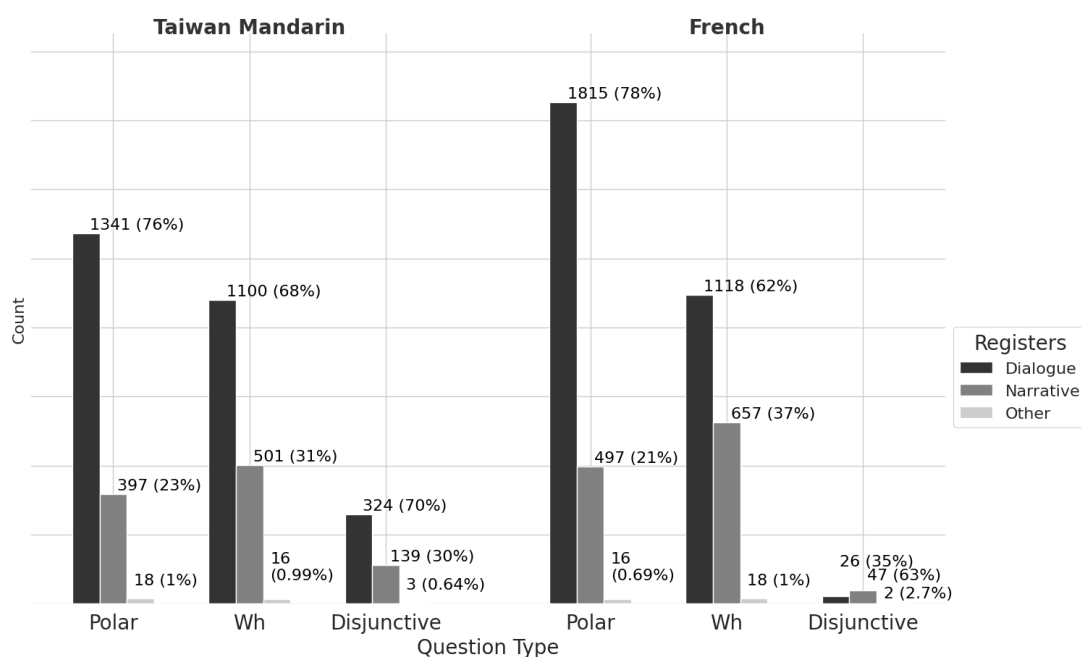


Figure 14 Question types by registers in Taiwan Mandarin in French by registers

Polar questions exhibit a similar distribution pattern in Taiwan Mandarin and French, with over 75% of occurrences in dialogue and approximately 22% in narrative. **Wh-questions also occupy a substantial proportion in dialogue in both languages.** However, their frequency is slightly higher in dialogue in Taiwan Mandarin (68%) compared to French (62%), whereas in narrative, *wh*-questions are more frequent in French (37%) than in Taiwan Mandarin (31%). By contrast, the distribution of **disjunctive questions reveals an opposite trend across the two languages.** While a majority of disjunctive questions in Taiwan Mandarin appear in dialogue (70%), in French they are more common in narrative (63%) than in dialogue (35%). Overall, despite minor differences in the distribution of *wh*-questions, **the register patterns are largely similar across question types, with most questions clustered in dialogue**, except for disjunctive questions in French.

To assess the influence of internal variable (i.e., works) and its interaction with registers and genres on question type distribution, a Bayesian multinomial mixed-effects model was fitted using the *brm* function (Bürkner 2021). The dependent variable (i.e., question type) was categorical with three levels, and the independent variables, used as fixed effects, were register and genre. Reference levels were set to ‘Disjunctive question’ (question type), ‘Dialogue’ (register), and ‘Detective’ (genre), with random intercepts for work, which serves as proxy for authorship, as each work approximately corresponds to a single author. The model outputs are shown in Table 58 and Table 59 for **Taiwan Mandarin**.

Table 58 Multinomial mixed-effects model output for question types in Taiwan Mandarin (Random effect: work)

Question type	Posterior mean of standard deviation	95% CI (lower, upper)
Polar	0.87	[0.37, 2.16]
<i>Wh</i>	0.7	[0.28, 1.6]

Formula= brm (formula = Type ~ Register + Genre + (1 | Work), data = data_MD1, family = categorical ())

Table 59 Multinomial mixed-effects model output for question types in Taiwan Mandarin (Fixed effects: register and genre)

Question type	Predictor	Posterior mean	95% CI (lower, upper)
Polar	Intercept	1.38	[-0.22, 2.66]
	Register: Narrative	-0.16	[-0.39, 0.09]
	Register: Other	0.61	[-0.58, 2.17]
	Genre: Fantasy	0.08	[-1.51, 1.98]
<i>Wh</i>	Intercept	1.33	[0.19, 2.44]
	Register: Narrative	0.35	[0.11, 0.58]
	Register: Other	0.74	[-0.48, 2.32]
	Genre: Fantasy	-0.27	[-1.72, 1.07]

Formula=brm(formula = Type ~ Register + Genre + (1 | Work), data = data_MD1, family = categorical ())

As shown in Table 58, **distribution of question types varies significantly across novels in Taiwan Mandarin** (Posterior mean of standard deviation = 0.87 for polar questions and 0.7 for *wh*-questions; 95% confidence intervals do not include zero). Accounting for work-level variation as a random effect, as shown in the formula under Table 59, both intercept-based comparisons (i.e., the reference levels) indicate that polar and *wh*-questions occur more frequently than disjunctive questions in dialogue within detective novels. However, only the difference between *wh*- and disjunctive questions is statistically significant. In narrative, polar questions are less frequent than disjunctive questions, whereas *wh*-questions occur more often. Notably, only the contrast between *wh*- and disjunctive questions is statistically significant, as the 95% CI does not include zero. Therefore, the **variation across novels exerts a stronger influence on the distribution of question types**. In addition, the **register *narrative* appears to have a significantly greater impact on the distribution of *wh*-questions**.

The results of the regression analysis for **French** are shown in Table 60 and Table 61.

Table 60 Multinomial mixed-effects model output for question types in French
(Random effect: work)

Question type	Posterior mean of standard deviation	95% CI (lower, upper)
Polar	0.34	[0.02, 1.2]
<i>Wh</i>	0.26	[0.01, 0.87]

Formula= brm (formula = Type ~ Register + Genre + (1 | Work), data = data_FR1, family = categorical())

Table 61 Multinomial mixed-effects model output for question types in French
(Fixed effects: register and genre)

Outcome Comparison	Predictor	Posterior mean	95% CI (lower, upper)
Polar	Intercept	4	[3.37, 4.64]
	Register: Narrative	-1.8	[-2.33, -1.28]
	Register: Other	-1.97	[-3.42, -0.09]
	Genre: Fantasy	1.15	[0.14, 2.27]
<i>Wh</i>	Intercept	3.53	[2.97, 4.1]
	Register: Narrative	-1	[-1.53, -0.49]
	Register: Other	-1.28	[-2.75, 0.57]
	Genre: Fantasy	1.22	[0.36, 2.22]

Formula= brm (formula = Type ~ Register + Genre + (1 | Work), data = data_FR1, family = categorical ())

Table 60 shows that the distribution of question types varies significantly across works, although the low estimated standard deviations values (0.34 for polar questions and 0.26 for *wh*-questions; 95% confidence intervals do not include zero) suggest only a limited impact on the use of question types. Compared to Taiwan Mandarin, **French novels exhibit even less variation in the distribution of question types**. After controlling for the variety of novels, the intercept results in Table 61 indicate that polar and *wh*-questions are significantly more frequent than disjunctive questions, particularly in dialogue within detective novels. Across registers, **polar and *wh*-questions occur less often in narrative and more frequently in dialogue than disjunctive questions**, as discussed above (see p. 201).

The regression models for both Mandarin and French suggest that **variation across works influences the distribution of question types**. Even when controlling for this variation as a random effect, **significant differences emerge across registers and genres**. In Mandarin, *wh*-questions occur significantly more frequently than disjunctive questions in dialogue within detective novels. In French, polar and *wh*-questions both dominate dialogue in detective novels, and their distribution differs significantly from that of disjunctive questions.

Cross-linguistically, the variation across works appears to exert a stronger influence on the distribution of question types in Taiwan Mandarin than in French. The difference may be from the corpus composition, as two of the novels in Taiwan Mandarin were written by the same author, while all French novels were produced by different authors. Regarding fixed effects, register significantly affects the distribution of question types in both languages, but in distinct ways. In Mandarin, this effect is observed only for *wh*-questions, which becomes more frequent in narrative than in dialogue. In contrast, in French, register influences both polar and *wh*-questions, and the same context leads to a decrease in their frequency. This contrast suggests that the two languages employ *wh*-questions differently in narrative of fiction. As for genre (fantasy vs. detective), no significant effect is found in Taiwan Mandarin, whereas an effect appears in French. **This discrepancy may result from the unbalanced number of genres in the Mandarin corpus or from the low frequency of disjunctive questions in French,** which may artificially amplify the observed effect. These cross-linguistic patterns are further examined in Section 5.11.

4.2.3 Morphosyntactic structures of questions in Taiwan Mandarin and in French

In addition to question types, the extracted questions were coded for morphosyntactic structure. As discussed in Sections 2.2.1 and 2.2.2, Mandarin and French differ in how questions are marked using morphosyntactic structures. This subsection is divided into two parts. Section 4.2.3.1 reports the results for Taiwan Mandarin, and Section 4.2.3.2 for French. Finally, Section 4.2.3.3 provides a summary offering contrastive insights into the morphosyntactic structures of Taiwan Mandarin and French. Results are analysed by question types, register and the syntactic distinction between typical and elliptical questions (i.e., those with incomplete sentential forms).

4.2.3.1 Morphosyntactic structures of questions in Taiwan Mandarin

The main strategies for marking questions in **Mandarin**, as discussed in Section 2.2.1, include **particles** as in (4.1), **question words** as in (4.2), **disjunctive structures** as in (4.3), and tags as in (4.4) (Li & Thompson 1981: 520–521).

(4.1) 你 好 嗎?
nǐ hǎo ma?
2SG good PTCL
‘How are you?’ (Li & Thompson 1981: 521)

(4.2) 誰 去?
shéi qù?
who go
‘Who’ll go?’ (Li & Thompson 1981: 520)

(4.3) 你 去 還是 他 來?
nǐ qù hái shì tā lái?
2SG go or 3SG come
‘Will you go or s/he will come?’ (Li & Thompson 1981: 521)

(4.4) 我們 去 吃 水果， 好-不-好?
women qù chī shuǐguǒ hǎo-bù-hǎo?
1PL go eat fruit good-NEG-good
‘Let’s go eat some fruit, OK?’ (Li & Thompson 1981: 546)

In addition to morphosyntactic strategies, intonation is also employed to mark polar questions. In the texts, this prosodic cue is manifested through declarative word order, without any of the conventional strategies mentioned earlier.

As shown in Table 62 and Figure 15, 81.4% of **polar questions** are in typical question form ($n = 1,429$), while 18.3% are elliptical ($n = 327$).

Table 62 Morphosyntactic structures of polar questions in Taiwan Mandarin
(raw frequencies)

	Typical question form			
	Dialogue	Narrative	Other	Total
Particle	641	180	7	828
No inversion (declarative)	451	143	7	601
Total of typical form				1429
	Elliptical form			
Noun	97	37	4	138
Tag	76	23	0	99
Verb	37	9	0	46
Adjective	15	1	0	16
Adverb	14	1	0	15
Preposition	3	3	0	6
Insubordination	4	0	0	4
Pronoun	3	0	0	3
Total of elliptical form				327
Total				1,756

Polar questions in Taiwan Mandarin

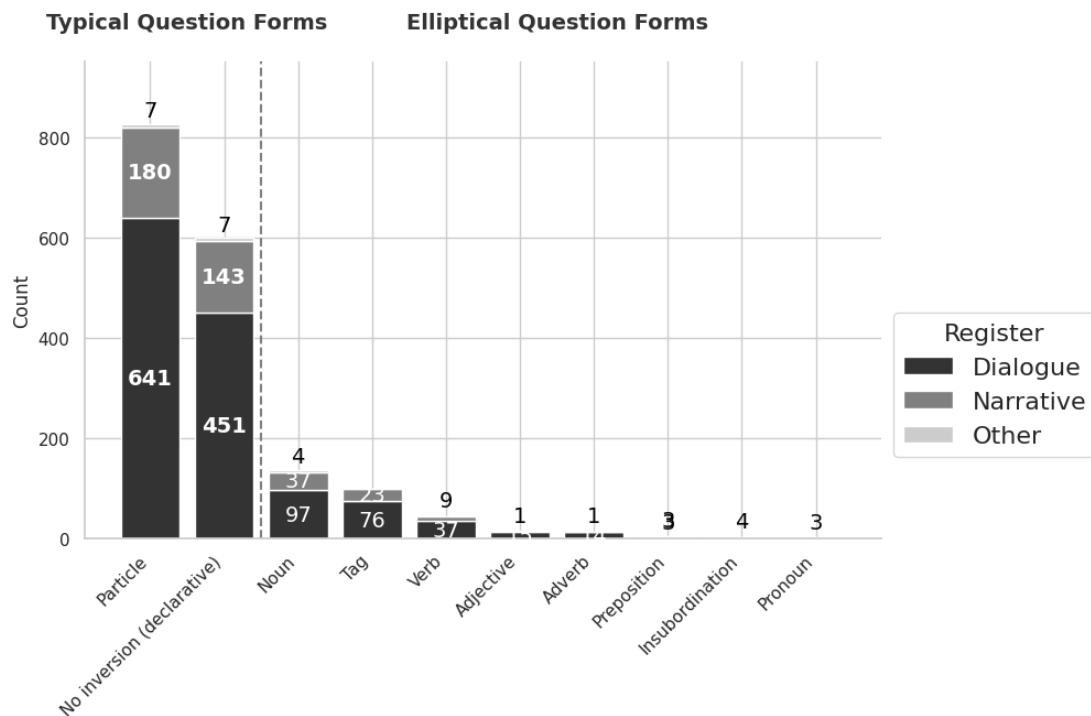


Figure 15 Morphosyntactic structures of polar questions in Taiwan Mandarin

Polar questions in Mandarin are usually marked by a particle (e.g., 嗎 *ma*), accounting for 47% of all polar questions, followed by 34% that use declarative word order. Elliptical forms each account for less than 10%. These fragment polar questions include: a noun as in (4.5), a tag as in (4.6), and a verb as in (4.7).

(4.5) 「可以，老規矩？」

kěyǐ lǎoguījǔ?

yes old.rule

‘Yes, as usual?’

(炒飯狙擊手 *Chǎo Fàn Jū Jí Shǒu*)

(4.6) 「這 一 兩 年 不 太 順 利 ， 是 吧 ？」

zhè yī liǎng nián bú tài shùnlì shì ba?

DEM one two year NEG so smooth COP PTCL

‘It’s been a rough couple of years, hasn’t it?’

(私家偵探 *Sī Jiā Zhēn Tàn*)

(4.7) 記得 嗎 ？

jìdé ma?

remember PTCL

‘Remember?’

(鬼地方 *Guǐ Dì Fāng*)

The particles used in elliptical forms, such as 吧 *ba* in (4.6) and 嗎 *ma* in (4.7), are also coded as such in the dataset. However, for reasons of methodological symmetry in contrastive analysis with French, the present study provides a more detailed analysis only of particles occurring in complete questions. This choice is motivated by the fact that given that the main particle in French (i.e., *est-ce que*) appears exclusively in complete questions. Particles occurring in elliptical forms are therefore beyond the scope of the present discussion and are reserved for future research.

As shown in Table 63 and Figure 16, 80% of *wh*-questions in Mandarin appear in typical forms ($n = 1,299$), and nearly 20% are elliptical ($n = 318$).

Table 63 Morphosyntactic structures of *wh*-questions in Taiwan Mandarin
(raw frequencies)

	Typical question form			
	Dialogue	Narrative	Other	Total
No particle	782	387	15	1,184
Particle	49	66	0	115
Total of typical form				1,299
	Elliptical form			
<i>Wh</i> -word	161	19	1	181
Noun	69	22	0	91
Adverb	12	0	0	12
Pronoun	10	1	0	11
Insubordination	5	5	0	10
Conjunction	5	1	0	6
Verb	6	0	0	6
Preposition	1	0	0	1
Total of elliptical form				318
Total				1,617

Wh-questions in Taiwan Mandarin

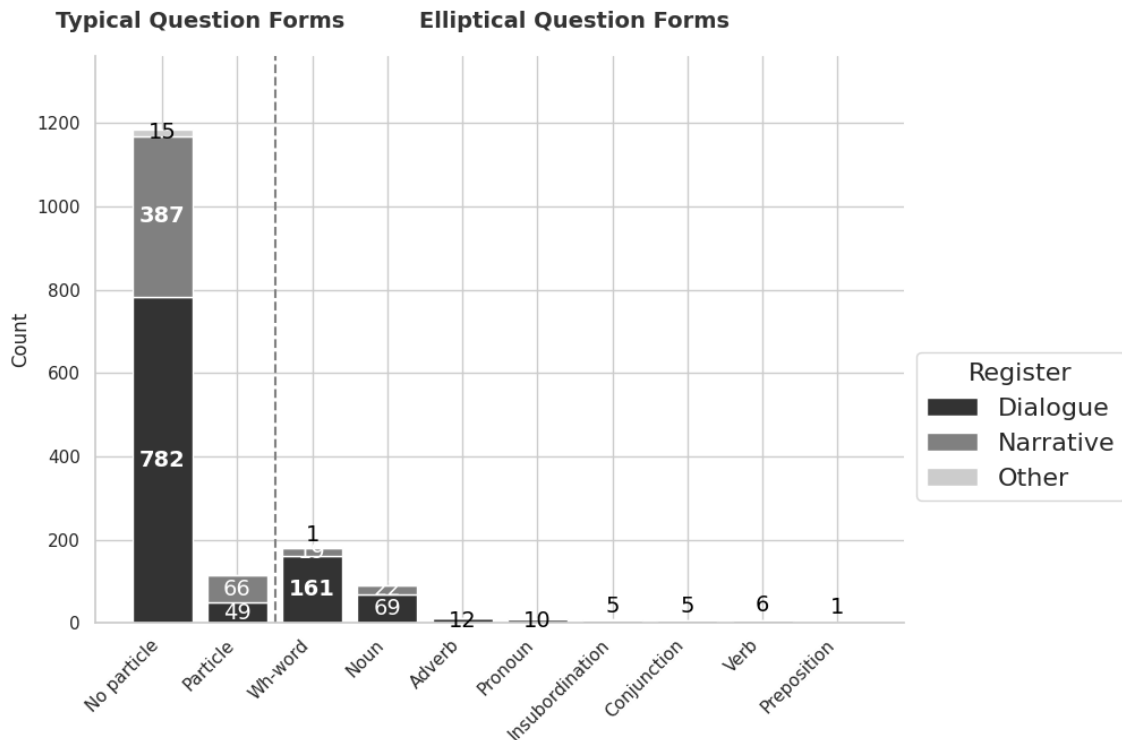


Figure 16 Morphosyntactic structures of *wh*-questions in Taiwan Mandarin

Unlike polar questions, ***wh*-questions in Mandarin are rarely marked by particles** (e.g., 呢 *ne*): only 7% are accompanied by a particle ($n = 115$), while 73% of questions are without particles ($n = 1,184$). Furthermore, particles in *wh*-questions occur more frequently in narrative. Among the elliptical forms, the presence of a single *wh*-word accounts for over 10% of the data, as illustrated in (4.8).

- (4.8) 四 個 出口， 哪 一 個？
sì ge chūkǒu nǎ yī ge?
 four CL exit which one CL
 ‘Which of the four exits?’ (炒飯狙擊手 *Chǎo Fàn Jū Jí Shǒu*)

Finally, **disjunctive questions** are relatively infrequent in Mandarin compared to the polar and *wh*-questions. Nonetheless, the distribution between typical and elliptical forms mirrors that of the other types. As shown in Table 64 and Figure 17, over 95% of disjunctive questions take typical forms ($n = 445$), while fewer than 5% are elliptical ($n = 21$).

Table 64 Morphosyntactic structures of disjunctive questions in Taiwan Mandarin
(raw frequencies)

	Typical question form			
	Dialogue	Narrative	Other	Total
A-not-A	253	90	2	345
還是 <i>háishì</i> 'or'	28	26	0	54
VP-Neg	27	4	1	32
Particle + A-not-A	3	8	0	11
Particle + 還是 <i>háishì</i> 'or'	0	3	0	3
Total of typical form				445
	Elliptical form			
Noun	6	5	0	11
Adjective	3	2	0	5
Verb	3	1	0	4
Adverb	1	0	0	1
Total of elliptical form				21
Total				466

Disjunctive questions in Taiwan Mandarin

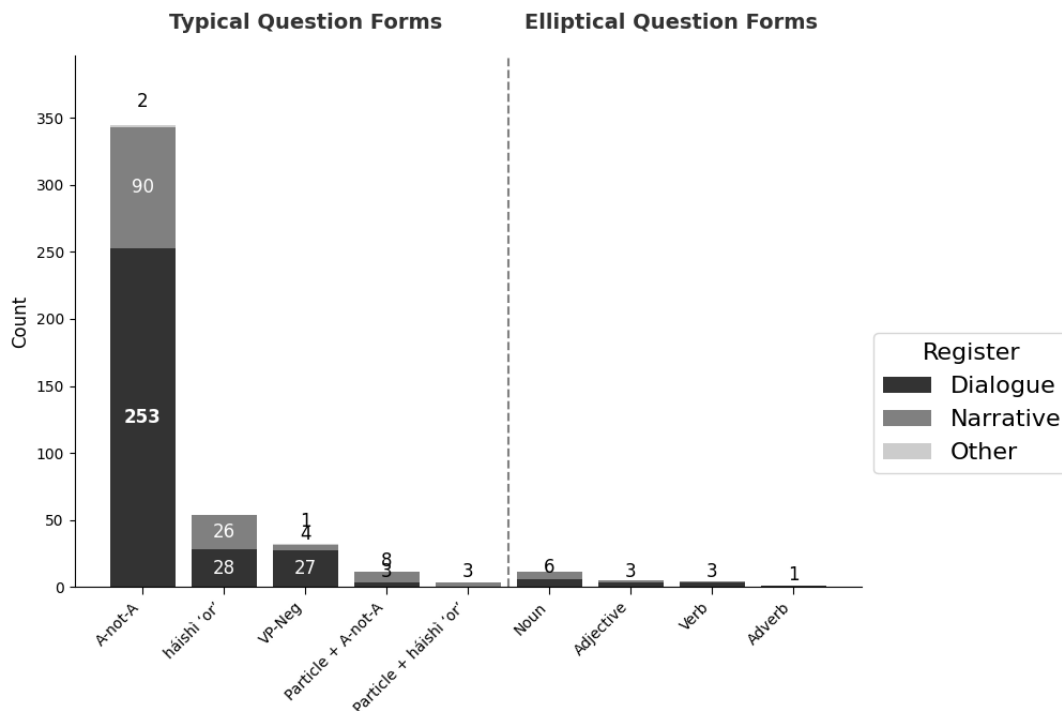


Figure 17 Morphosyntactic structures of disjunctive questions in Taiwan Mandarin

The **A-not-A structure**, exemplified by 餓不餓 *è bú è* ‘hungry or not’ in (4.9), **dominates the category**, appearing in 74% of cases ($n = 345$).

- (4.9) 肚子 餓 不 餓?
dùzi è bú è?
 stomach hungry NEG hungry
 ‘Are you hungry or not?’ (炒飯狙擊手 *Chǎo Fàn Jū Jí Shǒu*)

In contrast, the alternative structure using a connective 還是 *háishì* ‘or’ accounts for only 12% ($n = 54$). The total use of particle is minimal (3%, $n = 14$) and more commonly appears in the narrative, as is also the case for particles in *wh*-questions. The most frequent elliptical structure involves a noun phrase, as illustrated in (4.10).

- (4.10) 「長的 那 篇 還是 短的？」
chángde nà piān háishì duǎnde?
 long DEM CL or short
 ‘The long one or the short one?’ (複眼人 *Fù Yǎn Rén*)

In summary, **most questions in Mandarin follow typical morphosyntactic structures**, such as declarative word order, the use of particle, and the A-not-A structure,

while elliptical forms are relatively rare. Two notable observations emerge. First, **particles are more frequently used in polar questions**, as this latter often referred to as “particle questions” or “*ma* questions” (Li & Thompson 1981: 547, 1984: 54; Cheng 1984: 89). However, the use of particles is less favoured in *wh*- and disjunctive questions, where they are often omitted due to their optional nature (Pan & Paul 2016; Hsu 2020). Second, **both typical and elliptical forms are more prevalent in dialogue**, although particles in *wh*- and disjunctive questions tend to occur more often in narrative. These two observations will be discussed in detail in Section 4.5, which presents a comparable analysis of particles. The next section examines the morphosyntactic structures of questions in French and subsequently introduces a contrastive discussion of these morphosyntactic structures in Taiwan Mandarin and French.

4.2.3.2 Morphosyntactic structures of questions in French

Following the analysis of question structures in Mandarin, this subsection turns to the morphosyntactic structures of questions in French. As outlined in Section 2.2.2, the three main strategies to mark questions in French are **intonation**, **inversion**, and **particle** use (Harris 1998: 237; Guryev 2017: 24; Druetta 2018: 23). While intonation serves as a marker for polar questions in spoken language, it typically corresponds to declarative word order in written texts, such as novels.

As shown in Table 65 and Figure 18, 67% of French **polar questions** appear in typical forms ($n=1,562$), while 33% are elliptical ($n=766$).

Table 65 Morphosyntactic structures of polar questions in French (raw frequencies)

	Typical question form			
	Dialogue	Narrative	Other	Total
No inversion (declarative)	981	36	2	1,019
Inversion	228	255	0	483
Particle <i>est-ce que</i>	43	16	1	60
Total of typical form				1,562
	Elliptical form			
Noun	274	105	7	386
Tag	141	10	1	152
Verb	32	27	3	62
Preposition	30	20	2	52
Insubordination	32	13	0	45
Adverb	26	7	0	33
Adjective	18	5	0	23
Pronoun	9	3	0	12
Conjunction	1	0	0	1
Total of elliptical form				766
Total				2,328

Polar questions in French

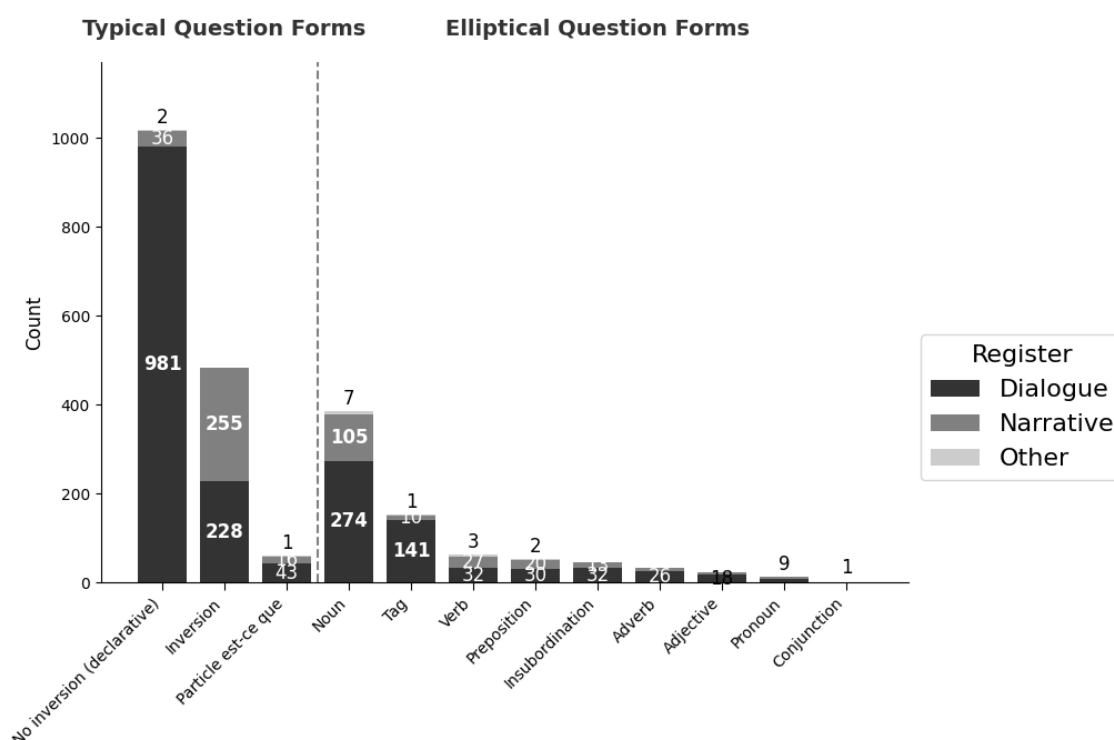


Figure 18 Morphosyntactic structures of polar questions in French

The most frequent structure is declarative word order, followed by inversion and the *est-ce que* particle. The disparity in frequency across these structures is notable. In terms of register, both declarative word order and particle occur primarily in dialogue. Conversely, inversion is more frequent in narrative. Within elliptical forms, noun phrases constitute nearly 17% of cases, as shown in the second question of A in (4.11).

(4.11) A: *Et vous, vous êtes qui ?*
 and 2SG 2SG COP who
 ‘And, who are you?’

B: *Marc Vitral.*
 ‘Marc Vitral.’

A: *Son mari ?*
 POSS.3SG husband
 ‘Her husband?’

(*Un avion sans elle*)

Table 66 and Figure 19 shows that 68% of *wh*-questions in French use typical question forms ($n = 1,222$), while 32% are elliptical ($n = 571$).

Table 66 Morphosyntactic structures of *wh*-questions in French (raw frequencies)

	Typical question form			
	Dialogue	Narrative	Other	Total
Inversion (<i>wh</i> -adverbial)	212	187	6	405
Inversion (<i>wh</i> -object)	163	138	5	306
Particle <i>est-ce que</i>	144	32	0	176
In-situ	106	8	0	114
No inversion (<i>wh</i> -subject)	51	56	1	108
No inversion (<i>wh</i> -adverbial)	79	7	2	88
No inversion (<i>wh</i> -object)	23	2	0	25
Total of typical form				1,222
	Elliptical form			
<i>Wh</i> -word	166	90	2	258
<i>Wh</i> -infinitive	35	73	0	108
Noun	55	16	0	71
Insubordination	15	33	1	49
Adverb	27	6	0	33
Preposition	13	5	1	19
Particle <i>ça</i>	17	0	0	17
Verb	2	4	0	6
Conjunction	4	0	0	4
Pronoun	4	0	0	4
Adjective	1	0	0	1
Particle <i>donc</i>	1	0	0	1
Total of elliptical form				571
Total				1,793

Wh-questions in French

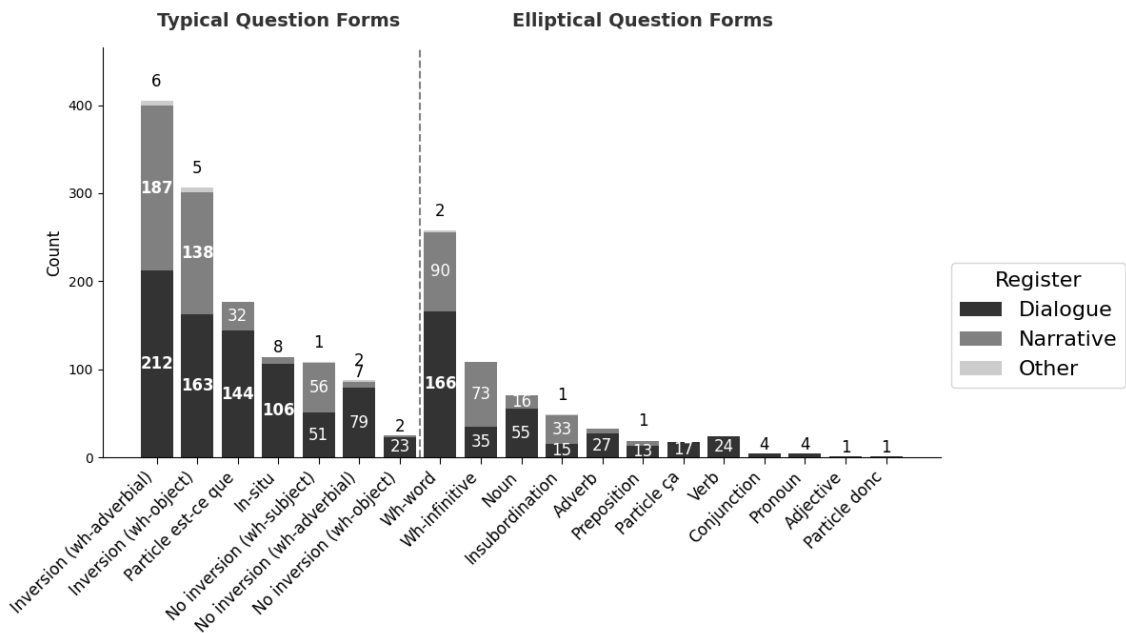


Figure 19 Morphosyntactic structures of *wh*-questions in French

Inversion is the most frequent structure (40%, $n = 711$), followed by *No inversion* (12%, $n = 221$), the *est-ce que* particle (10%, $n = 176$), and *wh*-in-situ (6.3%, $n = 114$). Among the elliptical forms, *wh*-word (14%, $n = 258$), as in (4.12), and *wh*-infinitive (6%, $n = 108$), as in (4.13) are most common. However, the former tends to appear in dialogue, while the latter is more frequent in narrative.

(4.12) *Des scientifiques ? Pourquoi ?*
 ART scientists why
 ‘The scientists? Why?’ (Gataca)

(4.13) *Et maintenant, quoi faire ?*
 and now what do
 ‘And now, what to do?’ (Trois jours et une vie)

Wh-infinitives, whose subjects are omitted and are typically the speaker or a general referent (e.g., *on* ‘we’), are often used to highlight an issue with an object or circumstance (Riegel, Pellat & Rioul 2009: 678). These questions do not necessary seek a response from the interlocutor; rather, they require an answer through mindful reflection, as deliberative questions (Abeillé & Godard 2021: 1323). Another elliptical form, insubordination, is mainly found in narrative contexts. It involves a subordinated conditional clause with an omitted main clause that implicitly raises a question ‘*What will*

happen?’ (Lombardi Vallauri 2016: 147; Fang & Hengeveld 2022: 890). This type of question is often seen in written language and often hypothetical in nature (Coveney 2011: 123), as shown in (4.14).

- (4.14) *Et si elle en trouvait une ?*
 and if 3SG it find.IMPART one
 ‘What if she found one?’ (Gataca)

Although **disjunctive questions** are relatively rare, they also tend to adopt typical forms, as shown in Table 67 and Figure 20.

Table 67 Morphosyntactic structures of disjunctive questions in French
 (raw frequencies)

	Typical question form			
	Dialogue	Narrative	Other	Total
Inversion+ <i>ou</i>	10	22	0	32
No inversion+ <i>ou</i>	12	0	0	12
Total of typical form				44
	Elliptical form			
Noun	2	17	1	20
Verb	2	5	0	7
Adjective	0	1	0	1
Preposition	0	0	1	1
Pronoun	0	1	0	1
Insubordination	0	1	0	1
Total of elliptical form				31
Total				75

Disjunctive questions in French

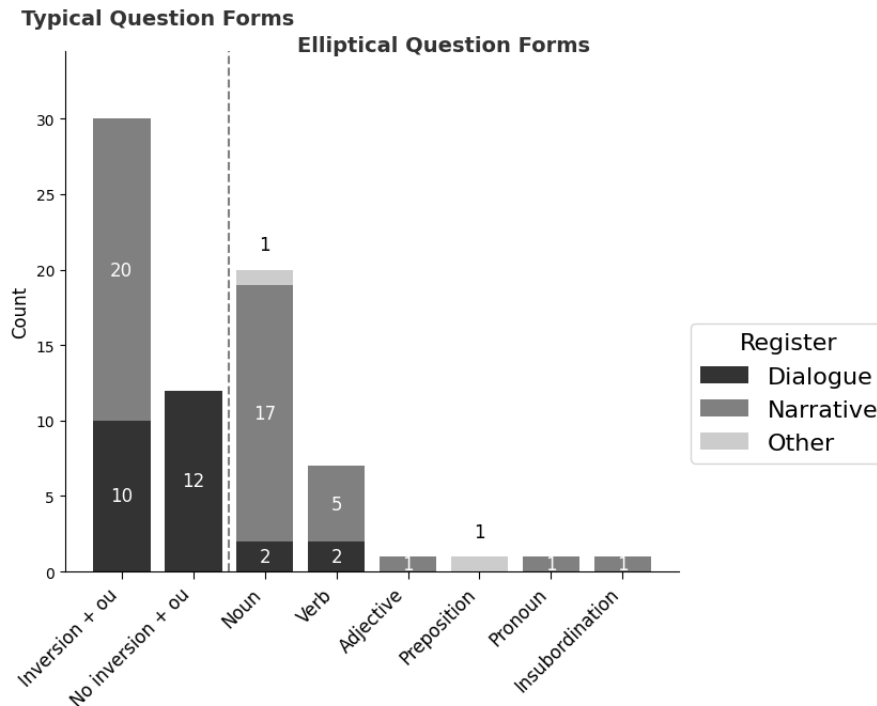


Figure 20 Morphosyntactic structures of disjunctive questions in French

Most instances appear in narrative ($n = 47$). However, *No inversion* appears exclusively in dialogue, as they essentially maintain the declarative word order, as in (4.15), a structure typically associated with spoken language, as evidenced in corpus studies (Coveney 2011: 126; Abeillé, Crabbé & Marandin 2012: 4; Guryev & Delafontaine 2022: 61).

- (4.15) *C' est un mariage ou c' est un mort ?*
 DEM COP ART marriage or DEM COP ART dead
 'It's a wedding or it's a dead?' (Un avion sans elle)

Elliptical disjunctive questions typically juxtapose options from the same lexical category (e.g., nouns, verbs, adjectives), linked by the connective *ou* 'or', as in (4.16).

- (4.16) *Un bandeau alors, ou une jaquette?*
 ART headband then or ART dust.jacket
 'A wrap-around band then, or a dust jacket?' (L'Anomalie)

In summary, **French questions predominantly employ typical morphosyntactic forms**, including inversion, the *est-ce que* particle, and declarative word order (no inversion). However, the distribution of **these structures varies across question types**:

polar questions tend to use declarative word order, whereas *wh*- and disjunctive questions more frequently involve inversion. Regarding register, while **most questions occur in dialogue**, certain structures, such as inversion in polar questions, *wh*-infinitives, and insubordinate conditional clauses, are more common in narrative, indicating that register influences morphosyntactic choices. Nevertheless, due to the complexity of morphosyntactic structures and the aim of purely describing the structures in the two languages, statistical tests are not implemented at this stage. A comprehensive statistical analysis of these morphosyntactic structures will be presented after all descriptive results and the analyses of the three topics, in Section 5.11.

4.2.3.3 Summary of morphosyntactic structures of questions in Taiwan Mandarin and French

Regarding the morphosyntactic structures of questions, **both languages predominantly employ typical forms across question types.** However, each question type has shown a preference for specific morphosyntactic structures. In Taiwan Mandarin, particles are frequently used in polar questions, while *wh*- and disjunctive questions often occur without particles. In French, polar questions typically retain declarative word order, whereas *wh*- and disjunctive questions mostly involve inversion.

With respect to register, although **most questions appear in dialogue in both languages**, register-based differences emerge when each structure is examined in detail. In Taiwan Mandarin, the particle is often found in dialogue for polar questions but tend to occur in narrative for *wh*- and disjunctive questions. In French, inversion seems to play a key role in differentiating registers: for polar and disjunctive questions, inversion tends to occur in narrative, whereas non-inversion structures are more frequent in dialogue. For *wh*-questions, however, both inversion or non-inversion structures are commonly found in dialogue.

Overall, this overview highlighted the morphosyntactic patterns of questions in the two languages. While **both show broadly similar tendencies, favouring typical form and occurring most often in dialogue, closer examination reveals register-specific differences in the use of particular structures.** Following this overview, the next section provides contrastive insights of question for the two languages from three perspectives, registers, question types, and morphosyntactic structures.

4.2.4 Contrastive insights of questions in Mandarin and French

This section examines the distribution of questions in Mandarin and French monolingually, taking into account registers, question types, and morphosyntactic structures. The main findings for the two languages considered separately are as follows: (1) questions are used more frequently in dialogue, (2) polar questions are the most common type, and (3) questions are mostly constructed in their typical forms for each question type.

From a cross-linguistic perspective, however, a notable contrast arises. For instance, **disjunctive questions display an opposite register preference**: they are more frequent in narrative in French but more commonly in dialogue in Taiwan Mandarin. This difference may relate to distinct strategies for mitigating face-threatening acts. In French, disjunctive questions in dialogue are often reduced to polar questions, with only one option explicitly stated, thereby softening the face-threatening effect (Levinson 2012: 23). In Mandarin, by contrast, disjunctive questions often use the A-not-A structure, which functions as a disjunctive structure without an explicit connective. This implicitness serves a similar purpose to reduce the face-threatening impact.

Additionally, the **two languages also diverge in how polar questions are formed**. Polar questions in Mandarin make frequent use of particles, whereas French polar questions typically favour declarative word order. Moreover, a substantial quantitative difference in the use of particles is also observed (Taiwan Mandarin: $n = 957$; French: $n = 236$). This contrast likely reflects typological differences in the encoding of discourse-pragmatic functions. Particles in Asian languages are necessary to convey such functions, while European languages tend to use intonation instead (Hancil, Post & Haselow 2015: 18).

In summary, while Taiwan Mandarin and French share general tendencies in the use and distribution of questions, they differ in the morphosyntactic realization of particular question types. These cross-linguistic contrasts are explored further in the following sections, which address three key topics: the taxonomy of questions, the *wh*-in-situ phenomenon, and the use of question particles. Each of these discussions also serves as the foundation for the bidirectional parallel and monolingual comparable analyses in Chapter 5.

4.3 Contrastive analysis: taxonomy of questions

From a typological viewpoint, questions are classified into two categories: polar and constituent. However, many monolingual studies treat disjunctive questions as a distinct type, as evidenced in Mandarin (Li & Thompson 1981: 520–521; Huang, Y.-H. Audrey Li & Li 2009; Wang & Sun 2015; Liou 2020) and similarly in French (Beyssade 2006; Riegel, Pellat & Rioul 2009; Larrivée & Guryev 2021). As such, a three-way distinction is widely adopted in the literature.

Despite the dominance of this traditional three-way distinction, Her, Che & Bodomo (2022) argue that the dichotomy suffices to account for all question types. Their argument is supported by pragmatic and semantic analyses. Pragmatically, there are confirmation-seeking and information-seeking questions. Semantically, the distinction lies in the number of set of propositions (singleton vs. two or multiple). Their refined taxonomy is also reflected in the syntactic structures associated with each question type. The theoretical validity of this dichotomy has been demonstrated across various languages, including Taiwan Mandarin, Xiang (Sinitic) and English (Her, Che & Bodomo 2022), Taiwan Southern Min (Hsiao & Her 2021) and Paiwan (Austronesian) (Huang & Her 2024). In addition, empirical evidence from experimental studies lends the support to the dichotomy of questions, particularly in Mandarin (Hsu, Chen, & Her, submitted) and French (Hsu, Allasonnière-Tang, & Her, in preparation). In line with the principle of Occam's razor, this empirically grounded binary distinction should therefore be favoured for its theoretical economy and explanatory adequacy.

By integrating a range of theoretical perspectives and experimental findings, this section seeks to address the following primary research question:

Research question: How can the distribution of question types be explained within individual and across languages?

In response to this question, two hypotheses are proposed, each situated at a distinct level of linguistic analysis, namely, the monolingual and cross-linguistic levels.

Hypothesis 1: Drawing on the theory of social economics of questions (Levinson 2012), it is hypothesised that polar questions are most frequently used in a given language in order to minimise the social cost associated with questioning.

Hypothesis 2: Based on the dichotomy of question types and the findings from two experiments of grammaticality judgement, it is posited that, although the dichotomy of questions exists in both Mandarin and French, French polar questions are only marked prosodically, lacking syntactic differentiation. Consequently, from a cross-linguistic perspective, the proportion of French polar questions should approximate the combined proportion of polar and disjunctive questions in Mandarin. This assumption follows from the fact that, from a syntactic perspective, French polar questions subsume certain disjunctive questions, since French lacks a clear syntactic distinction between the two types, whereas Mandarin overtly differentiates them.

The subsequent analysis addresses the research question by empirically testing these two hypotheses.

Table 68 and Figure 21 illustrate the distribution of question types in the two original languages.

Table 68 Question types in Taiwan Mandarin and French

(raw and relative frequencies)				
	Polar	<i>Wh</i>	Disjunctive	Total
Taiwan Mandarin	1,756	1,617	466	3,839
	(45.74%)	(42.12%)	(12.14%)	(100%)
French	2,328	1,793	75	4,196
	(55.5%)	(42.7%)	(1.8%)	(100%)

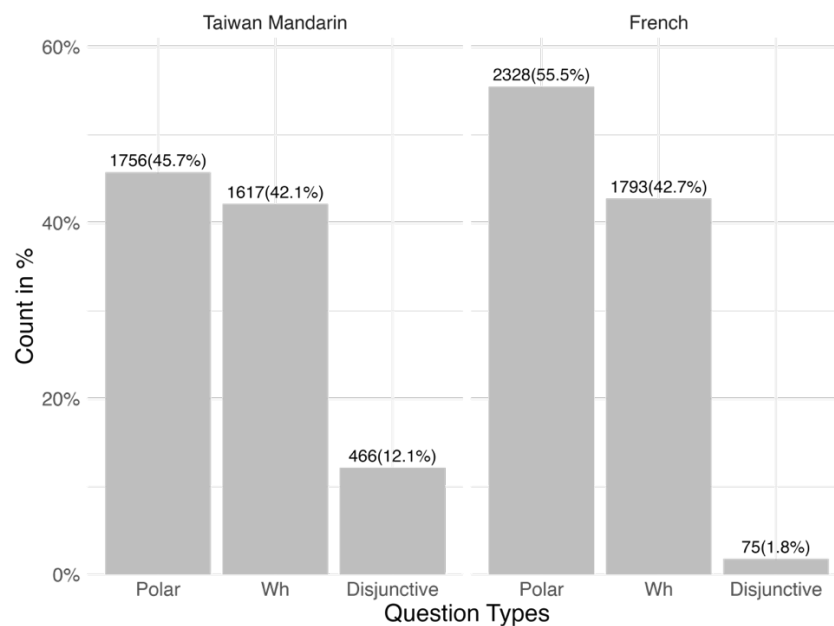


Figure 21 Question types in Taiwan Mandarin and in French

In both languages, **polar questions constitute the most prevalent type**, followed by *wh*-questions and disjunctive questions, **confirming the first hypothesis**. Chi-square tests indicate that the overall distribution of question types **differs significantly** in Taiwan Mandarin ($\chi^2 (2) = 783.8, p < .001$) and in French ($\chi^2 (2) = 1981.3, p < .001$). Pairwise chi-square tests were conducted to examine whether the differences in frequencies between question types were statistically significant. In **Taiwan Mandarin**, significant differences are found between polar and disjunctive questions ($\chi^2 (1) = 892.20, p < .0001$) and between *wh*- and disjunctive questions ($\chi^2 (1) = 741.05, p < .0001$), but not between polar and *wh*-questions ($\chi^2 (1) = 1.97, p = 0.161$). In **French**, all three types of questions differ significantly from each other: polar and *wh*-questions ($\chi^2 (1) = 40.59, p < .0001$), polar and disjunctive questions ($\chi^2 (1) = 3850.22, p < .0001$), and *wh*- and disjunctive questions ($\chi^2 (1) = 3021.37, p < .0001$). Overall, **both languages show a similar distribution pattern, with polar questions being the dominant types. However, in Taiwan Mandarin, the difference between the frequencies of polar and *wh*-questions is not statistically significant, unlike French, where all contrasts are significant.**

The predominance of polar questions as the preferred question type can be explained by the theory of social economics of questions (Levinson 2012). Questions inherently carry a potential social cost, as they may impose on the addressee by requiring a response and simultaneously expose the inquirer's lack of knowledge (Brown &

Levinson 1987: 66; Blum-Kulka 1987: 139–140; Levinson 2012: 16–17). Consequently, questions within conversational contexts can lead to potential social costs, including face loss and the risk of miscommunication (Levinson 2012: 20). According to this theory, polar questions are expected to occur with higher frequency due to their effectiveness in eliciting responses while reducing the amount of requested information and the associated social cost.

As discussed in Sections 2.1.1 and 2.3.1, questions can be distinguished according to the set of propositions they denote (Hamblin 1958: 166; Bhatt & Dayal 2020: 1124; Her, Che & Bodomo 2022: 262–265). Polar questions consist of singleton sets of propositions (with only one member), requiring confirmation of a single proposition through closed responses, such as *yes* or *no*. In contrast, *wh*-questions prompt open-ended answers, eliciting information from a broader set of possible propositions, known as non-singleton sets (with two or more members). In addition, the set of propositions associated with *wh*-questions may partially overlap with the singleton set of propositions in polar questions, since answering a *wh*-question entails answering the corresponding polar questions (Groenendijk & Stokhof 1997: 1094). Given that polar questions involve the confirmation of only one proposition, they effectively reduce the quantity of information requested. This reduction also helps to mitigate the risk of face loss. Consequently, it is predicted that polar questions will occur more frequently in conversations. This prediction is supported by empirical observation in spoken English (Stivers 2010) and by cross-linguistic studies on question-answer sequences based on the videotaped corpora of natural conversation in ten different languages (Enfield, Stivers & Levinson 2010). Hence, nine out of ten languages exhibit a higher frequency of polar questions compared to *wh*-questions in informal social interaction (Levinson 2012: 24). In addition to these findings based on spoken data, **the present study also supports this prediction**. Although the data analysed in this study come from written texts, a distinction between register in fiction (dialogue vs. narrative) has been made, which provides access to the “spoken language” represented in fiction. Accordingly, as illustrated in Figure 14 (reproduced here as Figure 22), **polar questions consistently outnumber other types in dialogues**.

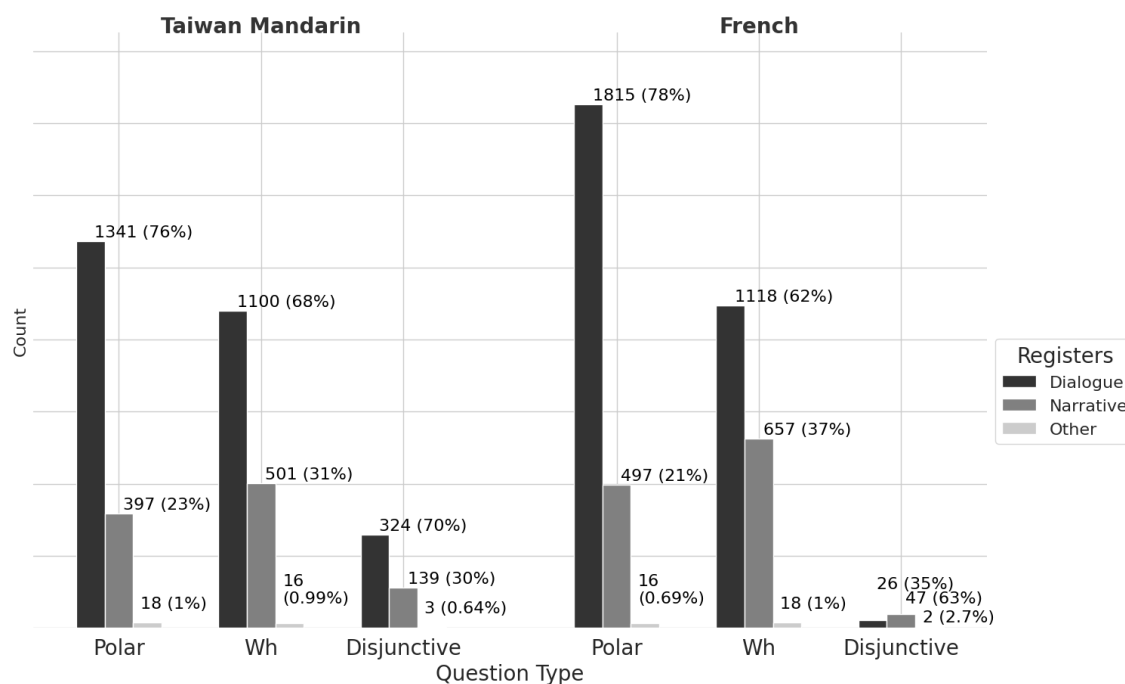


Figure 22 Question types by registers in Taiwan Mandarin in French by registers

Cross-linguistically, the distribution of question types shows notable divergence. As shown in Table 68 and Figure 21 (reproduced here as Table 69 and Figure 23), polar questions occur more frequently in French than in Mandarin, whereas Mandarin demonstrates a higher proportion of disjunctive questions in comparison to French. This observation is **statistically supported** by the results of a chi-square test of independence ($\chi^2 (2) = 356.63, p < .0001$), indicating a highly significant association between question types and languages.

Table 69 Question types in Taiwan Mandarin and French
(raw and relative frequencies)

	Polar	Wh	Disjunctive	Total
Taiwan Mandarin	1,756 (45.74%)	1,617 (42.12%)	466 (12.14%)	3,839 (100%)
French	2,328 (55.5%)	1,793 (42.7%)	75 (1.8%)	4,196 (100%)

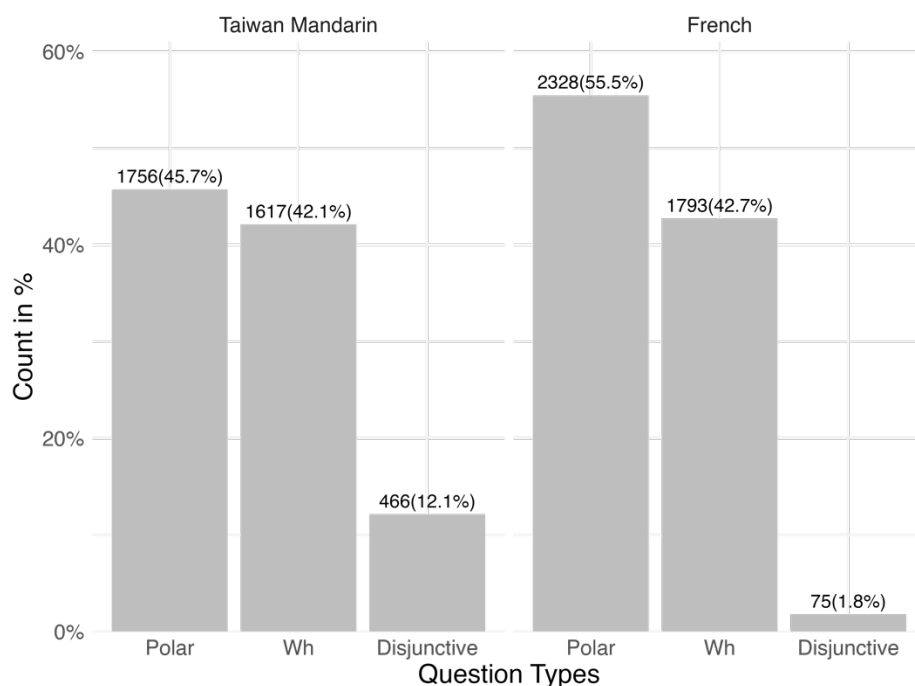


Figure 23 Question types in Taiwan Mandarin and in French

This pattern is further illustrated in the association plot in Figure 24, which offers a comprehensive visual representation of the relationship between question types and languages. In this plot, **blue** indicates frequencies higher than expected, while **red** signifies frequencies lower than expected.

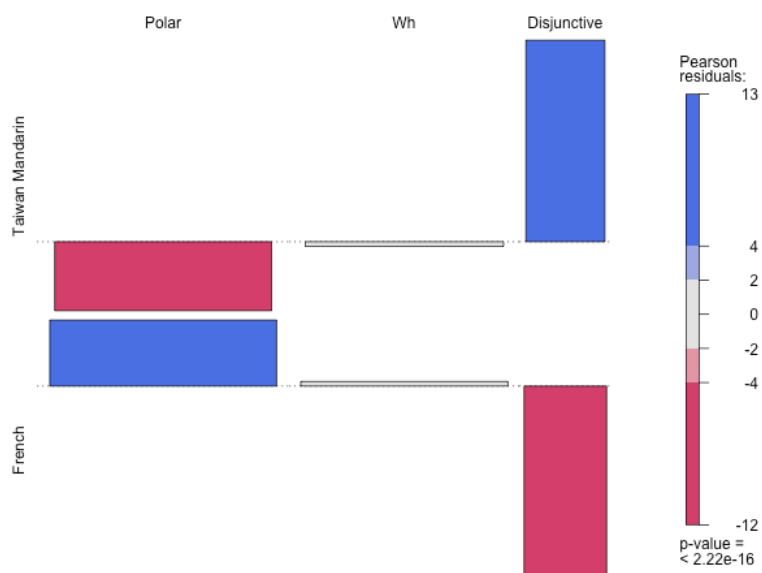


Figure 24 Association between question types and languages

Accordingly, polar questions are more frequent than expected in French, while disjunctive questions are underrepresented in the same language. Conversely, in Mandarin, polar questions occur less frequently than expected, while disjunctive questions, marked by a large blue box, display a strong positive association with the language. This suggests a **language-specific preference for particular question types**.

As illustrated above, the distribution of polar and disjunctive questions primarily accounts for the cross-linguistic differences in question types between Mandarin and French. The underlying reasons for this phenomenon may be interpreted from multiple perspectives. From a **sociocultural** standpoint, cross-cultural research in pragmatics has shown that norms governing speech act performance (e.g., apologies, requests, compliments) differ significantly across cultures (Blum-Kulka & Olshtain 1984: 197; Meier 2010: 75; House & Kádár 2021: 21; McConachy & Spencer-Oatey 2021: 734). More specifically, culture differences influence communication styles, such as the degree of directness (Blum-Kulka 1987; Bowe & Martin 2007: 47, 57; Tawalbeh & Al-Oqaily 2012; Ly 2016; Ge, Park & Pietromonaco 2022). In the context of requesting, although polar questions may reduce face-threatening effects for speaker and thereby lower the social costs (Levinson 2012), they may be less conducive to preserving interpersonal harmony, which is a core collectivist value emphasised in many Eastern societies (Yeung & Kashima 2012: 448; Kim & Park 2015). In contrast, individuals from Western cultures are often characterised as adopting more direct communicative styles than their Eastern counterparts (Hall 1976: 91; Hsieh 2011: 284–285; Park, Uchida & Kitayama 2016). Given the inherent directness of polar questions, they are more frequently used in French than in Mandarin. While the sociocultural perspective provides a plausible explanation for the observed cross-linguistic variation, it remains a necessarily simplified account. Broad generalizations concerning cultural norms and communication style risk oversimplification and may lead to problematic assumptions (Meier 2010: 80; Park et al. 2012; Spencer-Oatey & Kádár 2016: 75–77). Nevertheless, since this dissertation primarily focuses on morphosyntactic analyses, a detailed exploration of sociocultural factors falls outside its scope. The following discussion will therefore turn to a **structural** explanation, with particular emphasis on the second hypothesis, which aligns more directly with the theoretical aims of this study.

As evidenced by the data, **the second hypothesis is also supported: the proportion of French polar questions (55.5%) approximates the combined proportion of polar and disjunctive questions in Mandarin (57.8%)**. This numerical proximity is thus theoretically meaningful, as it mirrors how structural distinctions between question types manifest differently in the two languages. The experimental findings on the dichotomy of question types based on morphosyntactic features (see Section 2.3.2) indicate that Mandarin systematically distinguishes polar questions from constituent questions (i.e., *wh*- and disjunctive questions) through a range of morphosyntactic features (e.g., particles, adverbs, question forms). By contrast, French lacks morphosyntactic tools to reliably distinguish polar questions from another type. Since French does not reliably mark the distinction between polar questions and the other type by morphosyntactic features, this structural ambiguity is reflected in language use. In other words, some French polar questions in the corpus may contain another question type. This interpretation aligns with the statistical results: while the frequency and proportion of *wh*-questions are similar in both language, French displays a higher-than-expected number of polar questions. This apparent surplus of French questions may thus include disjunctive questions that are not formally marked as such. In summary, the corpus findings corroborate the second hypothesis derived from the experiment results. Therefore, the similarity in proportion empirically reflects the theoretical expectation that the lack of morphosyntactic differentiation in French causes polar questions to possibly encompass some disjunctive questions. A more detailed examination of the morphosyntactic properties of French polar questions may further illuminate how disjunctive questions can be subsumed under the structural form of polar questions.

The first piece of morphosyntactic evidence concerns the **use of the particle *est-ce que***. As discussed in Section 2.5.3, this particle can appear across all question types in French (Abeillé & Godard 2021: 1402), as shown in the polar question in (4.17), the *wh*-question in (4.18), and the disjunctive question in (4.19). The example for disjunctive question is cited from the literature, as no such instance is attested in the present corpus.

- (4.17) *Est-ce qu' on peint chez les elfes ?*
 PTCL 3SG paint PREP ART elves
 'Do elves paint?' (*La vie des elfes*)

(4.18) *Où est-ce que tu vas chercher des idées pareilles ?*
 where PTCL 2SG go search ART idea similar
 ‘Where do you get such ideas?’ (Le voleur d’ombres)

(4.19) *Est-ce que tu sors ou est-ce que tu préfères rester à la maison ?*
 PTCL 2SG go.out or PTCL 2SG prefer stay PREP ART house
 ‘Do you go out or do you prefer to stay at home?’ (Druetta 2021: 91)

These examples indicate that the presence of the *est-ce que* particle cannot be used to distinguish French polar questions from French disjunctive questions, as the particle is grammatically allowed in both.

Further morphosyntactic evidence lies in **the structure of disjunctive questions and the use of the French connective *ou* ‘or’**. As discussed in Section 2.2.2, disjunctive questions are often analysed as a subtype of polar questions, frequently described as a combination of two polar questions (Coveney 2011: 113; Larrivée & Guryev 2021: 11; Abeillé & Godard 2021: 1410). Conversely, the syntactic structure of polar questions is considered degenerate in comparison to that of disjunctive questions (Karttunen 1977: 383, 391–392). Moreover, the disjunctive structure is often treated as a strategy for marking polar questions (Siemund 2001: 1016; König & Siemund 2007: 297–298). This theoretical view allows for the possibility that polar questions may contain a covert disjunctive structure, such as *or not* in English in (4.20) or *ou pas* ‘or not’ in French in (4.21). When this covert structure is made explicit, the questions become formally equivalent to disjunctive questions. It should be noted that the two examples in (4.20) and (4.21) are illustrated of the theoretical point and are not drawn from the corpus, as the implicit element they exemplify are not overtly attested in natural data.

(4.20) *Do you want to come [OR NOT]?*

(4.21) *Est-ce que tu viens [OU PAS] ?*
 PTCL 2SG come [or NEG]
 ‘Do you want to come?’

In addition, the connective *ou* ‘or’ in French maintains the same form in both interrogative and non-interrogative contexts, a pattern also observed in English, Italian, and German (Mauri 2008: 28; Hölzl 2018: 65–66; Druetta 2021: 99–100). In these languages, the same connective is used to express both the standard disjunctions (i.e.,

inclusive reading), as in (4.22), and interrogative disjunctions (i.e., exclusive reading), as in (4.23) (Haspelmath 2007: 25–26).

(4.22) *Mike is a psychologist or a linguist* (inclusive or reading)

(4.23) *Marvin died on Tuesday or Wednesday* (exclusive or reading)

The key distinction between these two readings lies in the truth values of the propositions. Under an inclusive reading, both propositions may be simultaneously true or individually true. For example, as illustrated in (4.22), Mike may be either a psychologist or a linguist, or both. In contrast, the exclusive reading allows only one proposition to be true, as in (4.23), where Marvin died either on Tuesday or Wednesday, but not both. However, since both readings are conveyed by the same connective, they are not easily distinguished on a formal basis, but must instead be inferred pragmatically (Haspelmath 2007: 27). As Abeillé & Godard (2021: 1743) point out, the exclusive reading tends to prevail when the two propositions are presented in contrast, as illustrated in (4.24) from the corpus.

(4.24) *Et s' il l' était, serait-ce avant ou*
 and if 3SG that COP.IMPART COP.COND-DEM before or
après la disparition de Mme Courtin ?
 after ART disappearance of Mrs. Courtin
 ‘And if he was, would it be before or after Mrs. Courtin's disappearance?’
 (Trois jours et une vie)

The exclusive reading is preferred because the two prepositions *avant* ‘before’ and *après* ‘after’ are semantically opposed. In the absence of such contrast, the default interpretation of *ou* ‘or’ is inclusive (Abeillé & Godard 2021: 1743), as shown in example (4.25) from the corpus.

(4.25) *Donc, vous avez des démangeaisons, ou des brûlures*
 so 2SG have ART itchiness or ART burns
sur le visage ?
 PREP ART face
 ‘So, you have itching or burning on your face?’ (L’Anomalie)

In this example, both symptoms are allowed to co-occur as truth values. That said, the interpretation of the connective *ou* ‘or’ can at times be ambiguous, as illustrated by (4.26) and (4.27).

(4.26) *Marvin left on Tuesday or Wednesday?* (Haspelmath 2007: 26)

- (4.27) *Était-il en migration lui aussi, ou en fuite ?*
 be.IMPARF-3SG PREP migration 3SG also or PREP escape
 ‘Was he migrating too, or on the run?’ (Gataca)

In such cases, contextual cues become critical, especially when the propositions are not entirely in opposition, as in (4.27). Nevertheless, unless overridden by contextual information, the inclusive reading is generally preferred as the default interpretation (Abeillé & Godard 2021: 1743). Given that inclusive readings in interrogative context typically surface as polar questions, the association between polar questions and inclusive disjunction further supports the theory of social economics of questions (Levinson 2012). According to this theory, polar questions are favoured in discourse due to their potential to minimise social cost. This interpretation is consistent with the corpus data, which show that polar questions largely outnumber disjunctive questions in French, and certain instances containing the connective *ou* ‘or’ display no clear opposition between alternatives. Consequently, given the properties of the French connective *ou* ‘or’, a subset of polar questions in French may indeed be interpreted as disjunctive questions. These questions may, on the one hand, involve a covert disjunctive structure; on the other hand, exhibit overt use of *ou* ‘or’, introducing ambiguity between the inclusive and exclusive readings.

In sum, the analysis of the **first part of the research question and the first hypothesis** reveals that polar questions constitute the most frequent question types in both Mandarin and French, though in Mandarin they outnumber *wh*-questions only by a small margin. This observation is consistent with the theory of social economics of questions (Levinson 2012), which posits that polar questions are favoured due to their ability to reduce social cost in interaction. With regard to the **second part of the research question and the second hypothesis**, a cross-linguistic difference emerges in the distribution of polar and disjunctive questions. While Mandarin has a higher proportion of disjunctive questions, the number of polar questions in French is higher than expected, with their total nearly equivalent to the combined number of polar and disjunctive questions in Mandarin. This significant divergence can be attributed to both sociocultural (e.g., the directness of communication style) and morphosyntactic factors. The latter aspect is evidenced not only from the results of the experiment showing that in French, no syntactic features (e.g., adverbs, question form) effectively support a dichotomy of questions. This is further corroborated by the analysis of other syntactic properties, including the *est-ce*

que particle, disjunctive structure, and the connective *ou* ‘or’ in French. Collectively, the evidence suggests that a portion of polar questions in French may, in fact, represent disjunctive questions in disguise. This aligns with the theory of social economics of questions (Levinson 2012), as structural ambiguity allows speakers to reduce social cost. Nevertheless, in contexts where social cost is less relevant, such as narrative in the texts, disjunctive questions appear to be used more frequently.

In contrast, Mandarin displays a higher than expected frequency of disjunctive questions, which are distinct from polar questions by certain morphosyntactic features (e.g., particles, adverbs, and question form). In addition, Mandarin employs specific disjunctive structures (e.g., A-not-A and VP-Neg constructions) and morphologically distinguishes between inclusive (i.e., 或是 *huòshì*) and exclusive disjunctions (i.e., 還是 *háishì*). These morphosyntactic differences between Mandarin and French are clearly illustrated by the structural distribution of question types observed in the corpus data.

4.4 The analysis of *wh*-in-situ in French

As discussed in the previous section, a notable cross-linguistic distinction lies primarily in the use of polar and disjunctive questions. By contrast, the frequency of *wh*-questions is relatively comparable in Mandarin ($n = 1,617$, 42%) and French ($n = 1,794$, 43%). In addition to this similarity in frequency, both languages share a syntactic structure in the formation of *wh*-questions, namely, their question words can be found *in-situ*. This is illustrated in (4.28) for Mandarin and in (4.29) for French.

(4.28) 它 像 什麼 ?
 tā xiàng shénme?
 3SG like what
 ‘What does it look like?’ (私家偵探 *Sī Jiā Zhēn Tàn*)

(4.29) *Tu te souviens de quoi ?*
 2SG REFL remember PREP what
 ‘What do you remember?’ (*Un avion sans elle*)

In Mandarin, *wh*-in-situ is a genuine syntactic structure and represents the default strategy for forming *wh*-questions (Huang 1982; Cheng 1991: 10–11; Gryllia et al. 2020: 3; Pan 2022: 9; Chen 2024: 14). In French, however, *wh*-in-situ is only one of several available structural variants and is not obligatory (Cheng & Rooryck 2000; Coveney 2011; Larrivée & Guryev 2021: 13; Guryev & Delafontaine 2023: 300). Accordingly, **the following**

analysis focuses primarily on French *wh*-questions. I will present the distribution of *in-situ* across various French *wh*-words and examine the syntactic environments that either permit or restrict the use of this construction.

Although *wh*-in-situ is one of the structural variants available for *wh*-questions in French, this option is not uniformly applicable across all *wh*-words. As illustrated in Figure 25 (reproduced from Section 2.4.4), each question word in French exhibits a distinct tendency toward either *in-situ* or fronted position. With the exception of *wh*-words at the two extremes, most question words can occur in either position, though with varying degrees of preference, which is influenced, in part, by the syntactic relation between the question word and the main verb (Druetta 2018: 42; Fliessbach & Rockstroh 2024).

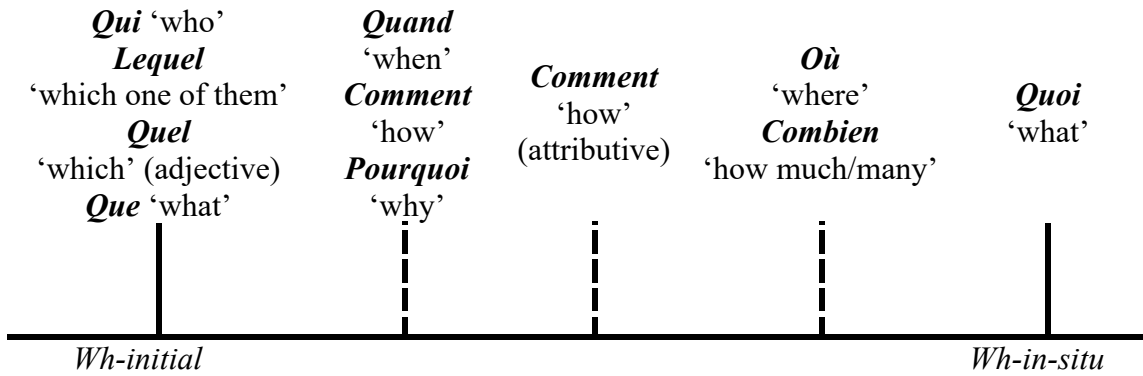


Figure 25 Distributional tendency of question words (Druetta 2018: 43)

In addition to the positional preferences associated with individual question words, previous studies have shown that the use of *wh*-in-situ in French is further constrained by both syntactic and pragmatic factors (Mathieu 1999; Cheng & Rooryck 2000; Coveney 2011: 130, 136–137; Shlonsky 2012; Guryev & Delafontaine 2023). That is, although certain question words are more likely to appear in-situ given their syntactic relation with the verb, the actual realization of *wh*-in-situ also depends on the broader structural and discourse context. Syntactically, constraints arise internally from structural domains, such as strong islands (e.g., complex noun phrases as relative clauses) and weak islands (e.g., negation, modal verbs, and universal quantifiers), as discussed in Section 2.4.3 and 2.4.4. Externally, structural constraints include the presence of focus marker (e.g., adverbs *seulement*), as in (4.30) (Mathieu 1999: 447) and the use of emphatic *wh*-expressions (e.g.,

wh-diable), as in (4.31) (Obenauer 1994: 300). Both restrict the acceptability of *wh*-in-situ.

(4.30) * *Seulement Jean arrive à faire quoi ?*
 only Jean arrive PREP do what (Mathieu 1999: 447)

(4.31) * *Il parle de quoi diable ?*
 3SG talk PREP what **hell** (Obenauer 1994: 300)

Pragmatically, *wh*-in-situ is typically licensed in presupposed contexts where a negative answer would be unexpected (Chang 1997: 42; Cheng & Rooryck 2000: 6), as exemplified in (4.32).

(4.32) A: *Marie a acheté quoi ?*
 Mary have buy.PTCP what
 ‘What did Mary buy?’

B:?? *Rien.*
 Nothing.

In other words, the answer to a *wh*-in-situ question is generally expected and highly answerable (Myers 2007: 153), given that the speaker assumes shared knowledge is represented in the proposition of *wh*-in-situ questions with addressee (Rosemeyer 2023: 231).

Note that *wh*-in-situ questions are employed not only for information-seeking purposes, but also as echo questions (Engdahl 2006: 104; Myers 2007: 152; Adli 2015: 179). The dialogue in (4.33) demonstrates the use of *wh*-in-situ in an information-seeking context, whereas example (4.34) illustrates its use in an echo question.

(4.33) A: *Ton fils, il lit quoi ?*
 2SG.POSS son 3SG read what
 ‘Your son, what does he read?’

B: *Des bandes dessinées.*
 ART comics
 ‘Comic books’

(Engdahl 2006: 104)

(4.34) A: *Mon fils, il lit [inaudible]*
 1SG.POSS son 3SG read
 ‘My son, he reads...’

B: *Il lit quoi ?*
 3SG read what
 ‘He reads what?’

(Engdahl 2006: 104)

However, *wh*-in-situ is generally less constrained by syntactic and pragmatic contexts when functioning as an echo question (Chang 1997: 17–18; Cheng & Rooryck 2000: 13). Accordingly, the following analysis will mainly focus on non-echo *wh*-questions. The discussion will address the following two research questions and corresponding hypotheses:

Research question 1: Does the distributional tendency of *wh*-words *in-situ* conform to the pattern reported by Druetta (2018: 43)?

Research question 2: Do the syntactic constraints proposed in the literature effectively restrict the placement of *wh*-words, or can counterexamples be identified in the current corpus?

Hypothesis 1: The *in-situ* tendencies of question words observed in the French comparable corpus will align with the patterns described in Druetta (2018: 43).

Hypothesis 2: In line with findings from previous empirical studies (Mathieu 2004; Adli 2006; Zimmermann & Kaiser 2019), the syntactic constraints proposed in the literature do not consistently limit the use of *wh*-in-situ.

These research questions and hypotheses will be discussed subsequently. First, I will present the overall frequency of *wh*-in-situ across different genres and individual French novels. I will then examine the distributional tendencies of specific question words *in-situ*, focusing on their occurrences between two registers (i.e., dialogue vs. narrative). Finally, I will analyse the syntactic contexts in which non-echo *wh*-in-situ questions appear.

Following the coding of the morphosyntactic structures of the French *wh*-questions, 6.4 % of the *wh*-questions ($n = 114$) involve *wh*-in-situ. Figure 26 provides an overview of the distribution of *wh*-in-situ across French novels and the two genres (detective vs. fantasy).

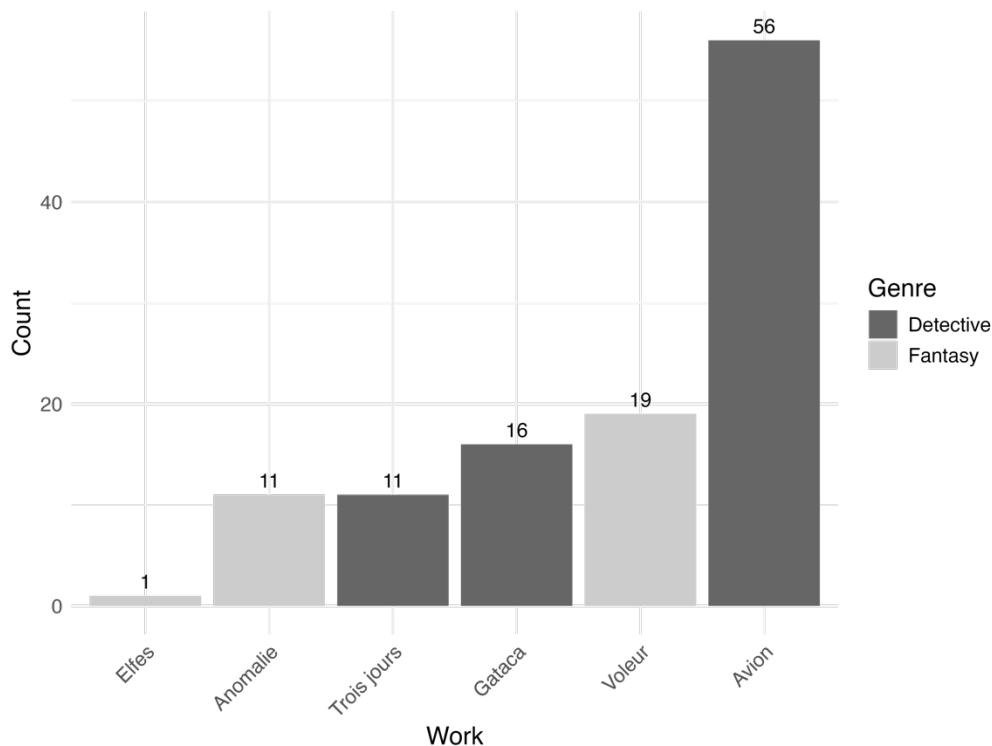


Figure 26 Distribution of *wh*-in-situ by genres and works

Over 70 % of the *wh*-in-situ appear in the detective genre ($n = 83$), while only 27 % are in the fantasy genre. To assess whether genre predicts the use of *wh*-in-situ, a generalised linear mixed-effects model was fitted using the *glmer* function. Genre was treated as a fixed effect, and individual works were included as a random effect to account for intertextual variation. The results reveal **no statistically significant difference in the distribution of *wh*-in-situ between the two genres** (z value= -0.5, $p = .618$). Nevertheless, variability in the use of *wh*-in-situ is attributed to differences across individual works (variance = 0.47). While minor variation exists in the frequency of *wh*-in-situ across novels, this variation is primarily linked to the nature of question words and the legitimate syntactic contexts in which they occur. Therefore, the following analyses will concentrate on these two aspects to address the research questions and test the corresponding hypotheses.

With respect to the first research question, Figure 27 shows the frequency of *in-situ* structure across the *wh*-words.

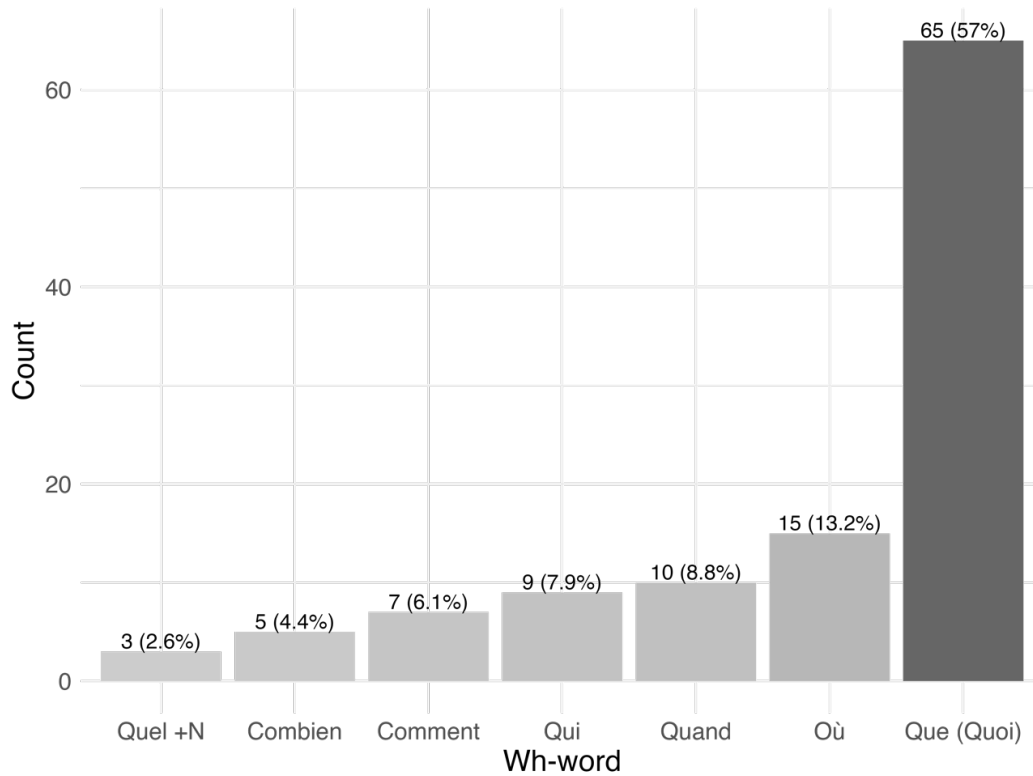


Figure 27 Distribution of *wh*-words *in-situ*

The distribution observed in the corpus supports the first hypothesis, as it closely parallels the patterns reported in the French spoken corpus (GARS) by Druetta (2018: 43). The question word *quoi* ‘what’ is the most frequently used *in-situ*, largely due to its syntactic requirement to remain *in-situ*, as shown in (4.35). The second most frequent *in-situ wh*-question word is *où* ‘where’, as shown in (4.36).

(4.35) *C' est quoi, ton nom ?*
 DEM COP **what** POSS.2SG name
 ‘What is your name?’ (Gataca)

(4.36) *Tu vas où ?*
 2SG go **where**
 ‘Where do you go?’ (Le voleur d'ombres)

Other question words, including *quand* ‘when’ in (4.37), *qui* ‘who’ in (4.38), *comment* ‘how’ in (4.39), *combien* ‘how much’ in (4.40), and *quel* ‘which’ functioning as a determiner in (4.41), exhibit comparatively fewer instances of *in-situ* usage.

(4.37) *On commence quand ?*
 3SG start **when**
 ‘When do we start?’ (L'Anomalie)

- (4.38) *C' est **qui**, cette Mélanie ?*
 DEM COP **who** DEM Mélanie
 'Who is it, this Mélanie?' (Un avion sans elle)
- (4.39) *Tu fais ça **comment** ?*
 2SG do that **how**
 'How do you do that?' (Un avion sans elle)
- (4.40) *Nous avons **combien** de Chinois?*
 1PL have **how.many** PREP Chinese
 'How many Chinese do we have?' (L'Anomalie)
- (4.41) *Vos employés finissent à **quelle heure** ?*
 POSS.2PL employee finish PREP **which hour**
 'What time do your employees finish?' (Gataca)

These question words are not typically used *in-situ*, but they do appear in a sentential final position in specific discursive contexts, particularly where the role of interlocutors, the range of the possible answers, and the expectations of the inquirers are highly constrained. In other words, when an inquirer seeks information within an established context, these *wh*-words are more likely to appear *in-situ* (Boucher 2010b: 108). For example, in (4.37), the context suggests that the patient has accepted the therapy and anticipates its immediate initiation. In (4.38), the name *Mélanie* has already been introduced, and the inquirer seeks additional information about her. Similarly, in (4.39), the question is posed with the presumption that the interlocutor has the necessary expertise to complete a challenging task. In (4.40), the consultant has previously mentioned the grouping of individuals by nationality at an airbase, and the president requests the specific number of Chinese nationals. Finally, (4.41) involves a police interrogation at a crime scene.

Although these question words may freely appear either *in-situ* or in clause-initial position (Druetta 2018: 41–42), their occurrences ($n = 920$) in the corpus predominantly exhibit clause-initial placement, with or without the subject-verb inversion, as shown in Table 70 alongside the corresponding *in-situ* frequencies.

Table 70 Frequency of *wh*-words by position in typical question forms

	<i>In-situ</i>	Fronted
<i>Que (Quoi)</i> ‘what’	65	226
<i>Comment</i> ‘how’	7	165
<i>Pourquoi</i> ‘why’	0	159
<i>Qui</i> ‘who’	9	126
<i>Quel</i> ‘which’	3	99
<i>Où</i> ‘where’	15	95
<i>Combien</i> ‘how much/many’	5	30
<i>Quand</i> ‘when’	10	20
Total	114	920

Note that *pourquoi* ‘why’, unlike other adverbial question words (e.g., *comment* ‘how’, *où* ‘where’, *quand* ‘when’), does not appear *in-situ* in the corpus. While this absence may be partially attributed to the limited size of the corpus, it nevertheless aligns with previous observation that *pourquoi* ‘why’ is rarely attested *in-situ* (Rizzi 1990: 47; Beyssade 2006: 181; Myers 2007: 153; Coveney 2011: 130; Druetta 2018: 41–42). The question word *pourquoi* ‘why’ is often associated with complex inversions (Grevisse & Goosse 2007: 496; Fairon, Simon & Grevisse 2018: 889), as shown in (4.42), which features both a full noun phrase (i.e., *Lucie*) and a pronoun (i.e., *elle* ‘she’) as the subjects of the question.

- (4.42) *Pourquoi Lucie était-elle là ?*
 why Lucie COP.IMPART-3SG there
 ‘Why Lucie was there?’ (*La vie des elfes*)

Since *wh*-in-situ is often regarded as a characteristic of colloquial and spontaneous speech (Adli 2006: 169–170; Coveney 2011: 130; Zimmermann & Kaiser 2019: 138), **it is predominantly found in dialogue** ($n = 106$) and rarely occurs in narrative ($n = 8$), as shown in Figure 28.

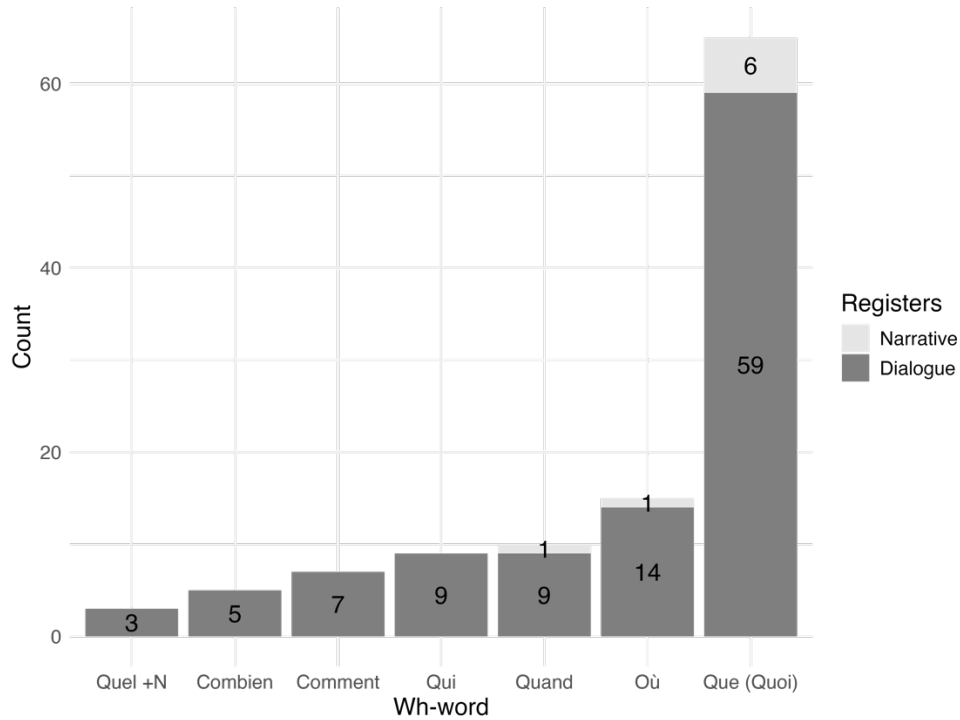


Figure 28 The distribution of *wh*-words *in-situ* across registers

Among the eight instances identified in narrative, most involve the question words *quoi* ‘what’ ($n = 6$), as in (4.43). One case involves *où* ‘where’, as in (4.44), and the other is *quand* ‘when’, as in (4.45).

(4.43) *Vous auriez fait quoi, à ma place ?*
 2PL have.COND do.PTCP **what** PREP POSS.1SG place
 ‘What would you have done in my place?’ (Un avion sans elle)

(4.44) *C’ est où, ça, Bodrum-Gumbet ?*
 DEM COP **where** that Bodrum-Gumbet
 ‘Where is Bodrum-Gumbet?’ (Un avion sans elle)

(4.45) *Le père d’ Antoine était parti quand ?*
 ART father PREP Antoine COP.IMPARF leave.PTCP **when**
 ‘When did Antoine’s father leave?’ (Trois jours et une vie)

These questions are interpreted either as direct addresses from a character to the reader, as illustrated in (4.43), or as representation of the character’s internal thoughts (i.e., internal speech (Leech & Short 2007: 270), as demonstrated in (4.44) and (4.45).

In response to the second research question and hypothesis concerning the syntactic constraints on the use of *wh*-in-situ, and following the approaches of Adli (2006) and Zimmermann & Kaiser (2019), this section examines the syntactic contexts in which *wh*-

in-situ occurs. The focus is placed primarily on internal structural constraints, including negation, modal verbs, universal quantifiers, and relative clauses. The aim is to identify any empirical evidence that may contradict the syntactic constraints proposed in the literature.

Upon analysis of the *wh*-in-situ instances in the corpus, **only two types of counterevidence were identified**. First, *wh*-in-situ may co-occur with modal verbs, such as *veux* ‘want’, as illustrated in (4.46) and (4.47). Second, *wh*-in-situ may also appear within embedded clauses, as shown in (4.48).

(4.46) *Tu veux savoir quoi ?*
 2SG want know **what**
 ‘What do you want to know?’ (Un avion sans elle)

(4.47) *Tu veux en venir où ?*
 2SG want that come **where**
 ‘What's your point?’ (Un avion sans elle)

(4.48) *Vous avez cru qu' il s' agissait de qui ?*
 2PL have believe.PTCP COMP
 3SG REFL act PREP **who**
 ‘Who did you think it was?’ (L'Anomalie)

However, the limited number of *wh*-in-situ instances in the corpus restricts the extent to which the full range of syntactic constraints proposed in the literature can be evaluated.

Overall, *wh*-in-situ is more frequently observed in detective novels (70%) than in fantasy novels, though this difference is not statistically significant. **The findings corroborate distributional tendencies of *in-situ* across French question words, as previously reported by Druetta (2018: 43)**. The *quoi* ‘what’ and *où* ‘where’ are most frequently found *in-situ*, whereas other question words are generally fronted and tend to occur *in-situ* only within contextually established discourse. Regardless of inherent positional tendencies, ***in-situ* question words occur primarily in dialogue**, consistent with their association with spoken French. Furthermore, while *wh*-in-situ is claimed to be constrained by specific syntactic contexts, both prior empirical studies and the present findings challenge some of these proposed restrictions, particularly the co-occurrence of *wh*-in-situ with modal verbs and occurrence in embedded clauses. Although *wh*-in-situ is one of the shared features between Mandarin and French, this section focuses on its distribution only in French. These findings provide a foundation for subsequent analyses,

such as examining how *wh*-in-situ structure is translated in bidirectional parallel study, and exploring differences between non-translated and translated French in monolingual comparable study.

4.5 Contrastive analysis: question particles

Regarding the previous section on French *wh*-in-situ, it is important to note that *wh*-in-situ in Mandarin is argued to be licensed through the feature-checking mechanisms involving question particles (Cheng 1991: 26; Tsai 1994b: 142–145; Cheng & Rooryck 2000: 2). In other words, the presence of these particles enables the question words to appear in-situ in *wh*-questions in Mandarin. Among the various question marking strategies employed in Mandarin, particles represent a particularly prominent feature. In addition, Mandarin has a rich inventory of particles (Chao 1968: 797–814; Li & Thompson 1981: 238; Li 2006: 6; Chu 2009: 282; Pan 2021; Yip & Rimmington 2021: 151; Fang & Hengeveld 2022: 879). The most common types of particles in questions (e.g., 嗎 *ma*, 吧 *ba*, 啊 *a*, and 呢 *ne*) have been discussed in Sections 2.5.1.1 to 2.5.1.4. Additional variants of the particle 啊 *a* (e.g., 呀 *ya*, 啦 *la*, 哇 *wa*) are also attested, along with less frequent particles, such as 哦 *ou*.

In contrast to Mandarin, as discussed in Section 2.5.3, French has only one dominant particle, namely *est-ce que*, in questions. In addition, two other particles *ça* and *donc*, are restricted to *wh*-questions. Despite the cross-linguistic differences in the use and distribution of particles, both Mandarin and French employ particles as a marking strategy for questions. Accordingly, this section mainly focuses on the syntactic properties of question particles in the two languages, with some references to their discursive and pragmatic functions where relevant. Moreover, although the focus of this study is on canonical question forms, it is important to note that certain particles can occur or only occur in elliptical questions, particularly the *ça* and *donc* in French. The research question and hypothesis guiding this section are as follows:

Research question: What is the distribution of question particles across different question types and registers in Mandarin and in French?

Hypothesis: Given that particles often carry pragmatic and discourse-related functions, they are expected to occur more frequently in dialogue.

In the following discussion, I will first examine the four main particles in Mandarin, followed by an analysis of the particles found in French (i.e., *est-ce que*, *ça* and *donc*).

4.5.1 Particles in Mandarin

In Mandarin, particles appear in 957 instances of typical question forms in Mandarin, accounting for 30% of all questions in this category. In addition, the use of particles varies significantly across question types, as demonstrated in Table 71 and Figure 29.

Table 71 Distribution of particles in typical form questions in Mandarin

	Polar	<i>Wh</i>	Disjunctive	Total
Particle	57.9 % (n = 828)	8.9 % (n = 115)	3.1 % (n = 14)	30.2 % (n = 957)
No particle	42.1 % (n = 601)	91.1% (n = 1184)	96.9% (n = 431)	69.8% (n = 2,261)
Total	100% 1,430	100% 1,300	100% 446	100% 3,173

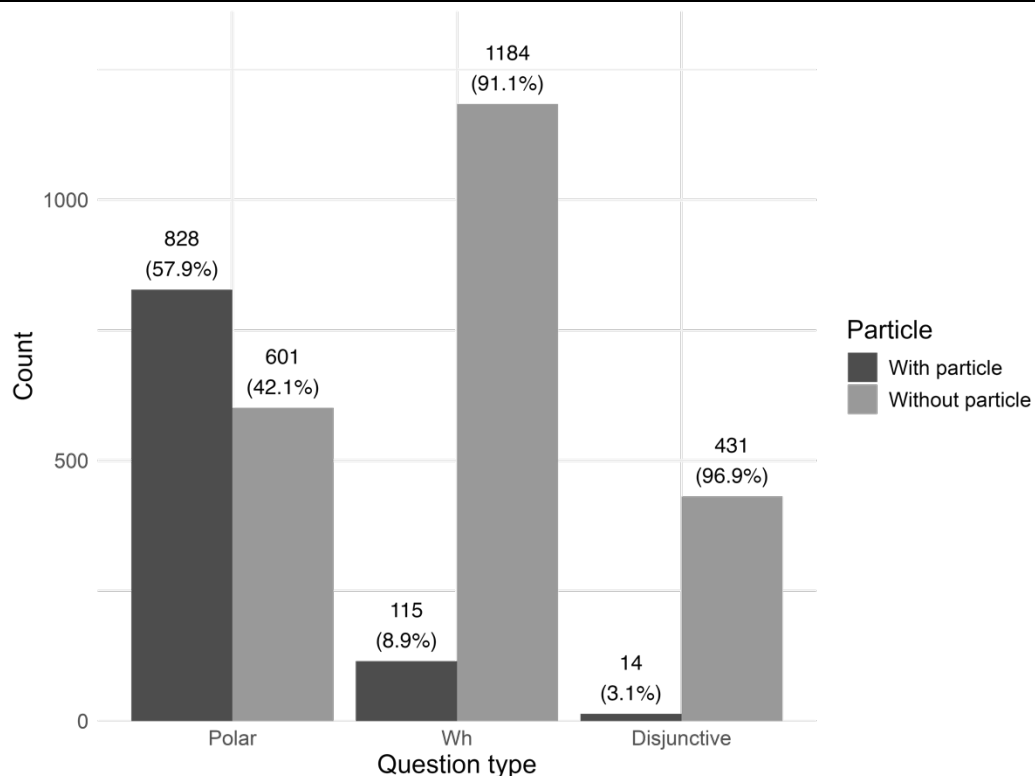


Figure 29 Distribution of particles in typical form questions in Taiwan Mandarin

Around 60% of the polar questions contain particles, while fewer than 10% of *wh*- and disjunctive questions are with particles. A chi-square test of independence confirms a

significant association between question types and the presence of particles ($\chi^2 (2) = 957.78, p < .001$). As discussed in Section 2.5.1, each question type may use different particles and certain particles are exclusive to particular question types (e.g., the 嗎 *ma* and 吧 *ba* particles only occur in polar questions). The detailed distribution of individual particles is provided in Table 72.

Table 72 Distribution of different particles across question types

	嗎 <i>ma</i>	吧 <i>ba</i>	哦 <i>o/lo</i>	啊 <i>a/la/ya</i>	呢 <i>ne</i>	Total
Polar	666	112	29	21	0	828
<i>Wh</i>	0	0	0	20	95	115
Disjunctive	0	0	0	0	14	14

In the light of the marked discrepancy in the distribution of particles across question types, as illustrated in Figure 29 and Table 72, a number of observations can be drawn.

First, **particles are rarely used in the information-seeking questions** (*wh*- and disjunctive questions). In these cases, the particle (i.e., 呢 *ne*) is optional and typically serves to reinforce the interrogative force. However, this force is already conveyed by the presence of question words in *wh*-questions or the disjunctive structure (e.g., A-not-A) in disjunctive questions, rendering the use of particle generally unnecessary. Exceptions arise when speakers aim to fulfil specific pragmatic functions, such as showing relevance to the previous context or signalling topic continuation, as illustrated in (4.49).

(4.49) A: [...]違反 了 就 會 變成 海膽 [...]
wéifǎn le jiù huì biànchéng hǎidǎn
 violate ASP just will become sea.urchin
 ‘If you violate it, you’ll become a sea urchin.’

B: 你 看 過 誰 變成 海膽 了 呢?
nǐ kàn guò shéi biànchéng hǎidǎn le ne?
 2SG see ASP who become sea.urchin ASP PTCL
 ‘Who have you seen turn into a sea urchin?’ (複眼人 *Fù Yǎn Rén*)

In this dialogue, the use of the 呢 *ne* by speaker B highlights the relevance of preceding topic (“becoming a sea urchin”) and invites further interaction. Thus, the 呢 *ne* particle is primarily associated with topic management and discourse progression. By contrast, the 嗎 *ma* particle plays an crucial role in polar questions, as it is closely tied to the epistemic

stance of the inquirer. This explains its much higher frequency compared to the 呢 *ne* particle in the corpus.

Secondly, although Fang & Hengeveld (2022: 884) argue that the particle 嗎 *ma*, used in polar questions, only serves to reinforce the interrogative force, it appears to function as an important marking strategy in Mandarin. In fact, polar questions in Mandarin are often referred to as “particle questions” or “*ma* questions” (Li & Thompson 1981: 547; Li & Thompson 1984: 54; Cheng 1984: 89), underscoring the central role of the particle. Moreover, the predictability of an answer may differ depending whether the particle is present (Li 2006: 34–35). Consider the two following polar questions from the corpus, (4.50) is a polar question with a particle, while (4.51) is without a particle.

(4.50) 你 知道 嗎?
nǐ zhīdào **ma**?
2SG know PTCL
‘Do you know that?’ (泡沫戰爭 *Pào Mò Zhàn Zhēng*)

(4.51) 你 知道?
nǐ zhīdào?
2SG know
‘You know that?’ (天橋上的魔術師 *Tiān Qiáo Shàng De Mó Shù Shī*)

According to the theory of predictability (Haan & Van Heuven 2003), polar questions with the typical particle 嗎 *ma* as in (4.50) show that the speaker holds significant uncertainty regarding the answer (Chao 1968: 800, 804). In such cases, the response is unlikely to be predicted and the inquirer has no clear expectation of how the interlocutor will reply. By contrast, polar questions without the particle 嗎 *ma*, which instead rely on declarative word order and prosodic cues, as in (4.51), are used when the speaker finds the answer relatively predictable. In these instances, the inquirer shows greater confidence in anticipating the response of the interlocutor. In addition to their high degree of predictability, declarative questions are highly context reliant (Poschmann 2008: 249; Stivers & Rossano 2010: 8–9; Wang 2021: 594). They often carry an epistemic bias, reflecting the speaker’s prior assumptions or stance (Gunlogson 2008: 103; Krifka 2015: 335–336; Kiss 2024: 11; Munteanu & Kiss 2025: 2). Specifically, speakers may lean toward expecting a negative response, revealing a form of bias in their questions. The question in (4.51) is then understood to convey a biased answer that the speaker appears

not to expect the interlocutor to know the answer at the moment of asking. Taken together, the contrast between polar questions with declarative word order and those marked with the 嗎 *ma* particle suggests that the latter constitutes the default and unmarked strategy in Mandarin. Polar questions with 嗎 *ma* particle typically seek neutral confirmation and involve minimal epistemic commitment from the speaker. This may explain the frequent use of 嗎 *ma* particle in polar questions, particularly when the speaker aims to avoid projecting bias or presuppositions.

Nevertheless, this finding contradicts the theory of social economics of questions, which predicts a preference for declarative word order in polar questions (Levinson 2012: 23). This prediction is further supported by cross-linguistic studies involving ten languages (Enfield, Stivers & Levinson 2010). However, my results are in line with those of Wang (2021: 595) based on a spoken corpus of Mandarin. Given that theory of social economics of questions is predominantly grounded in observations from Western languages, the discrepancy observed here may reflect deeper linguistic and cultural distinctions between Eastern and Western language systems. Linguistically, Western languages often use inversion (i.e., word order change) as an additional morphosyntactic strategy for questions, while Mandarin primarily relies on particles. In Western languages, the use of declarative word order polar questions is preferred because it closely resembles a statement, thereby reducing the social cost for the speaker. On the other hand, in Mandarin, polar questions marked with the particle 嗎 *ma* tend to convey a neutral epistemic stance. In comparison, polar questions with declarative word order in Mandarin often carry a bias toward a specific answer. Therefore, asking a polar question with the particle would conform to the norm of politeness strategy in Eastern societies. This type of polar question is perceived as more polite and less direct, making them more socially acceptable in contexts where deference and face-saving are valued.

Finally, particles in Mandarin are generally associated with spoken language or conversational contexts in written texts (Li & Thompson 1981: 238; Li 2006: 1; Fang & Hengeveld 2022: 873), as they serve various pragmatic and discursive functions that express emotive and epistemic nuances from the perspective of the speaker (Wu 2004: 2; Chu 2009; Simpson 2014: 157; Pan 2021: 8). Accordingly, **the present findings support the hypothesis that particles are often used in dialogue**. However, a closer look at the distribution reveals that this tendency primarily applies to particles used in polar question,

which predominantly appear in dialogue. By contrast, **particles found in *wh*- and disjunctive questions are more frequently used in narrative**, as shown in Figure 30.

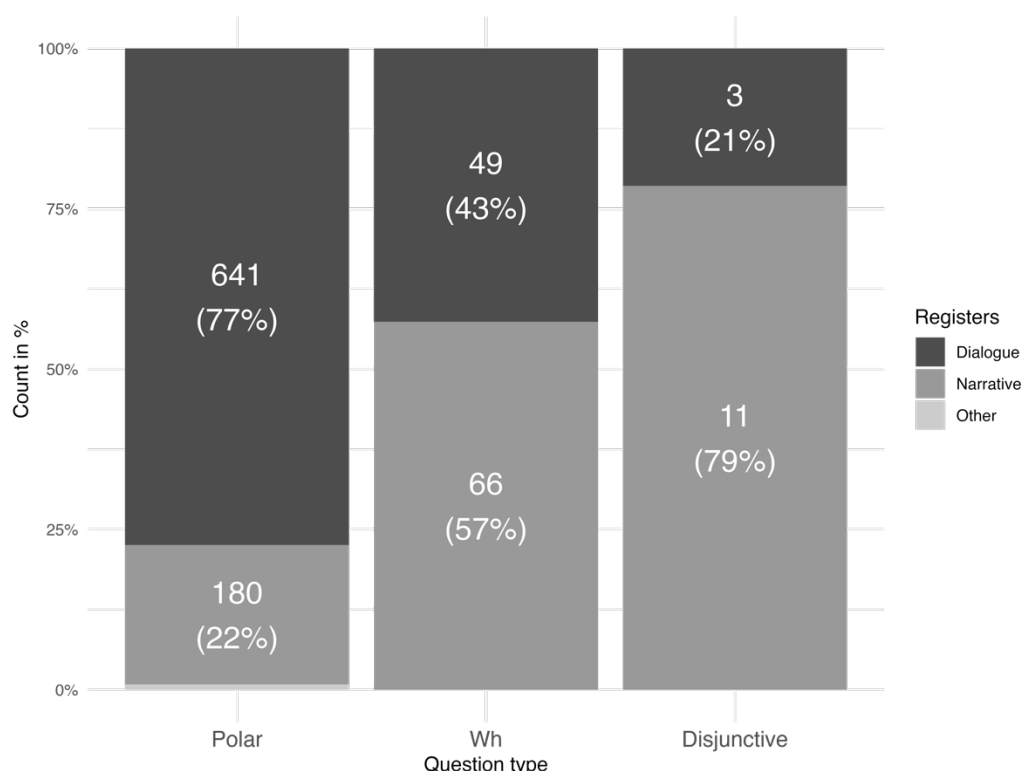


Figure 30 Distribution of particles by registers in Mandarin

This discrepancy can be attributed to two main reasons. First, as previously discussed, *wh*- and disjunctive questions are more likely to omit particles compared to polar questions. The particle in *wh*- and disjunctive questions often functions merely to reinforce the interrogative force, since this force is already inherently encoded by the presence of question words in questions. Accordingly, in spoken language context where communicative efficiency is prioritised, such particles are more likely to be omitted.

Secondly, unlike the particles used in polar questions (e.g., 嗎 *ma* and 吧 *ba*), the 呢 *ne* particle typically found in *wh*- and disjunctive questions serves primarily to enhance contextual relevance (Li & Thompson 1981: 306; Chu 2009: 298). In other words, the function of the 呢 *ne* particle is closely tied to discourse structure. It maintains coherence with preceding context, thereby facilitating the progression of ideas within a relevant thematic frame (Chu 2009). Moreover, it encourages participation from the addressee. In fictional texts, the 呢 *ne* particle plays a crucial role in driving the narrative forward, as illustrated in (4.52).

- (4.52) 小黑人 平常 都 在 想 什麼
 xiǎohēirén píngcháng dōu zài xiǎng shénme
 little.black.man usually all ASP think what
 事情 呢？
 shìqíng *ne?*
 thing PTCL
 ‘What’s usually on the little black man’s mind?’

(天橋上的魔術師 *Tiān Qiáo Shàng De Mó Shù Shī*)

This context explores the internal reflections of *Little black man*. The sentence preceding the question describes his contemplative posture, setting the stage for introspection. The use of the 呢 *ne* particle in (4.52) contributes to maintaining thematic continuity and facilitates the development of subsequent thoughts concerning the mental state of *Little black man*. Beyond its role in discourse progression, questions with the 呢 *ne* particle can also serve to engage readers directly, as illustrated in (4.53).

- (4.53) 他 想 從 屋頂 逃逸，
 tā xiǎng cóng wūdǐng táoyì
 3SG want from roof escape
 但是 往 哪 逃 呢？
 dànshì wǎng nǎ táo *ne?*
 but to where escape PTCL
 ‘He wanted to escape from the roof, but where?’ (私家偵探 *Sī Jiā Zhēn Tàn*)

The 呢 *ne* particle in (4.53) enhances the interrogative force by framing the question from a neutral narrative or character perspective, encouraging reader engagement to consider possible places of refuge sought by the character. The function of the 呢 *ne* particle aligns with the broader narrative technique observed in fictional texts, where questions in narrative do not necessarily seek information or elicit answers, particularly in the absence of a direct addressee (Solomon 1990: 91, 92). Rather, such questions are strategically employed to guide readers through a chain of thoughts and to foster deeper engagement with the unfolding narrative, which helps maintain their interest in the story (Solomon 1990: 98, 101).

In Mandarin, particles predominantly appear in polar questions, where they influence the epistemic stance of the speaker and the perceived predictability of the response. Conversely, ***wh-* and disjunctive questions often omit particles** due to their redundant role in merely reinforcing the interrogative force. However, **when *wh-* and**

disjunctive questions add a particle, its usage is often found in the narrative. This pattern aligns with functional role of such particles, particularly 呢 *ne*, in maintaining contextual coherence. In essence, they contribute to the narrative flow and serve to encourage reader engagement.

In the following section, the focus shifts to French particles. While French does not exhibit the same degree of particle diversity as Mandarin, several noteworthy parallels can be drawn between the two languages.

4.5.2 Particles in French

Particles in French include *est-ce que*, *ça*, and *donc*. The *est-ce que* particle occurs exclusively with typical question forms and is frequently compared with subject-verb inversion. In contrast, the *ça* and *donc* particles are restricted to use in fragment *wh*-questions, where they fulfil specific pragmatic functions. This subsection first examines the overall distribution of the *est-ce que* particle across different question types and registers. It then turns to a discussion of the use and function of *ça* and *donc* particles in fragment *wh*-questions.

In the present corpus, the *est-ce que* particle is attested only in polar and *wh*-questions, as shown in (4.54) and (4.55). In addition, it is predominantly found in *wh*-questions rather than in polar questions, as illustrated in Figure 31. Although the particle *est-ce que* can also be used in disjunctive questions (see Section 2.5.3.1), such usage is not observed in the current dataset, possibly due to the corpus size.

- (4.54) *Est-ce que* *tout* *est* *normal* ? (Polar question)
 PTCL all COP normal
 ‘Is everything normal?’ (*Gataca*)
- (4.55) *Qu’* *est-ce que* *tu* *fais* *ici* ? (*Wh*-question)
 what PTCL 2SG do here
 ‘What are you doing here?’ (*Un avion sans elle*)

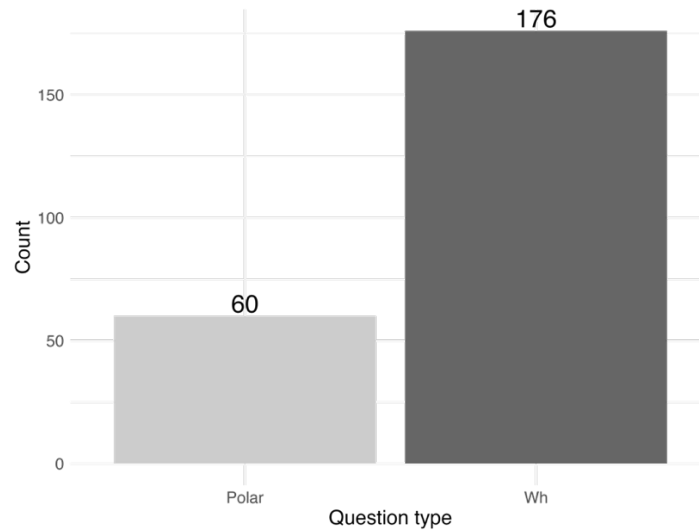


Figure 31 Distribution of the *est-ce que* particle across question types

The use of the *est-ce que* particle accounts for only 3.8% ($n = 60$) of polar questions and 14.4% ($n = 176$) of *wh*-questions. Although the relative frequency of the particle in *wh*-questions may seem low, it corresponds to about one *wh*-questions in seven, highlighting that *est-ce que* occurs with regularity. While the *est-ce que* particle does appear more frequently in *wh*-questions, its distribution is not uniform across different question words, as shown in Table 73.

Table 73 Distribution of the *est-ce que* across question words

	<i>Que</i> ‘what’	<i>Où</i> ‘where’	<i>Pourquoi</i> ‘why’	Total
Frequency	172	3	1	176

The most common pairing is with the question word *que* ‘what’, as in (4.56).

- (4.56) **Qu’** *est-ce que* *je* *dois* *dessiner ?*
what PTCL 1SG should draw
 ‘What do I have to draw’ (L’Anomalie)

While the *est-ce que* particle is also used with other question words, such as *où* ‘where’ in (4.57) and *pourquoi* ‘why’ in (4.58), these co-occurrences are comparatively rare (Adli 2015: 182; Rossi-Gensane et al. 2021: 170; Benzitoun 2022: 8).

- (4.57) **Où** *est-ce que* *tu* *vas* *chercher* *des* *idées* *pareilles ?*
where PTCL 2SG go find ART idea same
 ‘Where do you get ideas like that?’ (Le voleur d’ombres)

- (4.58) *Pourquoi est-ce que je reviens*
 why PTCL 1SG come.back
toujours à cette histoire ?
 always PREP DEM history
 ‘Why do I keep coming back to this story?’ (La vie des elfes)

Previous studies have demonstrated variations in the frequency of the *est-ce que* particle across both spoken and written French (see Section 2.5.3.1 and Guryev 2017: 114, 116–117). In the present study, **the overall frequency of the *est-ce que* particle is relatively low. This observation aligns with the findings from other studies based on written French**, such as text messages (Guryev 2017: 197; Guryev & Delafontaine 2022: 58) and novels (Reinhardt 2019: 260). Although the use of particle may be less common than other structures (e.g., declarative word order or inversion) in both polar and *wh*-questions, its occurrence is notably higher in *wh*-questions, as has been observed in previous studies (Guryev 2017; Reinhardt 2019: 260). Moreover, the frequent co-occurrence of the *est-ce que* particle with the question word *que* ‘what’ has been documented in other studies (Fromaigeat 1938: 28; Boucher 2010b: 114; Coveney 2011: 116; Dekhissi 2013: 114; Adli 2015: 191; Guryev 2017: 198). This strong association can be attributed to their long-standing usage, tracing back to ancient French (Rouquier 2002: 101). Additionally, regardless of its syntactic roles as a subject or an object, the use of *que* ‘what’ in clause-initial position poses greater structural challenges (see Section 2.4.2). While the use of the *est-ce que* particle may appear uneven across question types and question words, it is nevertheless **frequently observed in fictional dialogue**. This is reflected in Table 74 by question types and Table 75 by question words. Note that the *other* register is excluded from both tables, as it is not pertinent to the following analyses.

Table 74 Distribution of the *est-ce que* particle by registers and question types

Question type	Dialogue	Narrative	Total
Polar	43 (72.9%)	16 (27.1%)	59 (100%)
<i>Wh</i>	144 (81.8%)	32 (18.2%)	176 (100%)

Table 75 Distribution of the *est-ce que* particle by registers and question words

Question word	Dialogue	Narrative	Total
<i>Que</i> ‘what’	140 (81.4%)	32 (18.6%)	172 (100%)
<i>Où</i> ‘where’	3	0	3
<i>Pourquoi</i> ‘why’	1	0	1

Both tables highlight the predominance of the *est-ce que* particle in **dialogue**. This observation is consistent with prior research, which notes that this particle is frequently used in spoken French (Druetta 2002: 67, 2011: 29), while it is often considered inelegant in written French (Coveney 2011: 124). **The frequent use of the *est-ce que* particle in speech may be attributed to its discursive and pragmatic functions.** Beyond reinforcing the interrogative force, the *est-ce que* particle can be used to show politeness (Myers 2007: 177; Debaisieux & Boulton 2007: 36; Coveney 2011: 141) or to mark a shift in conversational topic (Boucher 2010a: 68, 2010b: 115;). These functions indicate that the *est-ce que* particle semantically conveys a weak presupposition (Coveney 2011: 142). For example, in example (4.59), the speaker addresses a school teacher by asking her important questions. The two polar questions introduced by the *est-ce que* particle may reflect politeness when addressing unfamiliar. In example (4.60), the speaker refers to having attended the funeral and then abruptly shifts the conversational topic from the funeral to the interlocutor’s present activity. This sudden transition illustrates the function of particle in marking topic shifts. However, since the study focuses primarily on morphosyntactic aspects and does not explicitly code for the particle functions, these two examples are included merely to illustrate the aforementioned points.

(4.59) *Excusez-moi, madame [...]*
excuse-me ma’am

Est-ce que les enseignants sortent pour surveiller les élèves ?
PTCL ART teachers go.out PREP watch.INF ART student

Est-ce que cette grille, là-bas, vous la fermez chaque fois ?
PTCL DEM gate there 2SG it close each time

‘Excuse me, ma’am [...] Do the teachers go out to supervise the students? Do you close that gate over there every time?’
(Gataca)

- (4.60) *J' étais présent aux obsèques de ta mère [...]*
 1SG IMPARF present PREP funeral PREP POSS.2SG mother
Qu' est-ce que tu es venu chercher
 what PTCL 2SG COP come.PTCP look.INF
dans cette cour d' école ?
 PREP DEM yard PREP school
 'I was present at your mother's funeral [...] What did you come to look for in this schoolyard?'
 (Le voleur d'ombres)

At the same time, given its relatively modest or low level of interactive potential (Guryev & Delafontaine 2022: 63), it is also plausible that the *est-ce que* particle appears in narrative of fictions, particularly in monologues where the interlocutor remains unidentified, as illustrated in (4.61) and in (4.62).

- (4.61) *Une autre idée pas si absurde :*
 ART other idea NEG so absurd
est-ce qu' une fausse vie existe après notre fausse mort ?
 PTCL ART false life exist after POSS.1PL false death
 'Another not-so-absurd idea: does a false life exist after our false death?'
 (L'Anomalie)
- (4.62) *Qu' est-ce que j' ai fait?*
 what PTCL 1SG have do.PTCP
 'What have I done?'
 (Trois jours et une vie)

In addition to the *est-ce que* particle, the particles *ça* and *donc* are also attested in the corpus. However, their frequencies are considerably lower in comparison, and their occurrences are exclusively confined to **dialogue**, as shown in Table 76. This restricted distribution may be attributed to their specific discourse functions and pragmatic force (Smirnova & Abeillé 2021: 260).

Table 76 Distribution of the *ça* and *donc* particle across registers

	Dialogue	Narrative	Other	Total
Particle <i>ça</i>	17	0	0	17
Particle <i>donc</i>	1	0	0	1

In the corpus, these particles are restricted to elliptical *wh*-questions, as illustrated in (4.63) and (4.64).

- (4.63) *Comment ça ?*
 how PTCL
 'What do you mean by that?'
 (Trois jours et une vie)

(4.64) *Pourquoi donc ?*

why PTCL

‘Why is that?’

(*L’Anomalie*)

Both particles can co-occur with nearly any type of question word (Lefevre 2018: 183), although certain tendencies emerge depending on the specific question words, as shown in Table 77.

Table 77 Distribution of the *ça* and *donc* particle across question words

	<i>Comment</i> ‘how’	<i>Qui</i> ‘who’	<i>Où</i> ‘where’	<i>Quoi</i> ‘what’	<i>Pourquoi</i> ‘why’	Total
Particle <i>ça</i>	13	2	1	1	0	17
Particle <i>donc</i>	0	0	0	0	1	1

The distribution observed in the corpus aligns with previous findings reported by Smirnova & Abeillé (2021: 246–247). In particular, the analyses indicate that ***comment* ‘how’ is among the most frequently used question words in *wh-ça*, while *pourquoi* ‘why’ is commonly found in *wh-donc* constructions.** In terms of syntactic structure, these elliptical *wh*-questions with particles follow two patterns. The first is *wh*-sluicing, as seen in (4.63) and (4.64). The second is known as *wh*-stripping, in which a remnant from the precedent context follows the elliptical question. This may take the form of a full clause, as in the interrogative sentence *nous sommes quoi ?* ‘What are we?’ in (4.65) or a noun phrase *tout le monde* ‘everyone’ in (4.66).

(4.65) *Comment ça, nous sommes quoi ?*

how PTCL 1PL COP what

‘What do you mean, we’re what?’

(*La vie des elfes*)

(4.66) *Qui ça, tout le monde ?*

who PTCL all ART world

‘Who, everyone?’

(*Un avion sans elle*)

In the present study, among the questions containing the *ça* particle, seven instances correspond to the first type (i.e., *wh*-sluicing), while ten fall under the second type (i.e., *wh*-stripping). In the latter cases, the particle *ça* functions as a device to establish coreference with the preceding discourse (Druetta 2018: 34; Lefevre 2018: 197). The repeated fragment serves to specify a referent from the prior context, prompting the interlocutor to provide further explanation. In other words, such questions are employed to narrow down and refine the topic under discussion. As demonstrated in the two

examples above, these fragments facilitate the rapid clarification of ambiguous or contextually underdetermined elements in the preceding utterance.

Regarding the *donc* particle, the only instance identified in the corpus appears in a *wh*-sluicing, as shown in (4.64). This elliptical form is the most frequently attested usage for the *donc* particle, with *pourquoi* ‘why’ being among the most commonly co-occurring question words (Smirnova & Abeillé 2021: 257). However, non-elliptical use of *donc* seems to be more widespread overall (Smirnova & Abeillé 2021: 254), as evidenced by the two cases in (4.67) and (4.68) found in the present corpus.

(4.67) *Qui donc l' a trompé ?*
 who PTCL him have deceive.PTCP
 ‘Who deceived him?’ (La vie des elfes)

(4.68) *À qui donc que tu parles ?*
 PREP who PTCL COMP 2SG talk
 ‘Who are you talking to?’ (La vie des elfes)

Regardless of whether its form is elliptical or non-elliptical, the use of the *donc* particle serves to primarily to signal the continuation of a discourse (Smirnova & Abeillé 2021: 261). For instance, the question *Pourquoi donc ?* ‘why’ in (4.64) is used by a host in response to a speaker’s comment on the shared responsibility in virtual spaces. The host expresses a desire to further explore the rationale behind this claim and invites the speaker to elaborate by asking the question in (4.64). This invitation is taken up in the subsequent turn, as shown in (4.69).

(4.69) *Parce que[...] ce test ne nous est pas destiné*
 because DEM test NEG us.OBJ COP NEG aim.PTCP
en tant qu' individus.
 as individual
 ‘Because this test is not designed for us as individuals.’ (L'Anomalie)

Overall, the particles (i.e., *est-ce que*, *ça*, and *donc*) used in French are largely observed in **dialogue**. This tendency can be attributed to the fact that the primary functions of these particle are discourse-related. Regarding their distribution across question types, although *ça* and *donc* occur exclusively in *wh*-questions, the ***est-ce que* particle also appears more frequently in *wh*-questions than in polar questions.** Moreover, the *est-ce que* particle **often co-occurs with the question word *que* ‘what’**, possibly because this pairing constitutes the earliest attested case of such co-occurrence

in the language. No instance of *est-ce que* is found in disjunctive questions, which may be due to the limited corpus size. These observations on the syntactic forms and pragmatic functions of French particles set the stage for broader comparison with Mandarin.

The use of particles is observed in both Taiwan Mandarin and French and their occurrence is most frequent in dialogue. However, the **two languages differ considerably in the number and distribution of particles**. While particles account for 30% occurrence ($n = 957$) in Taiwan Mandarin, they represent only 8% in French ($n = 236$). Furthermore, most particles in Taiwan Mandarin are clustered in polar questions ($n = 828$), whereas in French they are most often attested in French *wh*-questions ($n = 176$). Although particles predominantly appear in dialogue, they show a higher frequency in *wh*- and disjunctive questions in Taiwan Mandarin. This may be attributed to the fact that the particles used in these two question types (e.g., 呢 *ne*) primarily serve to maintain contextual coherence, a function that appears to be more closely associated with written language than spoken language.

The following section provides an overall summary of the analyses presented in this chapter. Based on these findings, the final section also offers insights and formulates hypotheses for the subsequent analyses in Chapter 5.

4.6 Contrastive analysis: summary and insights to parallel analyses

To summarise this chapter's comparable analysis, the main findings are organised around three key topics: question types, *wh*-in-situ, and question particles.

First, regarding **question types**, polar questions are the most prevalent question type in both corpora. However, a significant difference was observed in the overall distribution of the three question types across the two languages ($\chi^2 = 358.71$, $df = 2$, $p < .0001$). Specifically, Mandarin has fewer polar questions and more disjunctive questions than expected, while French shows the opposite pattern: more polar questions and fewer disjunctive questions. This discrepancy may be explained from two perspectives. From a sociocultural perspective, the communication style in Mandarin tends to be less direct, while French tends toward directness, frequently using polar questions to minimise social cost and reduce face-threatening acts for the speakers. From a morphosyntactic perspective, the distributional differences confirm the hypothesis grounded in the dichotomic theory of questions and the results from two experimental

studies, which suggest that some French polar questions may be disjunctive questions in disguise. Two pieces of evidence support this claim, owing to their structural similarities. First, the *est-ce que* particle can mark both polar and disjunctive questions in French. Second, the French connective *ou* ‘or’ can sometimes be covert in polar questions, thereby contributing to structural ambiguity. Moreover, the interpretation ambiguity of *ou* ‘or’, whether it conveys an exclusive or inclusive reading, further blurs the boundary between polar and disjunctive questions. However, since this ambiguity of interpretation arises from the use of the same connective form in both question types, corpus data alone cannot demonstrate it. Rather, such ambiguity can only be inferred from human judgments.

Regarding ***wh*-in-situ structures**, the analysis focuses primarily on French. The results show that majority of *wh*-in-situ instances occur in detective novels and are predominantly found in dialogue. This distribution aligns with previous findings that associate *wh*-in-situ with colloquial and spontaneous speech. Moreover, the pattern of question words found *in-situ* is consistent with the observations in Druetta (2018: 43). That is, the question words *quoi* ‘what’ and *où* ‘where’ are most frequently found *in-situ*. It is important to note, however, that *wh*-in-situ structure accounts for only 6.4% of all *wh*-questions in the French comparable corpus. This relatively low frequency is likely due to syntactic constraints traditionally argued to limit the use of *wh*-in-situ. Nevertheless, findings from previous studies (Adli 2006; Zimmermann & Kaiser 2019), as well as the present corpus data, suggest that these constraints may not be as rigid as previously assumed. For instance, *wh*-in-situ remains a possible structure even in the presence of modal verbs or when the question word is embedded in clauses.

Finally, this study examined **the use of question particles** in Mandarin and French. In Mandarin, particles appear in 30% of questions with typical form, with a substantial majority occurring in polar questions. This is because particles, particularly 嗎 *ma*, are essential for conveying the epistemic stance of the inquirer, such as (un)predictability of an answer. By contrast, most *wh*- and disjunctive questions do not include particles, as the particle in these cases primarily functions to maintain textual coherence rather than to express an epistemic stance. This functional distinction is also reflected in their register distribution. Particles in polar questions are frequently attested in dialogue, while those used in *wh*- and disjunctive questions more often appear in narrative. These findings suggest that Mandarin prefers polar questions with a particle, which contrasts with the

use of declarative word order anticipated by the theory of social economics of questions (Levinson 2012: 23).

This divergence may stem from both linguistic and cultural differences between Eastern and Western language systems. As for the particles in French (e.g., *est-ce que*, *ça*, *donc*), they account for only a small proportion of French questions and are mostly found in *wh*-questions, particularly with the question word *que* ‘what’. Although the distribution of particles across question types and question words may appear random, their use is largely concentrated in dialogue. The *est-ce que* particle contributes to politeness, signals a topic shift, and reinforces interrogative force. The particle *ça* particle helps establish co-reference to the prior discourse, while the *donc* particle marks the continuation of a conversation.

The corpus-based analyses presented in this chapter offer empirical insights that lay the groundwork for the hypotheses to be tested in Chapter 5, which will explore two approaches: bidirectional parallel and monolingual comparable analyses.

In the first part, which focuses on **bidirectional parallel analyses**, the hypotheses are formulated based on the specific topics under investigation. First, regarding **question types**, since some polar questions in French may represent covert disjunctive questions, it is expected that French polar questions will tend be translated with disjunctive questions in Mandarin. Conversely, given the frequent use of the distinctive A-not-A structure in Taiwan Mandarin, which has no direct corresponding structure in French, disjunctive questions in Mandarin are expected to be translated with polar questions in French, as shown in Figure 32.

Source language	Target language
French	Taiwan Mandarin
Polar question →	Disjunctive questions
Taiwan Mandarin	French
Disjunctive questions →	Polar questions

Figure 32 Hypotheses of question types translations in the two directions

Second, as for ***wh*-in-situ structures**, which can only be retrievable from the French data, the parallel analyses will remain exploratory. Without specific hypotheses, the analysis of *wh*-in-situ aims to examine how *wh*-in-situ is translated from Taiwan

Mandarin into French, and vice versa. This exploratory approach is expected to shed light on the contrastive behaviour of *wh*-in-situ structure from an empirical perspective, a topic that, to the best of my knowledge, has not yet been systematically addressed.

Third, with respect to **question particles**, the preliminary comparison between the two languages suggests that similarities in their syntactic distributions and pragmatic functions provide a solid *tertium comparationis* for the analysis of question particles in Taiwan Mandarin and French. Building on the comparable results, however, it appears that the two languages employ particles in distinct ways, and the frequency of particle occurrences also differs substantially. Therefore, it is hypothesised that question particles in the two languages are not directly translatable, particularly from Taiwan Mandarin into French. In such cases, question particles in Taiwan Mandarin are expected to be translated with other syntactic structures, as shown in Figure 33. Nonetheless, this does not exclude the possibility of translating particles with particles, given that certain pragmatic functions are shared across the two languages, albeit to a limited extent. Conversely, from French to Taiwan Mandarin, it is reasonable to hypothesise that question particles in French tend to be translated with particles in Taiwan Mandarin, since particles constitute the dominant marking strategy for questions in Mandarin.

Source language	Target language
Taiwan Mandarin	French
Question particles →	Inversion, No inversion, Declarative word order
French	Taiwan Mandarin
Question particles →	Particles

Figure 33 Hypotheses of question particles translations in the two directions

For the **monolingual comparable analyses**, this study examines how translated languages differ from their non-translated languages, namely, non-translated Taiwan Mandarin vs. translated Taiwan Mandarin, and non-translated French vs. translated French. Accordingly, two major translation effects are expected to emerge during the analyses: source language interference and normalisation (Toury 1979; Toury 2012: 303; Lefer & Vogeleeer 2013: 2). The interference of source language refers to the transfer of linguistic features from the source language into target language during the translation,

resulting in target-language outputs that resemble the source language and deviate from the norms of the non-translated language. Conversely, normalisation involves a tendency for translated languages to conform more closely to the linguistic norms of the target language, resembling its non-translated counterpart. The monolingual comparable analyses are therefore conducted on the basis of hypotheses grounded in these two translation effects.

First, regarding **question types**, polar questions are expected to remain the dominant type in both translated Mandarin and French, since this question type is the most frequent in the two original languages. However, the two languages show language specific preferences in the distribution of question types. That is, disjunctive questions occur more often than expected in Taiwan Mandarin, while polar questions are comparatively more frequent in French. Accordingly, under the influence of the two translation effects, the frequency of these two question types in translated language is expected to fluctuate compared with their non-translated languages, as shown in Table 78. Under interference of source language, the number of disjunctive questions in translated Taiwan Mandarin may decrease, while the number of polar questions increases. Conversely, source language interference may lead to an increase in disjunctive questions and a decrease in polar questions in translated French. Under normalisation, the tendencies are expected to reverse.

Table 78 Hypotheses based on two translation effects for question types

Hypothesis	Pair of comparison	Translation effect	Frequency of questions in translated language
I	Non-translated Taiwan Mandarin vs. Translated Taiwan Mandarin	Interference of source language	Disjunctive questions ↓ Polar questions ↑
	Non-translated French vs. Translated French		Disjunctive questions ↑ Polar questions ↓
II	Non-translated Taiwan Mandarin vs. Translated Taiwan Mandarin	Normalisation	Disjunctive questions ↑ Polar questions ↓
	Non-translated French vs. Translated French		Disjunctive questions ↓ Polar questions ↑

Second, given that *wh-in-situ* is the default structure in Mandarin, the monolingual comparable analysis of this syntactic structure will mainly focus on comparing non-translated and translated French. Since the *wh-in-situ* is an inherent syntactic structure in Mandarin, its frequency is expected to increase in the French translations of Taiwan Mandarin novels due to source language interference. By contrast, the number of *wh-in-situ* may decrease as a result of normalisation to the target language (i.e., French). The two corresponding hypotheses are summarised in Table 79.

Table 79 Two hypotheses based on two translation effects for *wh-in-situ*

Pair of comparison	Hypothesis	Translation effect	Frequency of <i>wh-in-situ</i> in translated language
Non-translated French vs. Translated French	I	Interference of source language	↑
	II	Normalisation	↓

Moreover, if the translated French is influenced by normalisation, the distribution of *wh*-in-situ across different question words in French translations is expected to be consistent with the patterns observed in the original French texts. Otherwise, under interference of source language, the distribution of *in-situ* question words may diverge from the original French patterns.

Finally, with respect to **question particles**, since particles constitute the dominant marking strategy for questions in Mandarin, source language interference may lead to an increased use of particle in translated French. On the other hand, the same translation effect may result in a decrease in the number of particles in translated Taiwan Mandarin, as French does not frequently employ particles. Under the normalisation effect, however, translated Taiwan Mandarin is expected to show an increased use of particles, whereas a decrease in particle use may be observed in translated French. The corresponding hypotheses are shown in Table 80.

Table 80 Two hypotheses based on two translation effects for question particles

Hypothesis	Pair of comparison	Translation effect	Frequency of question particles in translated language
I	Non-translated Taiwan Mandarin vs. Translated Taiwan Mandarin	Interference of source language	↓
	Non-translated French vs. Translated French		↑
II	Non-translated Taiwan Mandarin vs. Translated Taiwan Mandarin	Normalisation	↑
	Non-translated French vs. Translated French		↓

Above all, these hypotheses are formulated based on analyses of comparable data. It should be noted, however, that some hypotheses may be revised following the bidirectional parallel analyses presented in the first part of Chapter 5.

Chapter 5 Bidirectional parallel and monolingual comparable analyses of Mandarin and French

Building upon the findings based on comparable data in Chapter 4, this chapter adopts two additional approaches to investigate the bidirectional parallel corpus further: the **bidirectional parallel approach**, where the two translation directions are investigated, and a monolingual comparable approach, which contrasts original and translated language. Sections 5.1 through 5.4 employ a bidirectional parallel framework to examine three key phenomena: question types, *wh*-in-situ structures, and question particles. Section 5.5 provides a summary of the findings from bidirectional parallel analyses. Subsequently, Sections 5.6 through 5.9 apply a **monolingual comparable approach**, comparing original texts with their translations within the same language (e.g., original Taiwan Mandarin vs. translated Taiwan Mandarin). Section 5.10 summarises the results from the monolingual comparable analyses. The chapter concludes with a comprehensive statistical analysis of the overall distribution of question types and morphosyntactic structures, incorporating multiple factors, such as genre, works, translation status, authors, and translators.

5.1 Bidirectional parallel analysis

This section analyses the corpora using a bidirectional parallel approach, which involves comparing original texts and their translations in two directions (from Taiwan Mandarin to French and from French to Taiwan Mandarin). The following three subsections focus on the analysis of three main topics (i.e., taxonomy of questions, *wh*-in-situ, and question particles).

To present the results of the parallel analysis, two key metrics are used. The first is **mutual correspondence** (Lefer 2020: 264–265). This value measures “the frequency with which different (grammatical, semantic and lexical) expressions are translated into each other” (Altenberg 1999: 254). The calculation method is shown in Figure 34. At and Bt denote the frequencies of items A and B in the translations (i.e., target texts), while As and Bs represent their frequencies in the original texts (i.e., source texts). The mutual correspondence value captures how frequently the two items are translated into each other across languages. In other words, “the higher the mutual correspondence value is, the

greater the equivalence between the compared items is likely to be” (Altenberg & Granger 2002: 18).

$$\frac{(A \text{ in translations} + B \text{ in translations}) \times 100}{A \text{ in source texts} + B \text{ in source texts}} = n\%$$

Figure 34 Mutual correspondence formula (Altenberg 1999: 254)

In addition, the separate measure for each direction of correspondence is referred to as **translation bias**, which represents “the degree of unidirectional correspondence displayed in the [...] translations” (Altenberg 1999: 254–255). As illustrated in Figure 35, this value is calculated by identifying the number of occurrences where two items correspond across the source and target texts. This number is then divided by the total number of occurrences of item A in the source texts. A higher percentage indicates that item A in the source language is frequently translated with item B in the target language, suggesting a strong directional preference in the translation process, and vice versa.

Language X → Language Y

$$\frac{A \text{ in source texts of language X} = B \text{ in translations of language Y} \times 100}{A \text{ in source texts of language X}} = n\%$$

Language Y → Language X

$$\frac{B \text{ in source texts of language Y} = A \text{ in translations of language X} \times 100}{B \text{ in source texts of language Y}} = n\%$$

Figure 35 Translation bias formula (based on Altenberg 1999: 254)

The following analyses and discussions will primarily rely on these two metrics. While other values used in parallel studies, such as *reverse mutual correspondence* (Ebeling & Ebeling 2015: 79) and *source/target concentration* (Halverson 2017: 30–32) may also offer valuable insights, they are not directly relevant to the present investigation and are therefore not included in the current analysis. The following sections provide in-depth analyses of the three core topics introduced earlier. Each of them is examined individually in light of the bidirectional parallel data and the quantitative measures outlined above.

5.2 Bidirectional parallel analysis: Taxonomy of questions

In Section 4.3, the discussion centred on the taxonomy of questions based on comparable data (i.e., original languages). The corpus analysis revealed that, the distinct structures commonly used in Mandarin, disjunctive questions (e.g., A-not-A structure, VP-Neg) result in a significant asymmetry between Taiwan Mandarin and French in the distribution of polar and disjunctive questions. Building on this empirical observation, and drawing on the rationale based on the theory of question dichotomy and relevant experimental results in Chapter 4, I suggest that French lacks a morphosyntactic strategy for marking polar questions. As a result, certain French polar questions may at least theoretically, be disjunctive questions in disguise. These considerations motivate the central question to be explored in the context of parallel data analyses: **how are polar and disjunctive questions translated across the two directions?** In line with this objective, a working hypothesis is formulated regarding the ways in which different question types are translated between Taiwan Mandarin and French.

Hypothesis: Owing to the morphosyntactic differences between Mandarin and French, particularly the availability of the morphosyntactic structures for disjunctive questions in Mandarin, such as the A-not-A structure and VP-Neg, which are absent in French, the two languages differ fundamentally in how disjunctive questions are encoded. This cross-linguistic contrast is expected to surface in translations of question types. Specifically, my hypothesis is that **disjunctive questions in Mandarin tend to be frequently translated with polar questions in French. Conversely, French polar questions are likely to be translated with disjunctive questions in Mandarin.**

This hypothesis will be examined the following subsections. The structure of the discussion is as follows: Section 5.2.1 presents the results of question type translation from Taiwan Mandarin into French. Section 5.2.2 reports the results in the reverse direction, from French into Taiwan Mandarin. Finally, Section 5.2.3 offers a synthesis of the overall patterns observed in the translation of question types based on the bidirectional parallel data.

5.2.1 Translation of questions types from Taiwan Mandarin to French

In this subsection, I examine the translation of question types **from Taiwan Mandarin to French**, beginning with an overview of how questions are translated, as summarised in Table 81.

Table 81 Overview of translating patterns for question types, from Taiwan Mandarin to French

Questions in original texts (Taiwan Mandarin)	Questions in translations (French)	Raw Frequencies (%)
3,839 (100%)	Same question type	2,855 (74.3%)
	Different question type	582 (15.2%)
	Other	402 (10.5%)

This table presents the overall distribution of question type translations. **Over 70% of the questions in the original Taiwan Mandarin texts are translated using the same question types in French.** Approximately 15% are translated using a different question type, while around 10% are rendered without a corresponding question. In the following paragraphs, I provide a more detailed analysis of the translation patterns for each question type.

First, **polar questions** in Taiwan Mandarin are translated into French with different question types, as shown in Table 82.

Table 82 Translation patterns for polar questions from Taiwan Mandarin to French

Original (Taiwan Mandarin)	Raw frequencies (%)	Translation (French)	Raw Frequencies (%)
Polar question	1,756 (100%)	Polar question	1,493 (85%)
		<i>Wh</i> -question	80 (5%)
		Disjunctive question	19 (1%)
		Other	164 (9%)

The vast majority (85%) of the polar questions in Taiwan Mandarin are translated with polar questions in French, as shown in (5.1). Some are translated with *wh*-questions, as in (5.2), and disjunctive questions, as in (5.3), but these cases are quite rare (5% and 1% respectively).

(5.1) a. **Original Taiwan Mandarin**

他 知道 我 這個 人?
tā zhīdào wǒ zhège rén?
 3SG know 1SG DEM person

‘He knows me as this person?’

(炒飯狙擊手 *Chǎo Fàn Jū Jí Shǒu*)

b. **French translation**

Il est au courant de mon existence ?
 3SG COP PREP flow PREP POSS.1SG existence

‘He is aware of my existence?’

(*Le sniper, son wok et son fusil*)

(5.2) a. **Original Taiwan Mandarin**

小孩, 你們 有 事 嗎?
xiǎohái nǐmen yǒu shì ma?
 xiǎohái 2PL have thing PTCL

‘Children, you have something?’

(泡沫戰爭 *Pào Mò Zhàn Zhēng*)

b. **French translation**

Que se passe-t-il, les enfants ?
 what REFL pass-t-3SG ART children

‘What's happening, children?’

(*La guerre des bulles*)

(5.3) a. **Original Taiwan Mandarin**

你 兒子 要 結婚 啦?
nǐ érzi yào jiéhūn la?
 2SG son will marry PTCL

‘Your son is getting married?’

(炒飯狙擊手 *Chǎo Fàn Jū Jí Shǒu*)

b. **French translation**

Il va se marier ou quoi ?
 3SG go REFL marry or what

‘He is getting married or what?’

(*Le sniper, son wok et son fusil*)

As previously discussed, certain polar questions are not preserved as interrogatives in translation. In the present corpus, some are translated with declarative sentences, as in (5.4), while others are merged into one single question instead of being translated as two separate questions, as in (5.5). Additionally, omission is also observed, whereby the original question is left untranslated.

(5.4) a. **Original Taiwan Mandarin**

新聞 不 是 才剛 播完？
xīnwén bú shì zǎigāng bōwán?
 news NEG COP just finish
 ‘News isn’t just finished?’

(鬼地方 *Guǐ Dì Fāng*)

b. **French translation**

Le journal télévisé venait de se terminer.
 ART news television come.IMPART PREP REFL finish
 ‘The news had just finished.’
 (*Ghost Town*)

(5.5) a. **Original Taiwan Mandarin**

什麼 身分？ 私家 偵探？
shénme shēnfèn? sījiā zhēntàn?
 what identity private detective
 ‘What identity? A private detective?’

(私家偵探 *Sī Jiā Zhēn Tàn*)

b. **French translation**

De mon statut de détective privé ?
 PREP POSS.1SG status PREP detective private
 ‘My status as a private detective?’
 (*Rue du dragon couché*)

The translation patterns for **wh-questions** are presented in Table 83.

Table 83 Translation patterns for *wh*-questions from Taiwan Mandarin to French

Original (Taiwan Mandarin)	Raw frequencies (%)	Translation (French)	Raw frequencies (%)
<i>Wh</i> -question	1,617 (100%)	Polar question	138 (8.5%)
		<i>Wh</i>-question	1,310 (81%)
		Disjunctive question	1 (0.1%)
		Other	168 (10.4%)

Around 80% of *wh*-questions are translated with *wh*-questions in French, as shown in (5.6). Other question types are also used in translation, including polar questions in (5.7), accounting for about 8%, and one disjunctive question, given in (5.8).

(5.6) a. **Original Taiwan Mandarin**

你 女兒 幾 歲 ?
 nǐ nǚér jǐ suì?
 2SG daughter how.much/many age
 ‘How old is your daughter?’

(私家偵探 *Sī Jiā Zhēn Tàn*)

b. **French translation**

Quel âge a votre fille ?
 which age have POSS.2SG daughter
 ‘How old is your daughter?’

(*Rue du dragon couché*)

(5.7) a. **Original Taiwan Mandarin**

你們 目前 鎖定 哪 個 對象 ?
 nǐmen mùqián suǒdìng nǎ ge duìxiàng?
 2PL currently target which CL person
 ‘Who are you currently targeting?’

(私家偵探 *Sī Jiā Zhēn Tàn*)

b. **French translation**

Êtes-vous déjà fixé sur ceux que va
 COP-2PL already fixed PREP those REL go
viser votre action ?
 target POSS.2SG action
 ‘Have you already decided who your action will target?’

(*Rue du dragon couché*)

(5.8) a. **Original Taiwan Mandarin**

髮線 與 眉毛 的 距離 ?
 fàxiàn yǔ méimáo de jùlí?
 hairline and eyebrow DE distance
 ‘Distance between the hairline and the eyebrows?’

(鬼地方 *Guǐ Dì Fāng*)

b. **French translation**

Le front haut ou les cheveux implantés bas ?
 ART forehead high or ART hair implanted low
 ‘A high forehead or low-implanted hair?’

(*Ghost Town*)

In addition, some *wh*-questions are translated with declarative sentences, as in (5.9), or are subject to merging, where multiple questions are combined into a single one, as in (5.10). Together, these account for 10% of the occurrences.

(5.9) a. **Original Taiwan Mandarin**

你 怎麼 去 這麼 久？
 nǐ zěnmě qù zhème jiǔ?
 2SG how go so long.time

‘Why were you gone so long?’

(私家偵探 *Sī Jiā Zhēn Tàn*)

b. **French translation**

Vous en avez mis du temps !
 2PL that have put.PTCP ART time

‘You really took a while!’

(*Rue du dragon couché*)

(5.10) a. **Original Taiwan Mandarin**

要 說 什麼？ 該 說 什麼？
 yào shuō shénme? gāi shuō shénme?
 want say what should say what

‘What should (she) say? What ought (she) to say?’

(鬼地方 *Guǐ Dì Fāng*)

b. **French translation**

Qu' est-ce qu' elle aurait voulu ou dû
 what PTCL 3SG have.COND want.PTCP or must

lui dire ?

OBJ say

‘What would she have wanted or had to say to him?’

(*Ghost Town*)

Finally, Table 84 presents the distribution of **disjunctive questions** translated from Taiwan Mandarin into French.

Table 84 Translation patterns for disjunctive questions from Taiwan Mandarin to French

Original (Taiwan Mandarin)	Raw frequencies (%)	Translation (French)	Raw frequencies (%)
Disjunctive question	466 (100%)		324
		Polar question	(70%)
			20
		<i>Wh</i> -question	(4%)
			52
		Disjunctive question	(11%)
			70
		Other	(15%)

The results indicate that only 11% of disjunctive questions in Mandarin are translated with the same question type in French, as shown in (5.11).

(5.11) a. **Original Taiwan Mandarin**

你 到底 在 安慰 我 還是 糗 我 ?
 nǐ dàodǐ zài ānwèi wǒ háishì qiǔ wǒ?
 2SG finally ASP comfort 1SG or embarrass 1SG
 ‘Are you trying to comfort me or embarrass me?’ (私家偵探 *Sī Jiā Zhēn Tàn*)

b. **French translation**

*Vous cherchez à me reconforter ou à m’
 2PL look.for PREP OBJ.1SG comfort or PREP OBJ.1SG
 enfoncer ?*
 push.into
 ‘Are you trying to comfort me or push me down?’ (*Rue du dragon couché*)

In contrast, 70% are translated with polar questions in French, providing strong support for the hypothesis formulated above, namely that disjunctive questions in Mandarin are predominantly translated with polar questions in French, as in (5.12).

(5.12) a. **Original Taiwan Mandarin**

還是 有 人 在 學校 欺負 你 ?
 háishì yǒu rén zài xuéxiào qīfù nǐ?
 or have person PREP school bully 2SG
 ‘Or is someone bullying you at school?’ (私家偵探 *Sī Jiā Zhēn Tàn*)

b. **French translation**

Quelqu’un sur le campus se serait moqué
 someone PREP ART campus REFL COP.COND mock.PTCP
de toi ?
 PREP 2SG
 ‘Did someone on campus make fun of you?’ (*Rue du dragon couché*)

Notably, among these French polar questions, over 85% of them ($n=278$) are translated from Mandarin A-not-A disjunctive structures, such as 要不要 *yàobúyào* ‘want-not-want’ as shown in (5.13).

(5.13) a. **Original Taiwan Mandarin**

要 不 要 把 牠們 趕走?
yào bú yào bǎ tāmen gǎnzǒu?
want NEG want make 3PL drive.away
'Should (we) or should (we) not drive them away?'

(泡沫戰爭 Pào Mò Zhàn Zhēng)

b. **French translation**

Faut-il les chasser ?
must-3SG OBJ.3PL chase.out
'Should (we) chase them away?'
(*La guerre des bulles*)

Other translations are also observed, including *wh*-questions, as in (5.14), as well as non-interrogative renderings such as declarative sentences, as in (5.15) and merged translations, as in (5.16), accounting for 4% and 15% of the occurrences, respectively.

(5.14) a. **Original Taiwan Mandarin**

要 不 要 順便 去 基隆 吃 夜市
yào bù yào shùnbìan qù Jīlóng chī yèshì
want NEG want conveniently go Keelung eat night.market
的 營養 三明治?
de yíngyǎng sānmíngzhì?
DE nutritious sandwich
'Should (we) or should (we) not conveniently go to Keelung to eat the nutritious sandwiches at the night market?'
(炒飯狙擊手 Chǎo Fàn Jū Jí Shǒu)

b. **French translation**

Que dirais-tu d' un petit détour par Keelung pour
what say.COND-2SG PREP ARTsmall detour PREP Keelung for
y goûter les sandwichs gastronomiques du
there taste ART sandwiches gastronomic PREP
marché de nuit ?
market PREP night
'How about a little detour to Keelung to taste the gourmet sandwiches from the night market?'
(*Le sniper, son wok et son fusil*)

(5.15) a. **Original Taiwan Mandarin**

她 想 說, 吃 飽 沒?
tā xiǎng shuō chī bǎo méi?
3SG want say eat full NEG
'She wanted to say, "Have you eaten or not?"'
(鬼地方 Guǐ Dì Fāng)

b. French translation

Elle voudrait leur demander s' ils ont le
 3SG want.COND OBJ.3PL ask if 3PL have ART

ventre plein.

belly full

'She would like to ask them if they have full bellies.'

(*Ghost Town*)

(5.16) a. Original Taiwan Mandarin

該 說 是 偉大 呢？ 還是 多此一舉？

gāi shuō shì wěidà ne? háishì duōcǐyìjǔ?

should say COP great PTCL or unnecessary.action

'Should it be called great? Or a pointless action?'

(複眼人 *Fù Yǎn Rén*)

b. Translation French

Avait-ce été un exploit ou une peine inutile ?

have.IMPARF-DEM COP.PTCP ART feat or ART pain useless

'Had it been an achievement or a useless trouble?' (*L'homme aux yeux à facettes*)

The findings presented above indicate that the **majority of the questions in Taiwan Mandarin are translated into French using the same question types**. However, the **highest proportion of mismatches occurs with disjunctive questions**, which are frequently translated with polar questions in French, thus **supporting my hypothesis**. The following sub-section examines the other translation direction (i.e., from French into Taiwan Mandarin) to test the second part of the hypothesis, namely that French polar questions are likely to be translated with disjunctive questions in Taiwan Mandarin.

5.2.2 Translation of questions types from French to Taiwan Mandarin

Before turning to the analysis of individual question types, an overview of the overall translation patterns **from French to Taiwan Mandarin** is provided in Table 85.

Table 85 Overview of translating patterns for question types, from French to Taiwan Mandarin

Questions in original texts (French)	Questions in translations (Taiwan Mandarin)	Raw frequencies (%)
4,196 (100%)	Same question type	3,697 (88%)
	Different question type	409 (9.9%)
	Other	90 (2.1%)

The results show that **nearly 88% of questions are translated using the same question types**. Around 10% are translated differently, including both translations with mismatched question types and other strategies, such as declarative reformulations and omission. The following sections examine how each specific question type is translated from French into Taiwan Mandarin.

Table 86 presents the distribution of French **polar questions** and their translation into Taiwan Mandarin.

Table 86 Translation patterns for polar questions from French to Taiwan Mandarin

Original (French)	Raw frequencies (%)	Translation (Taiwan Mandarin)	Raw frequencies
Polar question	2,328 (100%)	Polar question	1,948 (83.7%)
		<i>Wh</i> -question	61 (2.6%)
		Disjunctive question	272 (11.7%)
		Other	47 (2%)

As shown in Table 86, over 80% are translated with the same question type, as in (5.17).

(5.17) a. **Original French**

Tu vois dans mon esprit ?
2SG see PREP POSS.1SG mind

‘You see into my mind?’

(*La vie des elfes*)

b. **Taiwan Mandarin translation**

妳 能夠 看 進 我 的 心 ?
nǐ nénggòu kàn jìn wǒ de xīn?
2SG can see into 1SG DE heart

‘You can see into my heart?’

(精靈少女 *Jīnglíng shàonǚ*)

The second most frequent translation is disjunctive questions, as in (5.18).

(5.18) a. **Original French**

Avait-elle mis le feu aux poudres ?
have.IMPART-3SG put.PTCP ART fire PREP powders

‘Did she spark an explosion?’

(*Gataca*)

b. **Taiwan Mandarin translation**

她 是 否 引起 軒然大波 了 ?
tā shì fǒu yīnqǐ xuānrándàbō le?
3SG COP NEG cause big.wave PTCL

‘Has she or not caused a big stir?’

(左撇子的殺意 *Zuǒpěizi de shāyì*)

Notably, more than half of these disjunctive questions ($n=152/272$) employ the A-not-A structure, such as 清不清楚 *qīngbùqīngchǔ* ‘clear-not-clear’ in (5.19). *Wh*-questions are also occasionally used in translation, as in (5.20).

(5.19) a. **Original French**

Est-ce que je me fais bien comprendre ?
PTCL 1SG REFL make well understand

‘Am I making myself clear?’

(*Trois jours et une vie*)

b. **Taiwan Mandarin translation**

我 講 得 清 不 清楚 ?
wǒ jiǎng de qīng bù qīngchǔ?
1SG speak DE clear NEG clear

‘Am I making myself clear or not clear?’

(三天一生 *Sāntiān yìshēng*)

(5.20) a. **Original French**

Je vous amène le carton de tracts ce soir ?
1SG OBJ.2PL bring ART box PREP flyers DEM evening

‘I bring you the box of flyers tonight?’

(*Un avion sans elle*)

b. Taiwan Mandarin translation

我 今天 晚上 送 傳單 去
wǒ jīntiān wǎnshàng sòng chuándān qù
 1SG today evening deliver flyer go

你 家 如何?
nǐ jiā rúhé?
 2SG home how

‘How about I bring the flyers to your place tonight?’

(沒有她的飛機 *Méiyǒu tā de fēijī*)

Other translations include declarative sentences, as in (5.21), and merged questions, as in (5.22).

(5.21) a. Original French

Ce livre va peut-être au-delà de
 DEM book go maybe beyond PREP

ce qu' on peut y lire ?
 DEM REL 3SG can there read

‘Perhaps this book goes beyond what one can read in it?’

(*Gataca*)

b. Taiwan Mandarin translation

這 本 書 代表 的 意義 也許 比
zhè běn shū dàibiǎo de yìyì yěxǔ bǐ
 this CL book represent DE meaning maybe than

我們 能 理解 的 還 深。
wǒmen néng lǐjiě de hái shēn.
 1PL can understand DE more deep

‘This book may mean more deeper than we can understand.’

(左撇子的殺意 *Zuǒpǎizi de shāyì*)

(5.22) a. Original French

Deux aspirants ensemble ?
 two trainees together

Alors que la mission est danger niveau quatre ?
 while COMP ART mission COP danger level four

‘Two interns together? When the mission is danger level four?’ (*L'Anomalie*)

b. Taiwan Mandarin translation

兩位 實習生 一起 處理 危險
liǎng wèi shíxíshēng yìqǐ chǔlǐ wéixiǎn
 two CL intern together handle danger

等級 四 的 任務？
děngjí sì de rènwù?
 level four DE mission

‘Two interns together handling a level-four danger mission?’ (異常 *Yicháng*)

Table 87 presents the translation patterns of *wh*-questions from French into Taiwan Mandarin.

Table 87 Translation patterns for *wh*-questions from French to Taiwan Mandarin

Original (French)	Raw frequencies	Translation (Taiwan Mandarin)	Raw frequencies (%)
<i>Wh</i> -question	1,793 (100%)	Polar question	61 (3.4%)
		<i>Wh</i>-question	1,684 (93.9%)
		Disjunctive question	8 (0.5%)
		Other	40 (2.2%)

Over 90% of *wh*-questions are translated using the same question type, as shown in (5.23).

(5.23) **a. Original French**

Qu' allez-vous faire, maintenant ?
 what go-2SG do now
 ‘What are you going to do now?’

(*Un avion sans elle*)

b. Taiwan Mandarin translation

現在, 你 打算 怎麼 辦？
xiànzài nǐ dǎsuàn zěnmē bàn?
 now 2SG plan how do

‘Now, what are you planning to do?’ (沒有她的飛機 *Méiyǒu tā de fēijī*)

Only a small number of instances are translated with other question types in Taiwan Mandarin, including polar questions in (5.24), and disjunctive questions in (5.25).

(5.24) a. **Original French**

Qu' est-ce qu' elle montre, cette photographie
what PTCL 3SG show DEM photograph
'What does it show, this photograph?' (Un avion sans elle)

b. **Taiwan Mandarin translation**

照片 上 有 什麼 特徵 嗎?
zhàopiàn shàng yǒu shéme tèzhēng ma?
photo PREP have what feature PTCL
'Is there any feature on the photo?' (沒有她的飛機 *Méiyǒu tā de fēijī*)

(5.25) a. **Original French**

Comment va ta schizophrénie ?
how go POSS.2SG schizophrenia
'How is your schizophrenia?' (Gataca)

b. **Taiwan Mandarin translation**

你 的 精神分裂症 好 點 沒?
nǐ de jīngshénfēnlièzhèng hǎo diǎn méi
2SG DE schizophrenia better bit NEG
'Is your schizophrenia any better or not?' (左撇子的殺意 *Zuǒpěizi de shāyì*)

The 'other' category accounts for a minor proportion of the data. In this category, French *wh*-questions are translated with declarative sentences, as in (5.26), and are merged into a single question, as in (5.27).

(5.26) a. **Original French**

Pourquoi n' aurait-il pas agi de même
why NEG have.COND-3SG NEG act.PTCP PREP same
avec les autres mères ?
PREP ART other mothers
'Why wouldn't he have acted the same with the other mothers?' (Gataca)

b. **Taiwan Mandarin translation**

且 難保 他 不 會 對 其他 母親
qiě nánbǎo tā bú huì duì qítā mǔqīn
moreover not.sure 3SG NEG will to other mother
也 做出 同樣 的 事 來。
yě zuòchū tóngyàng de shì lái.
also do same DE thing come
'And there is no guarantee that he will not do the same to his other mothers.'
(左撇子的殺意 *Zuǒpěizi de shāyì*)

(5.27) a. **Original French**

Napoléon Chimaux ? Qui est-ce?
 Napoléon Chimaux who COP-DEM
 ‘Napoleon Chimaux? Who is that?’ (Gataca)

b. **Taiwan Mandarin translation**

西莫 是 誰 ?
Xīmò shì shéi?
 Ximo COP who
 ‘Who is Chimaux?’ (左撇子的殺意 *Zuǒpěizi de shāyì*)

Finally, Table 88 presents the translation patterns of **disjunctive questions** from French into Taiwan Mandarin.

Table 88 Translation patterns for disjunctive questions from French to Taiwan Mandarin

Original (French)	Raw frequencies	Translation (Taiwan Mandarin)	Raw frequencies
Disjunctive question	75 (100%)	Polar question	6 (8%)
		<i>Wh</i> -question	1 (1%)
		Disjunctive question	65 (87%)
		Other	3 (4%)

As with the translation patterns found in the other question types, the majority of disjunctive questions are translated using the same question type, as shown in (5.28).

(5.28) a. **Original French**

À votre avis, suis-je l' original
 PREP POSS.2SG opinion COP-1SG ART original
ou la copie ?
 or ART copy
 ‘In your opinion, am I the original or the copy?’ (L'Anomalie)

b. Taiwan Mandarin translation

在您的看來, 我 是 正牌 還是 複製人?
zàinínkànlái wǒ shì zhèngpái háishì fùzhìrén?
 in.your.opinion 1SG COP original or clone
 ‘In your opinion, am I a real one or a copy?’ (異常 *Yìcháng*)

Translations involving other question types are relatively rare, including polar questions in (5.29), and *wh*-questions in (5.30).

(5.29) a. Original French

T' es bouché ou quoi ?
 2SG COP blocked or what
 ‘Are you clogged or what?’ (*Un avion sans elle*)

b. Taiwan Mandarin translation

你 是 白痴 嗎?
nǐ shì báichī ma
 2SG COP idiot PTCL
 ‘Are you an idiot?’ (沒有她的飛機 *Méiyǒu tā de fēijī*)

(5.30) a. Original French

Un projet, des expériences, un programme de recherche ?
 ART project ART experiments ART program PREP research
 ‘A project, experiments, a research program?’ (*Gataca*)

b. Taiwan Mandarin translation

是 什麼 方案、 實驗、 研究計畫?
shì shénme fāng'àn shíyàn yánjiūjìhuà?
 COP what project experiment research.plan
 ‘What is the program, experiment, research project?’
 (左撇子的殺意 *Zuǒpǎizi de shāyì*)

The ‘other’ category accounts for only a minimal portion of the data. In these cases, disjunctive questions are translated as declarative sentences in (5.31) or are merged into a single question in (5.32).

(5.31) a. **Original French**

<i>A-t-il</i>		<i>utilisé</i>		<i>un</i>	<i>prétexte</i>	<i>pour</i>	<i>les</i>	<i>aborder,</i>
have-t-3SG		use.PTCP		ART	pretext	PREP	OBJ.3PL	approach
<i>ou</i>	<i>s'</i>	<i>est-il</i>		<i>précipité</i>	<i>sur</i>		<i>eux ?</i>	
or	REFL	COP-3SG		rush.PTCP	PREP		OBJ.3PL	
<i>À</i>	<i>ce</i>	<i>stade, la</i>	<i>SR</i>	<i>est</i>	<i>dans</i>	<i>le</i>	<i>flou.</i>	
PREP	DEM	stage	ART	SR	COP	PREP	ART	unclear

‘Did he use an excuse to approach them, or did he rush at them? At this stage, the SR is unclear.’

(*Gataca*)

b. **Taiwan Mandarin translation**

他	有-沒-有	用	藉口	靠近	被害人,
<i>tā</i>	<i>yǒu-méi-yǒu</i>	<i>yòng</i>	<i>jièkǒu</i>	<i>kàojìn</i>	<i>bèihàirén</i>
3SG	have-NEG-have	use	excuse	approach	victim
還是	突然	快速	朝	他們	走去,
<i>háishì</i>	<i>túrán</i>	<i>kuàisù</i>	<i>cháo</i>	<i>tāmen</i>	<i>zǒuqù</i>
or	suddenly	quickly	toward	3PL.OBJ	walk.toward
目前	偵查組	還	不	清楚。	
<i>mùqián</i>	<i>zhēnchá zǔ</i>	<i>hái</i>	<i>bù</i>	<i>qīngchǔ</i>	
currently	investigation.team	still	NEG	clear	

‘Did he have an excuse to approach the victims, or did he suddenly and quickly walk towards them, it is not clear to the investigation team.’

(左撇子的殺意 *Zuǒpěizi de shāyì*)

(5.32) a. **Original French**

<i>Et</i>	<i>s'</i>	<i>il</i>	<i>en</i>	<i>parlait</i>	<i>à</i>	<i>son</i>	<i>père ?</i>
and	if	3SG	that	speak.IMPART	PREP	POSS.3SG	father
<i>Ou</i>	<i>aux</i>	<i>gendarmes ?</i>					
or	PREP	police					

‘What if he told his father? Or the police?’

(*Trois jours et une vie*)

b. **Taiwan Mandarin translation**

他	會	跟	他	爸爸	還是	巡警	說	嗎?
<i>tā</i>	<i>huì</i>	<i>gēn</i>	<i>tā</i>	<i>bàba</i>	<i>háishì</i>	<i>xúnjǐng</i>	<i>shuō</i>	<i>ma ?</i>
3SG	will	PREP	3SG	father	or	police	tell	PTCL

‘Will he tell his father or the police?’

(三天一生 *Sāntiān yìshēng*)

This subsection has shown that **nearly 90% of French questions are translated into Taiwan Mandarin with the same question type**, while approximately 10% are rendered differently. Among the three question types, *wh*-questions and disjunctive questions tend to retain their original type after translation. However, **a notable exception**

is observed in the case of French polar questions, almost 12% of which are translated with disjunctive questions in Taiwan Mandarin. More than half of these translations use the A-not-A structure. This pattern suggests that certain French polar questions may implicitly encode disjunctive questions, since polar questions can be analysed as a reduced form of disjunctive questions in which the negative alternative (i.e., *ou pas* ‘or not’) remains covert. When such French polar questions are translated into Taiwan Mandarin, this covert negative option becomes overt through the use of A-not-A structure, thereby making the underlying disjunctive structure explicit. Nevertheless, when compared with the degree of translation bias observed in the direction, namely, the tendency for Mandarin disjunctive questions to be translated as French polar questions (approximately 70%), the overall proportion remains relatively low. Therefore, **the data do not provide sufficient evidence to support the second part of the hypothesis**, which predicts that French polar questions tend to be translated with disjunctive questions in Taiwan Mandarin. In addition to this asymmetry, it is noteworthy that 8% of the French disjunctive questions ($n = 6$) are translated with polar questions in Taiwan Mandarin. Most of these cases arise from the ambiguity of French disjunctive connectives, whose interpretation can be either inclusive or exclusive. This semantic ambiguity may lead translators to select polar questions in Taiwan Mandarin when the disjunction does not need to be made explicit. In addition, one of these mismatches appears to reflect meaning-based rather than form-based translation choices, as illustrated in (5.29).

The next section broadens the discussion by integrating findings from the bidirectional parallel analysis of question types.

5.2.3 Discussion on bidirectional parallel analysis of question types

Following the separate analyses of translation correspondences in each direction, this subsection provides a more comprehensive account by examining translation bias and calculating mutual correspondence (Altenberg 1999: 254). As outlined at the beginning of this chapter, both measures assess the extent of correspondence of one item between source and target languages, in this case, from Taiwan Mandarin to French and from French to Taiwan Mandarin. The key distinction between the two lies in directionality: **translation bias** captures unidirectional correspondences, whereas the **mutual correspondence** reflects bidirectional relationships. Using the formulas presented in

Figure 34 and Figure 35, the results for each question type are summarised in the following tables: Table 89 for polar questions, Table 90 for *wh*-questions, and Table 91 for disjunctive questions.

First, the **mutual correspondence and translation bias in polar questions are highly similar in both directions**, as shown in Table 89. This indicates that polar questions are generally translated with the same question type, regardless of the translation direction.

Table 89 Mutual correspondence and translation bias of polar questions

Mutual correspondence			
Taiwan Mandarin → French			
French → Taiwan Mandarin			
$\frac{(1493+1948) \times 100}{(1756+2328)} = 84\%$			
Translation bias			
Taiwan Mandarin → French	%	French → Taiwan Mandarin	%
$\frac{1493 \times 100}{1756}$	85	$\frac{1948 \times 100}{2328}$	84

Secondly, in the case of *wh*-questions, the **mutual correspondence remains high**, suggesting that *wh*-questions are consistently translated with the same question type overall. Notably, *wh*-questions exhibit the highest mutual correspondence among the three question types. However, the **translation bias from Taiwan Mandarin to French is slightly lower** (81%) than in the reverse direction (93.9%).

Table 90 Mutual correspondence and translation bias of *wh*-questions

Mutual correspondence			
Taiwan Mandarin → French and French → Taiwan Mandarin			
$\frac{(1310+1684) \times 100}{(1617+1793)} = 87.8\%$			
Translation bias			
Taiwan Mandarin → French	%	French → Taiwan Mandarin	%
$\frac{1310 \times 100}{1617}$	81	$\frac{1684 \times 100}{1793}$	93.9

In contrast, **disjunctive questions display the lowest mutual correspondence (21.6%)** among all question types, as shown in Table 91. This low value can be partly attributed to the relatively small number of disjunctive questions in the French originals. More importantly, **a marked translation bias is observed in the translation of disjunctive questions**. The value of translation bias from Taiwan Mandarin to French is substantially lower (11%) than in the other direction (87%), indicating that most Mandarin disjunctive questions are not translated with the same type in French. One plausible explanation lies in the structural particularity of Mandarin disjunctive questions, especially A-not-A construction and VP-Neg. Since French lacks any morphosyntactic counterparts of these Mandarin structures, this structural mismatch leads to a pronounced asymmetry in translation bias across the two directions, particularly from Taiwan Mandarin to French. As a result, Mandarin disjunctive questions are often translated with polar questions in French. On the one hand, the French translators may not perceive the A-not-A construction as a disjunctive structure, given that French provides no formal cue corresponding to these Mandarin patterns. In addition, as shown in the contrastive analysis in Chapter 4, French generally favours the use of polar questions, while disjunctive questions are relatively rare in actual use. On the other hand, translators may intuitively rely on the fact that French polar questions can be analysed as a reduced form of disjunctive questions, with a negative alternative remaining covert. Taken together, these observations suggest that French polar questions may display a dual status, functioning as both polar and disjunctive questions at the same time. This line of reasoning is supported by the discussion in Section 2.3.2, where experimental results in French indicate that the language may lack clear morphosyntactic boundaries between polar and other question types, as well as by the analyses in Section 4.3 concerning the morphosyntactic overlap between French polar and disjunctive questions. Finally, the low level of translation bias in this context may also reflect that translators alternate between formal-based or meaning-based strategies, corresponding to the traditional distinction between literal (word-for-word) and free (sense-for-sense) translation (Munday, Pinto & Blakesley 2022: 26–28). As for the other direction, the higher translation bias from French to Taiwan Mandarin indicates that French disjunctive questions are more frequently translated with the same question type. Overall, these findings confirm that disjunctive

questions, especially from Taiwan Mandarin to French, are the least likely to be translated with the same form, due to structural asymmetries.

Table 91 Mutual correspondence and translation bias of disjunctive questions

Mutual correspondence			
Taiwan Mandarin → French and French → Taiwan Mandarin			
$\frac{(52+65) \times 100}{(466+75)} = 21.6\%$			
Translation bias			
Taiwan Mandarin → French	%	French → Taiwan Mandarin	%
$\frac{52 \times 100}{466}$	11	$\frac{65 \times 100}{75}$	87

The overall results of question type translation indicate that, **in both directions, questions are generally translated with the same type**. However, the proportions are not symmetrical: while 88% of French questions are translated with the same question types in Taiwan Mandarin, this rate drops to 74% in the other direction. This asymmetry is largely driven by a significant divergence in the translation of disjunctive questions from Taiwan Mandarin to French, as reflected in the both values of translation bias and mutual correspondence. In contrast, *wh*-questions exhibit the highest translation bias and mutual correspondence in both directions, pointing to a high degree of structural alignment between two languages. With regard to the hypothesis proposed for this parallel analysis, the findings offer only **partial confirmation**. Disjunctive questions in Taiwan Mandarin are frequently translated with polar questions in French, consistent with the prediction. However, only a limited proportion of French polar questions are translated with disjunctive questions in Taiwan Mandarin, offering only partial support for the expected pattern in the other direction. While morphosyntactic differences between the two languages certainly contribute to this asymmetry, they cannot fully account for the overall picture. In addition to cross-linguistic contrasts, whether consciously perceived by translators or not, the translation outcome is also shaped by the translation strategies adopted in translation process, such as direct vs. oblique translations (Vinay & Darbelnet 1963: 46). Moreover, translation is mediated by multiple agents, including revisors and editors, who play key roles in preparation and fashioning of translations (Munday, Pinto & Blakesley 2022: 201).

Building on this, the next subsection turns to a more fine-grained examination of a specific *wh*-construction, namely, *wh*-in-situ. Given its optional nature in French, the following analysis investigates how *wh*-in-situ is translated in both directions.

5.3 Bidirectional parallel analysis: *Wh*-in-situ

As discussed in Sections 2.4.3 and 2.4.4, *wh*-in-situ in Mandarin and French is argued to be licensed through distinct mechanisms. In Mandarin, *wh*-in-situ is the default configuration, explained either by covert movement under the quantificational theory (Huang 1982) or by a binding under the unselective binding theory (Aoun & Li 1993; Tsai 1994b). In French, by contrast, *wh*-in-situ is one of several available options. Theoretical accounts for this variation include covert movement (Bošković 1998), remnant movement (Boucher 2010a), binding relations (Boucher 2010b), and prosodic licensing (Cheng & Rooryck 2000). While determining the exact nature of syntactic movement is challenging within a corpus-based approach, it is still possible to examine translation correspondences through other overt morphosyntactic elements, such as particles.

According to the unselective binding theory, *wh*-in-situ constructions in Mandarin are licensed through a binding relation between an operator and a question word. In Mandarin, such an operator can be an overt *wh*-particle, such as 呢 *ne* (Cheng 1991; Tsai 1994b; Cheng & Rooryck 2000: 2; Huang, Y.-H. Audrey Li & Li 2009: 276). Since *wh*-in-situ in Mandarin can be licensed by particles, it is plausible to expect that when Mandarin *wh*-questions marked with an overt particle are translated, such a licensing mechanism of *wh*-in-situ can possibly be transferred to French translations. Namely, this transfer can also license the *wh*-in-situ in French and trigger the *wh*-questions to be *in-situ*. Conversely, it has been argued that *wh*-in-situ in French is licensed by intonation (Cheng & Rooryck 2000: 5–6). Given the absence of overt licensing elements in French, such as particle, it is expected that the French *wh*-in-situ is translated with *wh*-questions without particles in Taiwan Mandarin.

In the bilingual comparable analysis presented in Chapter 4, Section 4.4 examined the phenomenon of *wh*-in-situ, with particular attention to French, where this structure is optional, in contrast to its default usage in Mandarin. That section examined both the distribution of *wh*-in-situ across different question words and the morphosyntactic

constraints that may govern its usage. The analysis revealed that the French question words *que* ‘what’ and *où* ‘where’ are more frequently realised *in-situ*. Furthermore, *wh*-in-situ is more commonly observed in dialogue, which aligns with patterns found in spoken French. In terms of syntactic constraints, the data from both prior studies and the current corpus suggest that some previously assumed restrictions on *wh*-in-situ, such as its incompatibility with modal verbs or occurrence in embedded clauses, do not hold. These findings challenge traditional accounts and suggest greater flexibility in the syntactic licensing of *wh*-in-situ in French than previously assumed.

Following the results of the bilingual comparable analysis, this section turns to a bidirectional parallel examination of *wh*-in-situ, with the goal of identifying translation patterns between the two languages. As this is the first attempt to observe the cross-linguistic correspondences in *wh*-in-situ through parallel corpus, the analysis remains exploratory in nature. Two questions guide the investigation: **(i) what kind of structure of *wh*-questions in Taiwan Mandarin is translated into *wh*-in-situ in French? (ii) how French *wh*-in-situ, as analysed in Section 4.4, is translated into Taiwan Mandarin?** Based on the theoretical discussion presented earlier, **it is plausible that the presence of particles in Taiwan Mandarin may interact with the realization of *wh*-in-situ in French translations.** However, this possible link remains theoretical conjecture rather than an empirically established mechanism, and therefore no explicit hypothesis is formulated at this stage.

The analysis is organised as follows: Sections 5.3.1 and 5.3.2 examine translation from Taiwan Mandarin to French and from French to Taiwan Mandarin, respectively. Section 5.3.3 synthesises these findings and discusses the bidirectional correspondences revealed in parallel data. Finally, the present analysis focuses exclusively on the *wh*-questions with typical form (i.e., complete sentential form) in the source language and excludes fragment *wh*-questions (e.g., *wh*-word, noun phrases, *wh*-infinitive, see Section 3.4.2 for details). These fragment *wh*-questions are categorised as ‘other’ in the subsequent sections.

5.3.1 Translation of *wh*-questions from Taiwan Mandarin to French *wh*-in-situ

This section examines **how *wh*-questions with typical form in Taiwan Mandarin are translated into French.** Broadly speaking, *wh*-questions in Mandarin may appear

with or without a particle. The distribution of morphosyntactic structures among all such *wh*-questions ($n = 1,299$) is presented in Table 92.

Table 92 Distribution of morphosyntactic structures of *wh*-questions in Taiwan Mandarin (Typical forms)

No particle	1,184
Particle	115
Total	1,299

Among the 1,184 *wh*-questions without particle, only 953 of them are translated with *wh*-questions in French. The following discussion focuses exclusively on these cases of equivalent translation. These *wh*-questions employ a range of morphosyntactic structures in French, as summarised in Table 93.

Table 93 French morphosyntactic structures corresponding to *wh*-questions without particles in Taiwan Mandarin

Original (Taiwan Mandarin)	Raw frequencies (%)	Translation (French)	Raw frequencies	(%)
<i>Wh</i> -questions with No particle	953 (100%)	Inversion	453	47.53%
		Particle <i>est-ce que</i>	187	19.63%
		No inversion	148	15.53%
		In-situ	43	4.51%
		Other	122	12.8%
		Total	953	100%

The majority are translated using subject-verb inversion in French, as illustrated in (5.33), followed by constructions with the particle *est-ce que*, as in (5.34), and finally by structures without inversion, as in (5.35).

(5.33) a. **Original Taiwan Mandarin**

我 要 去 哪裡？
wǒ yào qù nǎlǐ?
 1SG want go where
 ‘Where do I want to go?’

b. French translation

Où dois-je aller à présent ?
 where must-1SG go.INF PREP present
 ‘Where must I go now?’ (泡沫戰爭 *Pào Mò Zhàn Zhēng*)

(5.34) **a. Original Taiwan Mandarin**

他 哪時 失蹤 的 ?
tā nǎshí shīzōng de?
 3SG when disappear DE
 ‘When did he disappear?’ (私家偵探 *Sī Jiā Zhēn Tàn*)

b. French translation

Quand est-ce qu' il a disparu?
 when PTCL 3SG have disappear.PTCP
 ‘When did he disappear?’ (*Rue du dragon couché*)

(5.35) **a. Original Taiwan Mandarin**

我 到底 跑 了 多久 ?
wǒ dàodǐ pǎo le duōjiǔ?
 1SG after.all run ASP how.long
 ‘How long did I run after all?’ (鬼地方 *Guǐ Dì Fāng*)

b. French translation

Combien de temps j' ai pu courir ?
 how.much PREP time 1SG have can.PTCP run.INF
 ‘How long was I able to run?’ (*Ghost Town*)

Although *wh*-in-situ in Mandarin is often analysed as being licensed by overt particles, the licensing may also occur through covert mechanisms (Cheng & Rooryck 2000: 2). This theoretical possibility offers an explanation for a small subset of cases in the corpus: approximately 5 % of *wh*-questions without particles in Taiwan Mandarin are translated into French using *in-situ* constructions, as shown in (5.36).

(5.36) **a. Original Taiwan Mandarin**

你們 需要 我 做 什麼 ?
nǐmen xūyào wǒ zuò shénme?
 2PL need 1SG do what
 ‘What do you need me to do?’ (私家偵探 *Sī Jiā Zhēn Tàn*)

b. French translation

Vous avez besoin de moi pour faire quoi?
 2PL have need PREP 1SG PREP do.INF what
 ‘What do you need me to do?’ (*Rue du dragon couché*)

Among the *wh*-questions in Taiwan Mandarin that are translated with *wh*-in-situ in French, the distribution of question words is presented in Table 94.

Table 94 Distribution of question words in Mandarin *wh*-questions translated with *wh*-in-situ

Question word	Raw frequencies
什麼 <i>shénme</i> ‘what’	12
哪裡 <i>nǎlǐ</i> ‘where’	9
怎麼 <i>zěnmē</i> ‘how’	7
誰 <i>shéi</i> ‘who’	5
什麼時候 <i>shénme shíhòu</i> ‘when’	4
多少/幾 <i>duōshǎo</i> ‘how much/many’	4
哪 <i>nǎ</i> ‘which one’	1
為什麼 <i>wèishénme</i> ‘why’	1
Total	43

This distribution closely aligns with the tendency observed in original French, discussed in Section 4.4, where certain question words occur more frequently *in-situ*. In other words, when Mandarin *wh*-questions are translated with *wh*-in-situ in French, the tendency of question words follows the same pattern in original French. A similar correspondence is found with respect to register. Given that *wh*-in-situ is characterised as a feature associated with spoken language (Adli 2006: 169–170; Coveney 2011: 130; Zimmermann & Kaiser 2019: 138), it is expected that these translations would predominantly appear in dialogue. This expectation is born out in the data: **most cases are found in dialogue** ($n = 37$), with only a small number occurs in narrative ($n = 6$). This register distribution mirrors the pattern observed in original French.

The category labelled ‘other’ in the tables refers to fragment *wh*-questions, such as *wh*-infinitive, as in (5.37), and a *wh*-word, as in (5.38), which fall outside the scope of typical sentential forms.

(5.37) a. **Original Taiwan Mandarin**

寫 故事 要 幹 什麼？
xiě gùshi yào gàn shénme?
 write story want do what

‘What (do you) want to do (by) writing a story?’ (複眼人 *Fù Yǎn Rén*)

b. **French translation**

Pourquoi écrire une histoire ?
 why write.INF ART story
 ‘Why write a story?’

(*L’homme aux yeux à facettes*)

(5.38) a. **Original Taiwan Mandarin**

為什麼 你 不 用？
wèishénme nǐ bú yòng?
 why 2SG NEG use
 ‘Why don’t you have to do it?’

(泡沫戰爭 *Pào Mò Zhàn Zhēng*)

b. **French translation**

Pourquoi pas toi ?
 why NEG 2SG
 ‘Why not you?’

(*La guerre des bulles*)

With regard to *wh*-questions containing particles in Taiwan Mandarin, only 101 instances are translated into French using the same question type. Although this number is relatively small, these questions are also translated with various morphosyntactic structures in French but not with *wh*-in-situ, as shown in Table 95.

Table 95 French morphosyntactic structures corresponding to *wh*-questions with particles in Taiwan Mandarin

Original (Taiwan Mandarin)	Raw frequencies (%)	Translation (French)	Raw frequencies	(%)
<i>Wh</i> -questions with Particle	101 (100%)	Inversion	46	45.54%
		Particle <i>est-ce que</i>	19	18.81%
		No inversion	18	17.82%
		Other	18	17.82%
		Total	101	100%

The most frequent strategy is inversion, accounting for over 45% of the cases, as illustrated in (5.39). This is followed by translations employing the *est-ce que* particle in (5.40) and those with no inversion, as shown in (5.41).

(5.39) a. **Original Taiwan Mandarin**

她 種 什麼 呢？
tā zhòng shénme ne?
3SG plant what PTCL
'What was she planting?'

(複眼人 *Fù Yǎn Rén*)

b. **French translation**

Que cultivait-elle ?
what grow.IMPART-3SG
'What was she cultivating?'

(*L'homme aux yeux à facettes*)

(5.40) a. **Original Taiwan Mandarin**

但 他 能 做 什麼 呢？
dàn tā néng zuò shénme ne?
but 3SG can do what PTCL
'But what could he do?'

(鬼地方 *Guǐ Dì Fāng*)

b. **French translation**

Mais qu' est-ce qu' il aurait pu faire ?
but what PTCL 3SG have-COND can.PTCP do.INF
'But what could he have done?'

(*Ghost Town*)

(5.41) a. **Original Taiwan Mandarin**

那 有 什麼 辦法 能 做到 呢？
nà yǒu shénme bànfǎ néng zuòdào ne?
then have what method can accomplish PTCL
'Then what method can make it work?'

(天橋上的魔術師 *Tiān Qiáo Shàng De Mó Shù Shī*)

b. **French translation**

Alors comment on fait pour y arriver ?
then how 3SG do PREP there arrive.INF
'So how do we manage to do it?'

(*Le Magicien sur la passerelle*)

As a result, it appears that **the findings do not conform with the expectation**. *Wh*-questions with particles in Taiwan Mandarin do not have any correspondence translated with *wh*-in-situ in French. In other words, **the licensing of *wh*-in-situ through particle does not transfer during the translation and trigger the use of *wh*-in-situ in French**. In contrast, it is the *wh*-questions without particle that are occasionally translated with *wh*-in-situ in French, albeit in less than 5% of cases. A more detailed discussion of these findings is provided in Section 5.3.3. The following section turns to the other direction of translation, examining how French *wh*-in-situ questions are translated in Taiwan Mandarin.

5.3.2 Translation of *wh*-in-situ from French to Taiwan Mandarin *wh*-questions

This section investigates **how French *wh*-in-situ questions are translated into Taiwan Mandarin**. Of the 114 *wh*-in-situ questions in original French, only 107 French *wh*-questions are translated into Taiwan Mandarin using the same question type. Consequently, the following discussion will focus primarily on these 107 *wh*-questions in original French. The distribution of their morphosyntactic structures in Taiwan Mandarin is summarised in Table 96.

Table 96 Taiwan Mandarin morphosyntactic structures corresponding to *wh*-in-situ questions in French

Original (French)	Raw frequencies (%)	Translation (Taiwan Mandarin)	Raw frequencies	%
<i>Wh</i> -questions with <i>In-situ</i>	107 (100%)	No particle	88	82.24%
		Particle	14	13.08%
		Other	5	4.67%

The most common structure involves **translations without the use of a particle**, as shown in (5.42), followed by the structure with a particle, as in (5.43).

(5.42) a. **Original French**

Vous en concluez quoi ?
 2SG of.it conclude what
 ‘What do you conclude from it?’ (L’Anomalie)

b. **Taiwan Mandarin translation**

所以 您 的 結論 是 什麼 ?
suǒyǐ nín de jiélùn shì shénme?
 so 2SG DE conclusion COP what (異常 Yícháng)
 ‘So, what is your conclusion?’

(5.43) a. **Original French**

Et entre vous deux, cela se passe comment ?
 and between 2PL two that REFL happen how
 ‘And how is it going between you two?’ (L’Anomalie)

b. Taiwan Mandarin translation

妳們 兩 人 之間 如何 呢 ?
nǐmen liǎng rén zhījiān rúhé ne?
 2PL two person between how PTCL

‘How are things between the two of you?’ (異常 *Yìcháng*)

Regarding the fourteen French *wh*-in-situ questions translated with a particle in Taiwan Mandarin, the particles consist of 呢 *ne* ($n = 5$), as in (5.43), and 啊 *a* ($n = 9$), as in (5.44).

(5.44) a. Original French

C' est qui cette fois ?
 DEM COP who DEM time
 ‘Who is it this time?’

(*Gataca*)

b. Taiwan Mandarin translation

這 次 是 誰 啊 ?
zhè cì shì shéi a?
 DEM time COP who PTCL
 ‘Who is it this time?’

(左撇子的殺意 *Zuǒpěizi de shāyì*)

The ‘other’ category includes non-canonical forms in translation, such as a single *wh*-word, as in (5.45), and noun phrase, as in (5.46).

(5.45) a. Original French

C' est pour quoi ?
 DEM COP PREP what
 ‘What is it for?’

(*Un avion sans elle*)

b. Taiwan Mandarin translation

什麼 事 ?
shénme shì?
 what matter
 ‘What’s the matter?’

(沒有她的飛機 *Méiyǒu tā de fēijī*)

(5.46) a. Original French

C' est quoi, ton nom ?
 DEM COP what POSS.2SG name
 ‘What’s your name?’

(*Gataca*)

b. Taiwan Mandarin translation

你 的 名字 ?
nǐ de míngzì?
 2SG DE name
 ‘Your name?’

(左撇子的殺意 *Zuǒpěizi de shāyì*)

These results align with the expectation that **French *wh*-in-situ questions are generally translated into Taiwan Mandarin without a particle**, as prosody is assumed to license *wh*-in-situ structures in French, leading the corresponding Mandarin translations to omit particles. Nonetheless, 13% of *wh*-in-situ questions are translated with particles, indicating that the absence of particles does not necessarily result from the translation of French *wh*-in-situ questions. The following section provides a synthesis of findings from both translation directions and discusses their broader implications.

5.3.3 Discussion on bidirectional parallel analysis of *wh*-in-situ

While *wh*-in-situ is an optional strategy in French, it is considered underlyingly obligatory in Mandarin. This bidirectional parallel analysis has therefore focused on identifying the types of morphosyntactic structures in Mandarin that correspond to *in-situ* constructions in French, as well as how French *wh*-in-situ is translated into Mandarin. In this context, the notion of mutual correspondence is not applicable as it is not possible to quantify the frequency of *wh*-in-situ in Mandarin, where it is the default configuration and not a marked syntactic option.

Through a detailed analysis of the morphosyntactic structures in both the original Mandarin and French data, the results reveal a clear pattern: **the expected relationship between particle and *wh*-in-situ is not as robust as anticipated**. While French *wh*-in-situ is more often translated into Taiwan Mandarin without particles, the presence of particles in Taiwan Mandarin does not appear to increase the likelihood of *wh*-in-situ in French. The following paragraphs explore possible explanations for these findings.

The expectation that the presence of a particle in *wh*-questions in Taiwan Mandarin consistently triggers the use of *wh*-in-situ in French translations is not supported, as the data show no instance in which a French *wh*-in-situ is translated from a Mandarin *wh*-questions containing a particle. This is despite previous claims that particles serve as evidence for the availability of *wh*-in-situ in Mandarin. Evidently, the syntactic conditioning provided by particles in Mandarin does not carry over in translation. Moreover, both data and theoretical accounts suggest that a covert particle can also license *wh*-in-situ in Mandarin (Cheng & Rooryck 2000). Therefore, the presence of an overt particle does not seem to be a necessary condition to *wh*-in-situ licensing in Taiwan Mandarin. Given that particles do not play an essential role in licensing the *wh*-in-situ in

Mandarin, they cannot be expected to exert any direct influence on French syntax in translation, particularly in triggering *wh*-in-situ. **This finding**, that the expectation fails, thus **underscores that the *wh*-in-situ works differently across the two language systems**. In French, two main factors have been identified in the studies as influencing the use of *wh*-in-situ: the nature of the question words and the discourse context. Certain question words, such as *que* ‘what’ and *où* ‘where’, are more likely to be used *in-situ* (Druetta 2018: 43), a tendency also observed in both parallel (see Section 5.3.1) and comparable analyses (see Section 4.4). The second factor relates to discursive context: *wh*-in-situ in French is more likely to appear in strongly presupposed contexts (Boucher 2010b: 116, 130; Coveney 2011: 125). However, since the present bidirectional parallel corpus does not include discursive level coding, this aspect cannot be empirically verified within the current framework.

When examining the translation direction from French to Taiwan Mandarin, the expectation somewhat holds. French *wh*-in-situ questions are most frequently translated into Taiwan Mandarin with the *wh*-questions without particles than with particles. Since *wh*-in-situ in French is licensed by prosody, specifically, rising intonation (Cheng & Rooryck 2000; Boucher 2010b: 123), this result suggests that **prosodic licensing in the source language may influence the morphosyntactic realization in the target language**. As a consequence, *wh*-questions Taiwan Mandarin translations tend not to feature overt morphosyntactic marking, such as particles. Nevertheless, some translated *wh*-questions in Taiwan Mandarin are marked with particles. Given the optional nature of particles in *wh*-questions in Mandarin, their presence may be driven by contextual needs, such as enhancing discourse coherence. Alternatively, it is also possible that linguistic markedness is being transferred from French into Mandarin. In French, *wh*-in-situ is generally considered a marked structure relative to subject-verb inversion, as its usage is often constrained by discourse conditions. Similarly, while particles in Mandarin are not syntactically required, their presence adds a level of interpretive specificity, thus rendering them marked as well. Accordingly, the marked syntactic nature of French *wh*-in-situ may influence the appearance of particles in the Mandarin translations. In this sense, markedness in the source language may find a parallel expression through similarly marked structures in the target language.

To conclude, although *wh*-in-situ in Mandarin and French share surface similarities and, to some extent, comparable syntactic mechanism (e.g., movement or binding relation), **the licensing conditions differ significantly**. In Mandarin, *wh*-in-situ is licensed through overt or covert particles; while in French, it is licensed by prosodic cues, discourse context, and additional factors, such as lexical identity of the verb (Fließbach & Rockstroh 2024). This structural resemblance belies deeper syntactic and pragmatic differences. As a result, cross-linguistic link strength (i.e., connectivity) (Halverson 2024: 28–29) of *wh*-in-situ construction might be weak. Notably, no instance of French *wh*-in-situ in the corpus is found to be translated from a Mandarin *wh*-question containing a particle. This may be due in part to the relatively low frequency of particle-marked *wh*-questions in the corpus, since particles are optional in Mandarin. Nonetheless, particles remain a key morphosyntactic strategy for marking questions in Mandarin, and they also appear as one of the strategies available in French. The following Section 5.4 will offer a more detailed parallel analysis of question particles in Mandarin and French.

5.4 Bidirectional parallel analysis: Question particles

In addition to *wh*-in-situ, another point of similarity between Mandarin and French lies in their use of particle in questions. Mandarin generally employs four particles in questions, while French makes use of three. Each particle exhibits distinct syntactic distributions and pragmatic functions, as discussed in Sections 2.5.1 for Mandarin and 2.5.3 for French. A preliminary cross-linguistic comparison is provided in Section 2.5.5, and the key contrasts are summarised in Table 97.

Table 97 Syntactic distribution and pragmatic functions of question particles in Mandarin and French

		Mandarin				French		
		<i>ma</i>	<i>ba</i>	<i>a</i>	<i>ne</i>	<i>est-ce que</i>	<i>ça</i>	<i>donc</i>
Syntactic distribution	Polar	✓	✓	✓	✗	✓	✗	✗
	<i>Wh</i>	✗	✗	✓	✓	✓	✓	✓
	Disjunctive	✗	✗	✓	✓	✓	✗	✗
Pragmatic function	Reinforce	✓				✓		
	Solicit agreement		✓					
	Reduce forcefulness			✓		✓		
	Enhance contextual coherence				✓		✓	
	Continuation				✓			✓

Mandarin and French share certain similarities in both distribution across question types and pragmatic functions. For example, the particle *a* in Mandarin and the *est-ce que* particle in French can both occur across three questions types. In terms of function, both *ma* particle in Mandarin and the *est-ce que* in French can serve to reinforce the interrogative force of a question. However, other particles tend to be restricted to specific question types and contribute unique pragmatic functions to questions. For instance, some particles function to indicate speaker attitude or mark discourse status, features that differ across the two languages. It has been argued that the kinds of functions expressed via particles in Asian languages (e.g., Mandarin) are more typically conveyed through prosody in Western languages (e.g., French) (Hancil, Post & Haselow 2015: 18). Accordingly, while surface-level similarities exist, particles in Mandarin and French generally do not have one-to-one correspondence, particularly in terms of pragmatic function and discourse usage.

This section investigates how particles are translated between Mandarin and French from a bidirectional parallel perspective. Based on the previous discussion and the results of bilingual comparable analyses, the following hypothesis can be formulated:

Hypothesis: Since particles in Mandarin often serve pragmatic or discourse marking functions that are typically conveyed through prosody in French, they are not necessarily

translated with overt particles. Instead, the particles in Mandarin tend to be translated with declarative word order in French (i.e., without subject-verb inversion), thereby relying on intonational cues. By contrast, French particles tend to be translated into Taiwan Mandarin using particles, as Mandarin particles not only constitute the dominant strategy for marking questions but also cover the full range of syntactic distribution and pragmatic functions of French particles, as shown in Table 97.

The following analysis will focus primarily on the particles that appear in typical question forms (i.e., questions with complete sentential form) in both source texts and their translations.

5.4.1 Translation of question particles from Taiwan Mandarin to French

A total of 957 particles are found in Mandarin questions with typical forms. The majority appear in polar questions ($n = 828$), followed by *wh*-questions ($n = 115$) and disjunctive questions ($n = 14$). The corresponding translations of these particles into French vary in morphosyntactic strategies, as summarised in Table 98. Of the 957 Mandarin particles, only 96 instances are translated with particles in French. **The vast majority are translated without particles**, suggesting a preference for prosodic or other syntactic encoding (e.g., inversion). Additionally, approximately 8% of Mandarin questions with particles are translated with declarative sentences, translation merging, and zero translation. They are categorised in ‘Other’.

Table 98 Overview of translating patterns for question particles, from Taiwan Mandarin to French

Structure in original texts (Taiwan Mandarin)	Structure in translations (French)	Raw frequencies (%)
957 (100%)	Particle	96 (10%)
	Non-particle	782 (81.7%)
	Other	79 (8.3%)

First, among the 96 instances where Mandarin particles are translated with particles in French, **the most frequently observed are the 嗎 *ma* in polar question**, as in (5.47),

the particle 呢 *ne* in *wh*-question, as in (5.48), and the particle 呢 *ne* in disjunctive question, as in (5.49).

(5.47) a. **Original Taiwan Mandarin**

我 有 放棄 她 嗎？
wǒ yǒu fàngqì tā ma?
 1SG have give.up 3SG PTCL
 ‘Did I give her up?’

(鬼地方 *Guǐ Dì Fāng*)

b. **French translation**

Est-ce que je l' ai abandonnée ?
 PTCL 1SG OBJ have give.up.PTCP
 ‘Did I give her up?’

(*Ghost Town*)

(5.48) a. **Original Taiwan Mandarin**

你們 想 說 什麼 呢？
nǐmen xiǎng shuō shénme ne?
 2PL want say what PTCL
 ‘What do you want to say?’

(泡沫戰爭 *Pào Mò Zhàn Zhēng*)

b. **French translation**

Qu' est-ce que vous voulez dire ?
 what PTCL 2SG want say
 ‘What do you want to say?’

(*La guerre des bulles*)

(5.49) a. **Original Taiwan Mandarin**

可 不 可以 看一下 大象 裝 呢？
kě bù kěyǐ kànyíxià dàxiàng zhuāng ne?
 can NEG can look elephant costume PTCL
 ‘Can or cannot take a look at the elephant costume?’

(天橋上的魔術師 *Tiān Qiáo Shàng De Mó Shù Shī*)

b. **French translation**

Est-ce qu' on pourrait d'abord voir
 PTCL 3SG can.COND first see

le déguisement d' éléphant ?
 ART costume PREP elephant

‘Could we first see the elephant costume?’

(*Le Magicien sur la passerelle*)

Other Mandarin particles, such as 啊 *a* and 吧 *ba*, are also occasionally translated with the French *est-ce que* particle, though their frequency is considerably lower compared to the particle 嗎 *ma*, as shown in Table 99.

Table 99 Distribution of Mandarin particles translated with French *est-ce que* particles

Particle	嗎 <i>ma</i>	呢 <i>ne</i>	啊 <i>a</i>	吧 <i>ba</i>	Total
Raw frequencies	70	17	5	3	95

Interestingly, there is one instance in which a Mandarin particle is translated with the particle *ça* in French. In this case, the question type shifts from a polar question in Mandarin to a *wh*-question in French, as illustrated in (5.50).

(5.50) a. **Original Taiwan Mandarin**

我 放 多 了 嗎?
wǒ fàng duō le ma?
 1SG put much ASP PTCL
 ‘Did I put too much?’ (炒飯狙擊手 *Chǎo Fàn Jū Jí Shǒu*)

b. **French translation**

Comment ça, je les sale trop ?
 how PTCL 1SG OBJ salt too.much
 ‘What do you mean, I salt them too much?’ (*Le sniper, son wok et son fusil*)

For the remaining 782 cases where Mandarin particles are not translated with particles in French, a variety of syntactic strategies are employed. These include declarative word order in polar questions, as in (5.51) ($n = 446$, 69%), inversion in *wh*-questions, as in (5.52) ($n = 46$, 56%), and inversion in disjunctive questions, as in (5.53) ($n = 2$). See Table 100.

(5.51) a. **Original Taiwan Mandarin**

公車 沒 有 到 商場 嗎?
gōngchē méi yǒu dào shāngchǎng ma?
 bus NEG have arrive mall PTCL
 ‘Doesn't the bus go to the mall?’
 (天橋上的魔術師 *Tiān Qiáo Shàng De Mó Shù Shī*)

b. **French translation**

Le bus ne va pas jusqu' au marché ?
 ART bus NEG go NEG until PREP market
 ‘Doesn't the bus go to the market?’ (*Le Magicien sur la passerelle*)

(5.52) a. **Original Taiwan Mandarin**

小弟 以前 是 住 哪 戶 呢?
xiǎodì yǐqián shì zhù nǎ hù ne?
 little.brother before COP live which CL.HOUSE PTCL
 ‘Where did the younger brother live before?’ (鬼地方 *Guǐ Dì Fāng*)

b. French translation

Lequel de ces appartements habitait-il ?
 which.one PREP DEM apartments live.IMPART-3SG
 ‘Which of these apartments did he live in?’ (Ghost Town)

(5.53) **a. Original Taiwan Mandarin**

那 算是 我 的 心 還是 他 的 心 呢 ?
nà suànshì wǒ de xīn háishi tā de xīn ne ?
 DEM count 1SG DE heart or 3SG DE heart PTCL
 ‘Would that count as my heart or his?’
 (天橋上的魔術師 *Tiān Qiáo Shàng De Mó Shù Shī*)

b. French translation

Était-ce mon cœur ou bien le sien ?
 COP.IMPART-DEM POSS.1SG heart or indeed ART POSS.3SG
 ‘Was it my heart or his?’ (Le Magicien sur la passerelle)

In addition to typical questions forms in French, some examples involve fragment questions, which are classified into ‘Other structures’ in Table 100, for example, a noun phrase in (5.54) and a tag in (5.55).

(5.54) **a. Original Taiwan Mandarin**

可以 吃 鍋貼 嗎 ?
kěyǐ chī guōtiē ma?
 can eat potsticker PTCL
 ‘Can (we) eat potstickers?’ (天橋上的魔術師 *Tiān Qiáo Shàng De Mó Shù Shī*)

b. French translation

Et des raviolis poêlés ?
 and ART dumplings pan-fried
 ‘(How about) potstickers?’ (Le Magicien sur la passerelle)

(5.55) **a. Original Taiwan Mandarin**

這樣 不 好 嗎 ?
zhèyàng bù hǎo ma?
 this.way NEG good PTCL
 ‘Isn’t it good?’ (天橋上的魔術師 *Tiān Qiáo Shàng De Mó Shù Shī*)

b. French translation

Pas mal, non ?
 NEG bad NEG
 ‘Not bad, right?’ (Le Magicien sur la passerelle)

Overall, when particles in Mandarin are translated into French, typical question structures are the most frequently observed. Less common are structures involving fragments

questions. These distributional patterns are summarised in Table 100 and visualised Figure 36. It should be noted that 48 original questions in Mandarin are translated into different question types in French; these are excluded from the summary data presented in the table and figure below.

Table 100 Non-particle translation by question types of Mandarin and French

Question types	Structures	Raw frequencies
Polar	Declarative order	446
	Inversion	120
	Other structures	84
Wh	Inversion	46
	No inversion	18
	Other structures	18
Disjunctive	Inversion	2
	Total	734

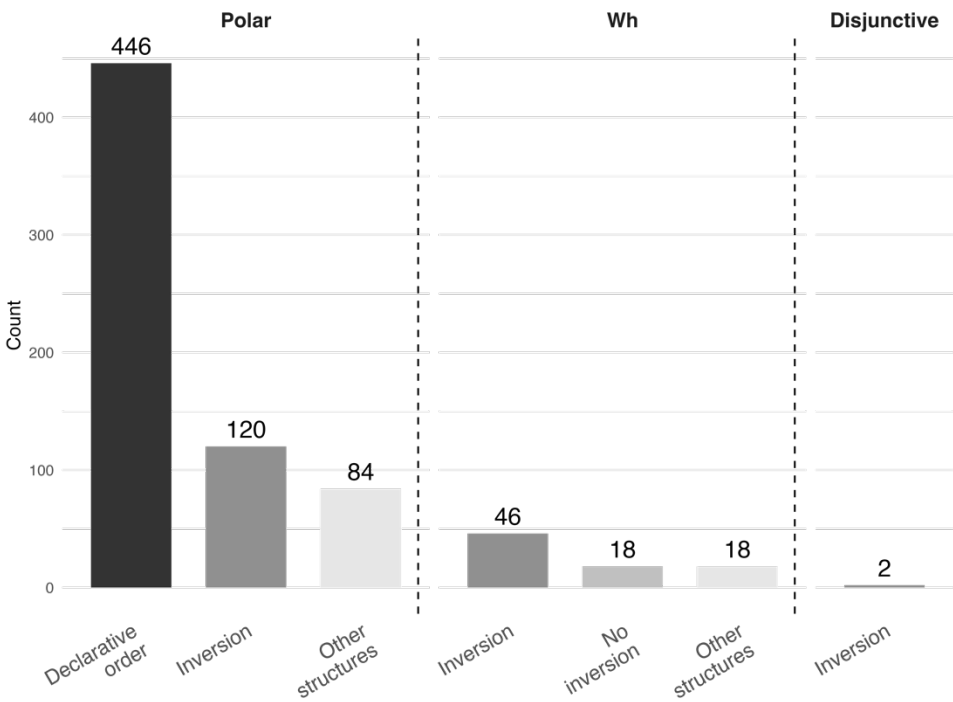


Figure 36 Particle translation by question types

Finally, Mandarin particles are also occasionally translated with declarative sentences, as in (5.56), translation merging, as in (5.57), and instance of zero translation.

(5.56) a. **Original Taiwan Mandarin**

你 怎麼 不 吃 自己 的 小狗 呢？
 nǐ zěnmē bù chī zìjǐ de xiǎogǒu ne?
 You how NEG eat self DE puppy PTCL
 ‘How come you don’t eat your own puppies?’ (泡沫戰爭 Pào Mò Zhàn Zhēng)

b. **French translation**

Tu aurais dû manger tes chiots.
 2SG have.COND must.PTCP eat.INF POSS.2SG puppies
 ‘You should have eaten your puppies.’ (La guerre des bulles)

(5.57) a. **Original Taiwan Mandarin**

我 是 誰 呢？ 我 是 誰 呢？
 wǒ shì shéi ne? Wǒ shì shéi ne?
 1SG COP who PTCL 1SG COP who PTCL
 ‘Who am I? Who am I?’ (複眼人 Fù Yǎn Rén)

b. **French translation**

Qui je suis ?
 who 1SG COP
 ‘Who am I?’ (L’homme aux yeux à facettes)

Following the analysis of particle translation from Taiwan Mandarin to French, the results indicate that **only approximately 10% of Mandarin question particles are translated with the *est-ce que* particle in French**. Among these, **the majority correspond to the particle 嗎 *ma* in the original Mandarin**. By contrast, over 80% of Mandarin particles are translated with other typical French question structures, such as declarative word order and inversion, both of which occur frequently across various question types. These findings suggest that the **question particles in Taiwan Mandarin do not map neatly onto the particles in French**, which explain the low rate of particle-to-particle translatability. The next section turns to another direction of translation, examining how French question particles are translated into Taiwan Mandarin.

5.4.2 Translation of question particles from French to Taiwan Mandarin

The use of particle *est-ce que* in original French texts is notably less frequent than the use of question particles in original Mandarin. Nevertheless, as shown in Table 101, **the majority of *est-ce que* particles are not translated with particles in Taiwan Mandarin** ($n = 171$). Only 27% of these instances ($n = 63$) are translated with a particle, and two cases are translated with declarative sentences, categorised as ‘Other’ in the table.

Table 101 Overview of translating patterns for question particles, from French to Taiwan Mandarin

Particle <i>est-ce que</i> in original texts (French)	Structure in translations (Taiwan Mandarin)	Raw frequencies (%)
236 (100%)		63
	Particle	(27%)
		171
	Non-particle	(72%)
	Other	2 (1%)

Among the 236 French particles, most occur in *wh*-questions ($n = 176$), as shown in (5.58), followed by polar questions ($n = 60$), as shown in (5.58). No cases are found in disjunctive questions.

(5.58) a. **Original French**

Qu' est-ce que tu vas imaginer ?
 what PTCL 2SG go imagine.INF
 ‘What are you going to imagine?’ (Le voleur d'ombres)

b. **Taiwan Mandarin translation**

你 在 想 什麼 啊?
nǐ zài xiǎng shénme a?
 2SG ASP think what PTCL
 ‘What are you thinking about?’ (偷影子的人 *Tōu yǐngzi de rén*)

(5.59) a. **Original French**

Est-ce que tu aimes dessiner ?
 PTCL 2SG like draw.INF
 ‘Do you like drawing?’ (L'Anomalie)

b. **Taiwan Mandarin translation**

妳 喜歡 畫畫 嗎?
nǐ xǐhuān huàhuà ma?
 2SG like draw PTCL
 ‘Do you like drawing?’ (異常 *Yìcháng*)

In the Mandarin translations, a range of particles are used, as shown in Table 102, and their distribution varies across question types.

Table 102 Mandarin particle correspondences in the translation of French *est-ce que*

Particle	嗎 <i>ma</i>	呢 <i>ne</i>	啊 <i>a</i>	吧 <i>ba</i>	Total
Raw frequencies	41	15	6	2	64

This suggests that **the translation of *est-ce que* particle into Mandarin is sensitive to the type of question**. While *wh*-questions account for the majority of the *est-ce que* particle uses in the French source texts, the most frequent Mandarin particle used in translation is 嗎 *ma*, which is primarily associated with polar questions. This is reflected in Table 102, which presents the overall distribution of particles in the Mandarin translations. Therefore, despite the predominance of particle use in French *wh*-questions, **one-to-one particle translations occur more frequently in polar questions**, indicating a structural and functional mismatch between the two languages in terms of particle use.

Most French particles are translated without any particle in Taiwan Mandarin. For example, polar questions are translated without particles in (5.60), and *wh*-questions in (5.61).

(5.60) a. **Original French**

Est-ce qu' il est vraiment... mort ?
 PTCL 3SG COP really dead
 'Is he really... dead?' (Trois jours et une vie)

b. **Taiwan Mandarin translation**

難道 他 真的 死 了 ?
nándào tā zhēnde sǐ le?
 don't.tell.me 3SG really die ASP
 'Could it be that he's really... dead?' (三天一生 *Sāntiān yìshēng*)

(5.61) a. **Original French**

D' où est-ce que ça vient ?
 PREP where PTCL DEM come
 'Where does that come from?' (Gataca)

b. **Taiwan Mandarin translation**

這 畫 打 哪裡 來 的 ?
zhè huà dǎ nǎlǐ lái de?
 DEM painting from where come DE
 'Where did this painting come from?' (左撇子的殺意 *Zuǒpěizi de shāyì*)

Additionally, fragment questions are also attested, such as *wh*-word, as shown in (5.62).

(5.62) a. **Original French**

Qu' est-ce qu' il y a ?
what PTCL 3SG there have
'What is going on?' (Gataca)

b. **Taiwan Mandarin translation**

怎麼 了 ?
zěnmě le?
how ASP
'What's wrong?' (左撇子的殺意 *Zuǒpěizi de shāyì*)

In addition to *est-ce que*, two other particles *ça* and *donc* are also examined. Both are structurally restricted to *wh*-questions, and only 18 instances are found in the bidirectional parallel corpus (17 with particle *ça* and one with particle *donc*). **None of these French particles are translated with particles in Taiwan Mandarin.** Thirteen of these are translated with a single *wh*-word, as shown in (5.63) for the particle *ça*, and the particle *donc* in (5.64). The remaining 5 instances are translated without any particle, as in (5.65).

(5.63) a. **Original French**

Où ça ?
where PTCL
'Where?' (Un avion sans elle)

b. **Taiwan Mandarin translation**

查 哪裡 ?
chá nǎlǐ
check where
'Where to check?' (沒有她的飛機 *Méiyǒu tā de fēijī*)

(5.64) a. **Original French**

Pourquoi donc?
why PTCL
'Why on earth?' (L'Anomalie)

b. **Taiwan Mandarin translation**

為什麼 ?
wèishénme?
why
'Why?' (異常 *Yìcháng*)

(5.65) a. **Original French**

Comment ça, depuis le temps que ça te ronge?
how PTCL since ART time COMP DEM OBJ.2SG bother
'What? Since when has it been bothering you?' (Un avion sans elle)

b. **Taiwan Mandarin translation**

什麼 叫 你 老早 就 良心 不安?
shénme jiào nǐ lǎozǎo jiù liángxīn bùān?
what mean 2SG long.ago already conscience uneasy
'What does it mean that your conscience has been uneasy for a long time already?'
(沒有她的飛機 Méiyǒu tā de fēijī)

The results of this section suggest that **direct correspondence between the particles in French and Mandarin is relatively rare**. In most cases, particles are not retained in French to Mandarin translation. However, when a French particle is translated with a particle in Taiwan Mandarin, it typically originates from polar questions involving *est-ce que*. Lastly, the translation of *ça* and *donc* show no particle correspondence at all, likely due to their limited distribution and low frequency in the corpus. In the following section, I will provide an overview of particles translation in both directions, with a particular focus on translation bias and mutual correspondence.

5.4.3 Discussion on bidirectional parallel analysis of question particles

The two preceding sections have examined how particles are translated bidirectionally between Taiwan Mandarin and French. The results indicate a general lack of systematic one-to-one correspondence. In other words, particles are rarely translated as such in the target language, regardless of the translation direction. To provide more quantitative support for this observed asymmetry, two measures are introduced: translation bias and the mutual correspondence. These values offer a more robust assessment of the degree of translational equivalence between the particles in the two languages.

Specifically, 957 particles are found in the source texts in Taiwan Mandarin, of which only 96 are translated with particles in French. In the opposite direction, 236 particles appear in the French source texts, but only 64 are translated with particles in Mandarin. These figures are summarised in Table 103.

Table 103 Translation bias and mutual correspondence of particles

Mutual correspondence			
Taiwan Mandarin → French			
French → Taiwan Mandarin			
$\frac{(96+64) \times 100}{(957+236)} = 13.4\%$			
Translation bias			
Taiwan Mandarin → French	%	French → Taiwan Mandarin	%
$\frac{96 \times 100}{957}$	10	$\frac{64 \times 100}{236}$	27

As shown by the mutual correspondence, **the overall translatability of particles between Taiwan Mandarin and French is very weak**: particles are rarely translated with particles in either direction. The translation bias confirms this tendency. In the direction from Taiwan Mandarin to French, Mandarin particles are unlikely to be translated into French with particles, which confirms the first part of the hypothesis: particles in Taiwan Mandarin are not generally translated with overt French particles, but rather with declarative word order in French (i.e., without subject-verb inversion), as demonstrated in Section 5.4.1. Conversely, given that French particles, primarily *est-ce que*, are often translated without particles in Taiwan Mandarin, the results do not support the second part of the hypothesis, namely that French particles tend to be translated with particles in Taiwan Mandarin. Despite this general lack of direct correspondence, **a light translation bias is noted: the value is higher in the direction from French to Taiwan Mandarin**, which indicates that *est-ce que* seems somewhat more likely to be translated with a particle in Mandarin than the reverse. Two possible explanations may account for this tendency: (1) Mandarin possesses a wider range of particles than French, and (2) Mandarin particles are frequently used to convey modality or speaker attitude. On the one hand, this discourse-pragmatic function of Mandarin particles may encourage translators to retain particle usage in the Taiwan Mandarin translations. On the other hand, because French typically conveys such functions through intonation, this typological difference may also lead translators to introduce a Mandarin particle in order to render nuances that are not overtly marked in the French source text. Importantly, however, the present study does not systematically analyse instances in which particles are added during translation,

and this remains an issue for future investigation. In sum, although both metrics point to a general lack of direct translatability between particles in the two languages, **two additional tendencies emerge when translation bias is examined by question type in each direction**. To explore these tendencies in more detail, I examine the frequency of particles in cases where translated questions retain the same question type from Taiwan Mandarin to French. The results are presented in Table 104.

Table 104 Particle raw frequencies and translation bias by question type from Taiwan Mandarin to French

Type	Taiwan Mandarin originals	French translations	Translation bias
Polar question	828	69	$\frac{69 \times 100}{828} = 8\%$
<i>Wh</i> -question	115	19	$\frac{19 \times 100}{115} = 17\%$
Disjunctive question	14	0	0%

Notably, most instances in which a particle appears in the French translation correspond to polar questions in the Taiwan Mandarin source texts. However, despite the high frequency of particles in polar questions in Taiwan Mandarin, the translation bias for this question type remains relatively low (8%) compared to *wh*-questions (17%). This suggests that **particles used in *wh*-questions in Taiwan Mandarin are more likely to be translated with *est-ce que* in French than those in polar questions**. This tendency may be attributed to syntactic and pragmatic similarities between the two languages. On the one hand, both the Mandarin particle 呢 *ne* in *wh*-questions and French *est-ce que* are highly optional syntactic elements. On the other hand, both serve to soften the illocutionary force of the question, often signaling politeness or a reduction in directness. These shared syntactic similarity and pragmatic functions may motivate the use of particles when translating Mandarin *wh*-questions containing particles into French.

In the opposite direction, I also examine cases where questions in French are translated into the same question type in Taiwan Mandarin. The results are presented in Table 105.

Table 105 Particle *est-ce que* raw frequencies and translation bias by question type from French to Taiwan Mandarin

Type	French originals	Taiwan Mandarin translations	Translation bias
Polar question	60	38	$\frac{38 \times 100}{60} = 63\%$
Wh-question	176	17	$\frac{17 \times 100}{176} = 9.7\%$

Although most particles in French source texts appear in *wh*-questions, they are not frequently translated with particles in the corresponding *wh*-questions in Taiwan Mandarin. In contrast, **particles in French polar questions are more likely to be translated with particles in Taiwan Mandarin**. The tendency is further supported by the value of translation bias across question types. The higher translation bias in polar questions indicates a strong likelihood of maintaining the particle in translation, whereas this is not observed for *wh*-questions. This asymmetry may be explained by the shared pragmatic function of particles in polar questions in both languages, namely, to reinforce the interrogative force. Such functional overlap may increase the likelihood of particle preservation during translation between polar questions.

To sum up, the mutual correspondence values indicate that **particles are unlikely to be directly translated as such in either direction between Taiwan Mandarin and French**. However, when each direction is examined individually, **the translation bias is higher from French into Taiwan Mandarin**. This can be attributed to the fact that Mandarin has a richer inventory of particles, each with specific syntactic distribution and pragmatic functions, many of which are typically expressed through intonation in French rather than morphosyntactic means. Interestingly, **when analysing particle translations by question types, an opposite trend emerges across the two directions**. From Taiwan Mandarin to French, the translation bias is higher for particles in *wh*-questions, while from French to Taiwan Mandarin, it is higher for particles in polar questions. This divergence appears to be closely tied to the pragmatic and typological nature of each language.

In the direction from Taiwan Mandarin to French, particles are more likely to be translated when both source and target particles are highly optional, such as 啊 *a* and 呢 *ne* in Taiwan Mandarin and *est-ce que* in French. Based on their theoretical functions, such

particles are expected to mitigate the forcefulness of the question and thereby enhance politeness in the translation. Conversely, from French to Mandarin, particles that serve to reinforce the interrogative force in the source text are likely to be translated with overt particle in Taiwan Mandarin, transferring this pragmatic emphasis. It should be noted that, however, the present corpus does not systematically annotate pragmatic functions. Thus, these conclusions are based on theoretical functional similarities rather than direct empirical evidence. These functional tendencies in particle translation may also imply that the direction of translation interacts with the socio-cultural communication style of the source language. As discussed in Section 4.3, Mandarin, as a language that tends toward indirectness in communication, tends to trigger the use of mitigating structures (e.g., the *est-ce que* particle) in French translations. In contrast, French, being more direct in socio-cultural interaction, appears to influence Mandarin translations toward more marked and reinforced question forms, especially through the use of the 嗎 *ma* particle.

5.5 Summary of bidirectional parallel analysis of questions in Taiwan Mandarin and French

Sections 5.1 to 5.4 have investigated the translation of questions between Taiwan Mandarin and French. The analysis focused on three main phenomena: question types, *wh*-in-situ constructions, and question particles. In the following summary, I present my key findings, drawing on evidence from the bidirectional parallel data analyses presented above.

First, the translation of **question types** appears to be generally stable across the two directions. **In most cases, the original question type is preserved in translation.** However, this cross-linguistic correspondence is more pronounced in the direction from French into Taiwan Mandarin (88%) than into the other direction (74.3%). The difference can be attributed partly to morphosyntactic divergence between the two languages, especially in disjunctive questions, and partly to the translation strategies adopted in translation. Given that the present study focuses on the structural level, a closer examination of individual question types reveals more nuanced patterns. For instance, a striking difference emerges across the two languages: nearly 70% of **disjunctive** questions in Taiwan Mandarin are translated with **polar** questions in French. Notably, many of these Mandarin disjunctive questions involve the A-not-A structures, which do not exist in French. Accordingly, the Mandarin A-not-A structures, a distinctive structure

without a direct counterpart in French, is frequently translated with French polar questions. This suggests that translators, possibly guided by functional considerations, choose to translate A-not-A disjunctive questions with French polar questions. Such translation choices are consistent with the theoretical claim that polar and disjunctive questions in French are syntactically overlapping. A similar tendency, though far less frequent, is observed in the other direction: over 10% of French polar questions are translated with disjunctive questions using A-not-A structure in Taiwan Mandarin. Overall, **the mutual translatability of disjunctive questions between the two languages is remarkably low, especially in the direction from Taiwan Mandarin to French.** Moreover, these findings partially confirm the hypothesis: disjunctive questions in Taiwan Mandarin are frequently translated with French polar questions. However, French polar questions are less frequently translated with disjunctive questions in Taiwan Mandarin.

Second, with regard to **wh-in-situ** structures, given that *wh*-in-situ is the default strategy in Mandarin but an option in French, the parallel analysis was exploratory and focused on how *wh*-in-situ is translated to and from Taiwan Mandarin. Based on the theoretical discussion, I aimed to explore a possible connection between the use of particle in Taiwan Mandarin and the choice of *wh*-in-situ in French. However, the results showed no evidence for such a relationship: the presence of particles in Taiwan Mandarin does not trigger the use of *wh*-in-situ in French. In the other direction, French *wh*-in-situ questions are often translated into Taiwan Mandarin without particles, but with particles to a lesser extent. Taken together, the results suggest that ***wh*-in-situ may be regulated independently within each language by the internal licensing system rather than by cross-linguistic influence.** Although the French translations of Mandarin questions exhibit a distribution pattern of *in-situ* questions words similar to that in original French, this appears to reflect that language-internal preference rather than an effect driven by the source texts in Taiwan Mandarin.

Third, drawing on the tentative comparison of **particles** presented in Section 2.5, I investigated how question particles are translated between Mandarin and French. Given that Mandarin often uses particle to express speaker modalities, whereas French relies more on prosody to convey similar meanings, it was hypothesised that particles would rarely be translated directly between the two languages. The findings, supported by a very

low mutual correspondence (13.4%), confirm this hypothesis: **particles in Mandarin and French are generally not mutually translated**, despite some overlaps in their syntactic distribution and pragmatic functions. Nonetheless, the analyses across question types revealed some **notable directional tendencies**. **From Taiwan Mandarin to French**, particles in *wh*-questions are more likely to be translated with particles, as reflected in a relatively higher translation bias (17%), compared with polar questions (8%). Conversely, in the direction **from French to Taiwan Mandarin**, particles in polar questions are more frequently translated with particles in Taiwan Mandarin, as indicated by a much higher translation bias (63%). This opposing trend in the two directions suggests that the communicative norms of the source language can influence the choice of using particles or not in the target language. Based on the tentative comparison of particles in terms of their syntactic and pragmatic distributions, I argue that the choice made in translating particles may be related not only to syntactic factors but also to the pragmatic functions shared across the two languages. In this sense, socio-cultural preferences regarding directness or politeness in communication may influence whether translators to use particles in the target language. It should be noted, however, that this remains a speculation grounded in theoretical reasoning and in structural tendencies observed in the translations, as the present study does not include a systematic coding of pragmatic functions.

This bidirectional parallel approach offered a deeper and more nuanced comparison between Mandarin and French than what can be achieved through bilingual comparable analysis alone. While the bilingual comparable approach highlighted surface-level similarities, the analysis of parallel data revealed fundamental structural and functional differences. Specifically, **the licensing mechanisms of *wh*-in-situ are not mutually transferable and particles are rarely translated across the two languages**. **In contrast, question types tend to be preserved in translation, with the notable exception of disjunctive questions from Taiwan Mandarin**, which are frequently rendered as polar questions in French. Importantly, however, the patterns observed in the parallel analyses do not only reflect morphosyntactic discrepancies between the two languages. They also suggest the influence of translator's cognitive selection process related to the connectivity of establishing translation pairs (Halverson 2024: 22). In the present study, although both Mandarin and French employ particles in questions, the connectivity between them is not

strongly entrenched. As a result, translators may not readily perceive particles as a natural translation pair (Halverson 2024: 28–29). In this sense, translational choices arise not only from structural differences but also from the varying degrees of cross-linguistic link activated during translation.

To account for these translation-induced influences, the final analysis in the following section adopts a monolingual comparable perspective. By comparing translated with non-translated language, this approach helps distinguish effects that stem from translation choices (e.g., source language interference and normalisation) from those that arise from inherent cross-linguistic discrepancies. As in previous sections, three topics will be examined under this framework: question types, *wh*-in-situ, and question particles.

5.6 Monolingual comparable analysis

The final section of this chapter adopts a monolingual comparable approach, focusing on **comparisons between non-translated and translated varieties**, specifically, original Taiwan Mandarin vs. translated Taiwan Mandarin, and original French vs. translated French. Building on the results of the parallel analyses presented above, it is clear that discrepancies in morphosyntactic structures across languages can influence translation outcomes, while individual translator choices also play a crucial role. Accordingly, a monolingual comparable analysis enables the identification of translation-specific effects by disentangling them from cross-linguistic structural differences, thereby complementing the patterns observed in translated texts. More specifically, the monolingual comparable approach aims to determine whether the phenomena observed in translated texts arise from the influence of the source text/language or rather from a tendency to adhere to target language norms, as represented by their non-translated counterparts. In contrast to the parallel approach, which focuses on specific choices, the monolingual comparable approach provides a more general and quantitative view of translated languages, capturing broader distributional patterns rather than individual instances (Bernardini 2022: 489). In the following, I will first review the features (e.g., explicitation, simplification) and effects of translation, including interference and normalisation.

5.6.1 Translation features: Interference and normalisation in translated language

The **monolingual comparable approach** has been central to corpus-based translation studies since its inception (Bernardini 2022: 489, Lefer 2025:2). In particular, following Baker (1993; 1996), translations can be conceptualised as essential communicative events worth studying in their own right. From this perspective, translated texts **should be analysed independently of their source texts** and systematically compared with non-translated texts in the same language. Building on this view, Baker (1993; 1996) proposes that translated texts have a set of recurrent features, which she refers to as ‘**translation universals**’ (now termed ‘**translation features**’ or ‘translation properties’, Lefer 2025:2). These are defined as “features which typically occur in translated text rather than original utterances and which are not the result of interference from specific linguistic systems” (Baker 1993:243). Four such features are identified: simplification, explicitation, normalisation, and levelling out (Baker 1996: 176–177). **Simplification** refers to the tendency to simplify the language used in translation, for instance by avoiding redundancy, reducing lexical scope and lexical density, and simplifying sentence structures (Hu & Zheng 2024: 222–224). **Explicitation** denotes the tendency to spell things out rather than leave them implicit in translation, such as through the addition of explanatory notes, discourse markers (e.g., *because*, *therefore*), and optional complementisers (e.g., *that* in English). As a result, text length in translations is usually longer than their originals (Baker 1996: 180). **Normalisation** captures the tendency to exaggerate features of the target language and to conform to its typical patterns, including grammatical and collocational structures and clichés (Baker 1996: 183). Finally, **levelling out** refers to the tendency of steering a middle course between any two extremes, converging towards the centre (Baker 1996: 184). In other words, translated texts are more homogeneous and convergent, while non-translated texts are more heterogeneous and divergent (François & Lefer 2022: 65, Hu & Zheng 2024: 227). Therefore, these features suggest that, in comparison to non-translated texts, translated texts tend to be simpler, more explicit, more conventional, and more homogeneous (Oster 2022: 8; Bernardini 2022: 492). By positing such regularities, the notion of translation universals/features establishes translated language as a legitimate object of study in its own right, with properties that are not reducible to source-language influence. However, this idea of excluding source-text influence from translations has been subject to

substantial critics (see e.g., Becher 2011: 42-43, Oster 2022: 9; Hu & Zheng 2024: 217). Notably, even Baker (1996: 183) acknowledges the possible influence from the source text in her discussion of normalisation. Consequently, Chesterman (2004: 39) proposed a distinction between **S-universals** (i.e., characteristics of the way in which translators process the source text) and **T-universals** (i.e., characteristics of the way translators use the target language when they translate). This distinction highlights that some universal features are more source-oriented, while others are more target-oriented. For example, explicitation has often been explored as an S-universal by comparing the frequency of specific elements between source and target texts as an indicator of the relative explicitness. In contrast, simplification is typically investigated as a T-universal, for instance through measures, such as lexical density and sentence length in translated texts vs. non-translated texts (Castagnoli & Kajzer-Wietrzny 2024: 4–5). Although these studies have contributed to our understanding of translation and offer a new perspective on it, **the idea of universality**, namely the assumption that all translations share a set of common features regardless of language, **has been called into question** (Pym 2008: 319). Empirical studies have shown that such tendencies are not consistently attested across languages. For example, Puurtinen (2004) demonstrates that explicitation is difficult to be maintained as a universal feature, as no significant difference is found in the frequency of connective use between translated and non-translated Finnish children’s literature. Similarly, after examining linguistic dimensions (e.g., type-token ratio, word length, sentence length, and keywords analysis) in translated Chinese, Wang & Qin (2010: 179) find that the hypotheses of translation universals are not fully supported. In particular, simplification is not observed, as translated Chinese uses more lexical types and longer segments than original Chinese, although explicitation of logical relations and co-reference devices is attested in Chinese translations. Xiao & Hu (2015) likewise re-evaluate the translation universal hypotheses from Chinese perspective. Although explicitation can be identified in translated Chinese at lexical and grammatical levels, translated Chinese may display greater complexity, as reflected in longer word length and sentence segment length in translated Chinese. Consequently, these findings cast doubt on the universality of such translation universals. Accordingly, the term *translation universals* has gradually been abandoned in favour of the more neutral notion of *translation features* (Granger & Lefer 2022:27).

Despite the controversial status of translation universals, the concept of translation features has attracted considerable attention and continues to be an important research strand in corpus-based translation studies, with explicitation in particular receiving sustained focus (Granger & Lefer 2022:27). However, this strong research focus may be disproportionate, as it stems from a very tentative list of suggestions for a starting-point of corpus-based translation investigation (Baker 1993: 243; see De Sutter & Lefer 2020). At the same time, several fundamental issues remain unresolved, such as the nature of explicitation, which remains difficult to account for, as it is still impossible to predict when translators will (or will not) use linguistic devices to encode explicitness (De Sutter & Lefer 2020: 2). Moreover, the proposed universals partially overlap with other theoretical accounts regarding translated language, in particular the Toury's laws of standardisation and interference (Pym 2008: 318). According to Toury, translated language is generally understood to be governed by two universal laws: the law of **interference** and the law of growing **standardisation** (Toury 1979; Toury 2012: 303). **The law of interference** concerns the influence of the source text, whereby linguistic features of the source language are transferred into the target language (Toury 2012: 310–311). In other words, the source language “shines through” the translation (Teich 2003: 145). This law is specific to translation and other language contact situations and does not apply to the non-translated (original) language (Lefer & Vogeleer 2013: 2–3). Conversely, **the law of growing standardisation** refers to the modification of textual relations found in the original text “in favour of [more] habitual options offered by a target repertoire” (Toury 2012: 304). This law of standardisation closely resembles *normalisation à la* Baker (1996: 176) and is widely regarded as a core feature of translated language, reflecting a tendency of translators to conform to the typical patterns of the target language. As noted by Pym (2008), normalisation largely overlaps conceptually with the law of standardisation. The other proposed translation features, such as explicitation and simplification, have also been criticized for downplaying the role of source-text interference and for treating these phenomena as separate tendencies, while they may in fact reflect different manifestations of normalisation in translated texts (Pym 2008: 318). In other words, an exclusive focus on translation features risks constraining the scope of investigation by privileging target-oriented explanations without systematically inspecting the potential effects of the source text. A comprehensive account of translation

therefore requires consideration of both explanatory mechanisms: the tendency towards the target language systems (i.e., law of standardisation or normalisation) and the influence of the source text (i.e., interference or source language influence). Importantly, these explanatory dimensions are not mutually exclusive. Explanations based on source-language interference are not necessarily opposed to those invoking translation features, since interference itself may be considered a fundamental law of translation (Mauranen 2004; Cappelle & Loock 2013: 253). Consequently, a growing body of corpus-based translation studies has approached translated language in terms of these two translation effects.

Pang & Wang (2020), for example, examine adversative conjunctions in the Chinese translations from English, including 雖然 *suīrán* ‘although’, 但是 *dànshì* ‘but’, 然而 *rán’ér* ‘however’. Despite the accent on diachronic variation, their results consistently show a higher occurrence of such conjunctions in translated Chinese than non-translated Chinese. This pattern is attributed to the interference from the source language (i.e., English). A similar effect has been observed for passive constructions. Their frequency in Chinese translated from English is higher than in non-translated Chinese, as a result of interference from the source language (Xiao, He & Yue 2010: 205). In addition to studies focusing on one translation direction, Cappelle & Loock (2013) adopt a bidirectional approach to examine existential constructions in French (i.e., *il y + avoir*) and in English (i.e., *there + be*). Given that existential constructions occur more frequently in original English, translated French exhibits an increased use of such structures compared to non-translated French, indicating source-language interference. Likewise, the same translation effect is attested in translated English, since the frequency of existential constructions is lower than in non-translated English. However, these effects are not symmetrical. While existential constructions increase in translated French compared to non-translated French, such an increase is weaker than the decrease observed in translated English. This asymmetry suggests that translated French also conforms to target-language norms, namely, a tendency towards normalisation, given that non-translated French typically uses fewer existential constructions (Cappelle & Loock 2013: 267). Adopting a similar approach, this thesis focuses on the two translation laws described above: interference and normalisation, to investigate bidirectionally the distribution of questions and three related issues, including question types, *wh*-in-situ,

and particles, across translated and non-translated varieties of the two languages investigated.

Despite the primarily linguistic focus of the present study, translation outcomes may also be affected by other crucial factors, such as register and translation expertise (De Sutter & Lefer 2020: 4). In this respect, translation can be viewed as one variety of constrained communication (Van Rooy & Kotze 2024: 275), as it is produced in language contact situations where one or more of the constraints or factors play a role in affecting communication. These constraints include (1) language activation (i.e., monolingual vs. bilingual), (2) modality and register (i.e., spoken vs. written vs. multimodal), (3) text production (i.e., independent vs. dependent), (4) proficiency (i.e., native vs. non-native), and (5) task expertise (i.e., expert vs. non-expert) (Kotze & Van Rooy 2024: 5). Such constraints model the outcome of the language production the same way as cognitive and socio-pragmatic processes, jointly shaping linguistic output. In other words, translation should not be understood merely as a product, but a production arising from the interaction of multiple factors. Although the translation analyses in this thesis do not directly implement the constrained language framework in examining question distribution, they do take into account key sources of variation. In particular, variation across translators and register (i.e., dialogue vs. narrative), operationalised here as random and fixed effects, respectively, in the statistical analysis presented in Section 5.11, following the recommendations of Halverson & Kotze (2024: 14).

This section is organised as follows. Section 5.6.2 presents the overall frequency of questions in translations, specifically in translated Taiwan Mandarin and translated French. Sections 5.7 through 5.9 provide detailed analyses of three main topics (i.e., question types, *wh*-in-situ, and question particles) to compare original and translated varieties in each language. Finally, Section 5.10 summarises the findings from the monolingual comparable analyses.

5.6.2 Overall frequency of questions in each translated language

The question marks were retrieved using the advanced Concordance function in Sketch Engine (searching for the character: ?) (Kilgariff et al. 2014). In principle, each question mark is assumed to correspond to a single question. However, one instance in the French sub-corpus contains two consecutive question marks for emphasis, as in (5.66).

(5.66) ...*Famille ?!?*
‘Family?!?’

(*Le sniper, son wok et son fusil*)

To maintain consistency, such cases with duplicated question marks are counted as a single occurrence of a question. The following analyses are based on these adjusted counts. Frequencies of questions in non-translated and translated languages is presented in Table 106 below.

Table 106 Frequencies of question in non-translated and translated languages

Languages	Type	Raw Frequencies	Relative Frequencies per 100,000 tokens	Corpus size (tokens)
Taiwan Mandarin	Non-translated	3,921	762	514,540
Taiwan Mandarin	Translated	4,433	737	601,139
French	Non-translated	4,224	653	647,180
French	Translated	4,188	600	697,504

As shown in Table 106, **the two translated varieties display lower relative frequency of questions compared to their non-translated counterparts**. This is likely due to the tendency of translated texts to expand through additional lexical or structural material, a phenomenon commonly associated with the explicitation universal in translation studies (Frankenberg-Garcia 2009). To quantify the expansion observed in the corpora of translated languages, a corpus-level coefficient of expansion was calculated by dividing the increase in corpus size of each language (translated minus non-translated corpus size) by the size of the non-translated corpus of each language. According to this measure, translated French exhibits a 7.8% of expansion, while translated Taiwan Mandarin shows a substantially larger expansion of 16.8%. **This expansion or increase in textual volume directly affects the relative frequency calculation, explaining the lower relative frequency of questions observed in translated corpora**. In addition, when comparing the difference of relative frequency, the difference of relative frequency is more pronounced in French, where the translated French shows 53 fewer questions per 100,000 tokens, compared to a reduction of 25 questions in Taiwan Mandarin. In proportional terms, translated French contains 8.1% fewer questions per 100,000 tokens than non-translated French, while the decrease in Taiwan Mandarin is relatively modest at 3.3%.

This reduction can be partly attributed to expansion mechanisms in both languages. However, expansion alone cannot explain the cross-linguistic difference in the degree of reduction. Such differences are likely related to the translation strategies employed. As shown in the parallel analysis of question types in Sections 5.2.1 and 5.2.2, changes in sentence type occur in 10% of questions in translated French, compared to only 2% of questions in translated Taiwan Mandarin. This further suggests that **the larger reduction in relative frequency observed in translated French may be related to more frequent alternations of sentence types during translation.** In addition to these overall trends, variation across individual translated works should also be taken into account. Examining the distribution of question frequency at the textual level can help determine whether the observed pattern reflects a broader translation tendency or whether it is more closely linked to specific textual characteristics, such as works and genres. The following tables provide language-specific comparisons of question frequency across individual translated works. Table 107 presents the data for translations into Taiwan Mandarin, and Table 108 for translations into French.

Table 107 Metadata and frequencies of question marks for novels in translated Taiwan Mandarin (Raw and relative frequencies)

Novels in translated Taiwan Mandarin

Translator	Work	Genre	Text size (number of tokens)	Raw frequencies of question mark (per 10,000 tokens)
Chang Chiao-wen	<i>Gataca</i> 左撇子的殺意 <i>Zuǒpěizi de shāyì</i>	Detective	148,862	1,458 (98)
Liang Sophie	<i>Un avion sans elle</i> 沒有她的飛機 <i>Méiyǒu tā de fēijī</i>	Detective	150,661	1,211 (80)
Chen Yung-Wei	<i>L'Anomalie</i> 異常 <i>Yìcháng</i>	Fantasy	90,810	659 (73)
Miao Yung-hua	<i>Trois jours et une vie</i> 三天一生 <i>Sāntiān yìshēng</i>	Detective	66,350	448 (68)
Duan Yun-ling	<i>Le voleur d'ombres</i> 偷影子的人 <i>Tōu yǐngzi de rén</i>	Fantasy	66,175	402 (61)
Wu Qing-zhu/ Duan Yung-ling/ Nian Geng-jia	<i>La vie des elfes</i> 精靈少女 <i>Jīnglíng shàonǚ</i>	Fantasy	78,288	255 (33)

Table 108 Metadata and frequencies of question marks for novels in translated French
(Raw and relative frequencies)

Novels in translated French

Translator	Work	Genre	Text size (number of tokens)	Raw frequencies of question mark (per 10,000 tokens)
Alexis Brossollet	炒飯狙擊手 <i>Chǎofàn jūjīshǒu</i> <i>Le sniper, son wok et son fusil</i>	Detective	97,615	1,031 (106)
Emmanuelle Péchenart	私家偵探 <i>Sījiā zhēntàn</i> <i>Rue du dragon couché</i>	Detective	159,545	1,219 (76)
Emmanuelle Péchenart	鬼地方 <i>Guǐdìfāng</i> <i>Ghost Town</i>	Fantasy	149,209	684 (46)
Gwennaël Gaffric	泡沫戰爭 <i>Pàomò zhànzhēng</i> <i>La guerre des bulles</i>	Fantasy	100,695	533 (53)
Gwennaël Gaffric	複眼人 <i>Fùyǎnrén</i> <i>L'homme aux yeux à facettes</i>	Fantasy	116,228	474 (41)
Gwennaël Gaffric	天橋上的魔術師 <i>Tiānqiáoshàng de móshùshī</i> <i>Le Magicien sur la passerelle</i>	Fantasy	74,270	247 (33)

Poisson regression models were applied using the *glmer* function (Bates et al. 2015) to examine the number of questions (i.e., dependent variable) across genres and individual works within each translated language. Genre was specified as a fixed effect, while work

was included as a random effect to account for variability among individual texts. In addition, to control for differences in text length, the total number of tokens per novel was incorporated as an offset in the model (i.e., $\text{offset} = \log(\text{Total_Translation})$), thereby adjusting the question counts to a comparable scale across texts. The results of the regression models are shown in Table 109 for Taiwan Mandarin translations and Table 110 for French translations.

Table 109 Poisson regression output for frequencies of question marks in translated Taiwan Mandarin

	Estimate	Standard Error	z value	p value
(Intercept)	-4.8151	0.1509	-31.917	< .0001 ***
Genre Fantasy	-0.4337	0.2144	-2.023	0.0431 *

Formula= `glmer(Count_translation ~ Genre+ (1 | Work), family = poisson, offset = log(Total_translation), control = glmerControl(optimizer = "bobyqa"), data = .)`

Table 110 Poisson regression output for frequencies of question marks in translated French

	Estimate	Standard Error	z value	p value
(Intercept)	-4.7141	0.1148	-41.047	< .0001 ***
Genre Fantasy	-0.7407	0.1419	-5.219	< .0001 ***

Formula= `glmer(Count_translation ~ Genre+ (1 | Work), family = poisson, offset = log(Total_translation), control = glmerControl(optimizer = "bobyqa"), data = .)`

The results of both regression models indicate that the distribution of question marks **differs significantly** across two genres in Taiwan Mandarin translations ($z = -2.023$, $p = 0.0431$) and in French translations ($z = -5.219$, $p < .0001$). More specifically, compared with detective fiction, **fantasy novels contain significantly fewer questions in both translated languages**, with a particularly strong effect observed in French. These findings suggest that genre consistently influences the distribution of questions in translated novels across the two languages. As for the random effect of work, the estimated variance is minimal in both models (0.067 for Taiwan Mandarin and 0.025 for French), indicating that **inter-textual variation among individual works contributes negligibly to the overall distribution of questions in translation**.

The analyses above demonstrate that the relative frequency of questions decreases in both translated languages compared with their non-translated languages. One possible explanation is that translated languages tend to use more lexical and syntactic material, a phenomenon referred to as ‘expansion’ in translation. This expansion can consequently reduce the relative density of questions, consistent with the explicitation hypothesis in translation studies. In addition to expansion, the baseline frequency in non-translated (original) language provides a reference point for the expected relative frequency in translation: since the relative frequency of questions in original Taiwan Mandarin is higher than in original French, one might expect translated French to retain a higher relative frequency if all questions were translated directly from Taiwan Mandarin. However, the larger reduction in relative frequency observed in translated French is likely due to translation strategies, particularly the tendency to alter sentence types during translation. In other words, **French translations more frequently modify sentence type, which contributes to a more pronounced reduction in the relative frequency of questions.** Additionally, the Poisson regression models of individual translated works reveal that the **genre is a significant factor influencing the distribution of questions**, with translated fantasy novels in the corpus containing fewer questions than translated detective fictions. Taken together, these findings suggest that **the observed decrease in relative frequency in translated language results primarily from a combination of textual expansion, translation strategies, and genre-specific tendencies**, while baseline difference in the original language helps contextualise the expected relative frequency in translation.

Having examined the overall frequency of questions in each translated language, the following sections (Sections 5.7 to 5.9) offer detailed monolingual comparable analyses focusing on three specific linguistic features (i.e., question types, *wh*-in-situ, and question particles). These analyses aim to investigate whether the translated languages exhibit **source-language interference** or **target-language normalisation**. Accordingly, each discussion will be framed around two hypotheses corresponding to these distinct translation effects. A total of 23 cases in translated Taiwan Mandarin and 64 cases in French are excluded from the following analyses, as they are non-qualifying cases (e.g., interjections and questions in other languages, see Section 3.4 on data coding).

5.7 Monolingual comparable analysis: Taxonomy of question types

The first topic investigated concerns the distribution of question types. Based on the results in Section 4.3 from the comparable corpus, polar questions are the most frequent types in both Taiwan Mandarin and French. However, a divergence emerges when comparing polar and disjunctive questions. On the one hand, polar questions occur much more frequently in original French; on the other hand, disjunctive questions are used more often in original Taiwan Mandarin, but not in original French. These tendencies observed in the source languages can ultimately influence their translations into the other language, as reflected in the parallel corpus results discussed in Section 5.2. Although most question types tend to be translated with the same types, the particularly low mutual correspondence value of disjunctive questions indicates that their translations rarely correspond across the two languages. Building on these findings, **two relevant hypotheses** can be proposed when comparing translated languages with non-translated languages. Under the source language interference hypothesis, translated Taiwan Mandarin is expected to contain fewer disjunctive questions, whereas translated French is expected to contain more. By contrast, from the normalisation perspective, the frequency of disjunctive questions would be expected to increase in translated Taiwan Mandarin but decrease in translated French. These two hypotheses are summarised in Table 111.

Table 111 Two hypotheses based on two translation effects for question types

Hypothesis	Pair of comparison	Translation effect	Frequency of disjunctive questions in translated language
I	Non-translated Taiwan Mandarin vs. Translated Taiwan Mandarin	Interference of source language	↓
	Non-translated French vs. Translated French		↑
II	Non-translated Taiwan Mandarin vs. Translated Taiwan Mandarin	Normalisation	↑
	Non-translated French vs. Translated French		↓

In what follows, Section 5.7.1 focuses on non-translated and translated Taiwan Mandarin, while Section 5.7.2 presents the results for French. Finally, an overall discussion is provided in Section 5.7.3.

5.7.1 Comparison of question types in non-translated and translated Taiwan Mandarin

This section begins with an analysis of the distribution of question types in non-translated and translated Taiwan Mandarin, as shown in Table 112 and Figure 37.

Table 112 Question types in non-translated and translated Taiwan Mandarin (raw frequencies)

Languages	Type	Polar question	<i>Wh</i> -question	Disjunctive question	Total
Taiwan Mandarin	Non-translated	1,756	1,617	466	3,839
		45.7%	42.1%	12.1%	100%
Taiwan Mandarin	Translated	2,133	1,897	380	4,410
		48.4%	43%	8.6%	100%

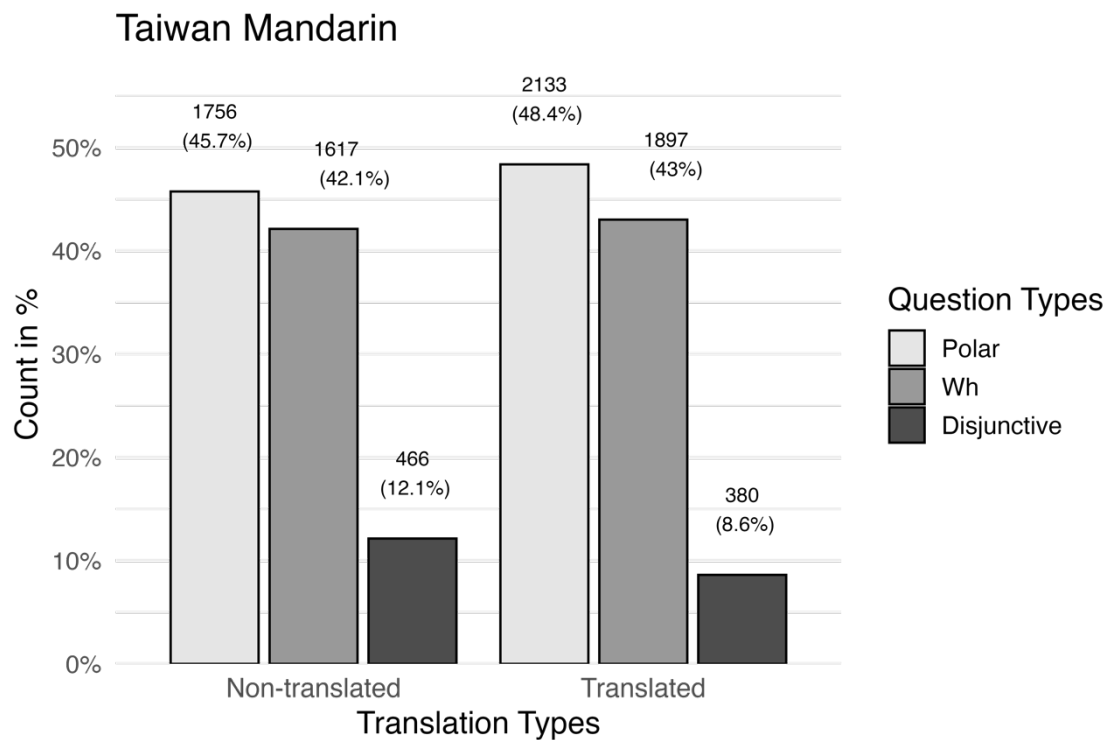


Figure 37 Question types in non-translated and translated Taiwan Mandarin

Results indicate that there is not much difference between the two varieties in terms of question type distribution. In Taiwan Mandarin, *wh*-questions occur at nearly the same proportion across both translation types. However, disjunctive questions appear more frequently in non-translated Taiwan Mandarin, showing a difference of 3.5%, respectively, whereas polar questions appear more frequently in translated Taiwan Mandarin, with a difference of 2.7%. The chi-square test further reveals a **significant difference** ($\chi^2(2) = 28.21, p < .0001$), **particularly for polar and disjunctive questions**, as reflected in the adjusted residuals presented in Table 113.

Table 113 Adjusted residuals of question type by translation status

Type	Non-translated	Translated
Polar	-2.38	+2.38
<i>Wh</i>	-0.82	+0.82
Disjunctive	+5.26	-5.26

Accordingly, translated Taiwan Mandarin shows an increase in polar questions and a greater decrease in disjunctive questions. These findings may be attributed to **interference** from French. On the one hand, disjunctive questions are not frequently used in French, leading to their underuse in translated Taiwan Mandarin. On the other hand,

the number of polar questions is significantly higher in original French, which in turns results in a higher frequency of polar questions in translated Taiwan Mandarin.

In sum, question types appear less divergent between non-translated and translated Taiwan Mandarin overall, yet evidence of **source-language interference is observed in the case of polar and disjunctive questions**. This “shining through” effect reduces the use of disjunctive questions while increasing the use of polar questions in translated Taiwan Mandarin.

5.7.2 Comparison of question types in non-translated and translated French

For non-translated and translated French, the distribution of question types is presented in Table 114 and in Figure 38.

Table 114 Question types in non-translated and translated French (raw frequencies)

Languages	Type	Polar question	<i>Wh</i> -question	Disjunctive question	Total
French	Non-translated	2,328	1,793	75	4,196
		55.5%	42.7%	1.8%	100%
French	Translated	2,334	1,692	98	4,124
		56.6%	41%	2.4%	100%

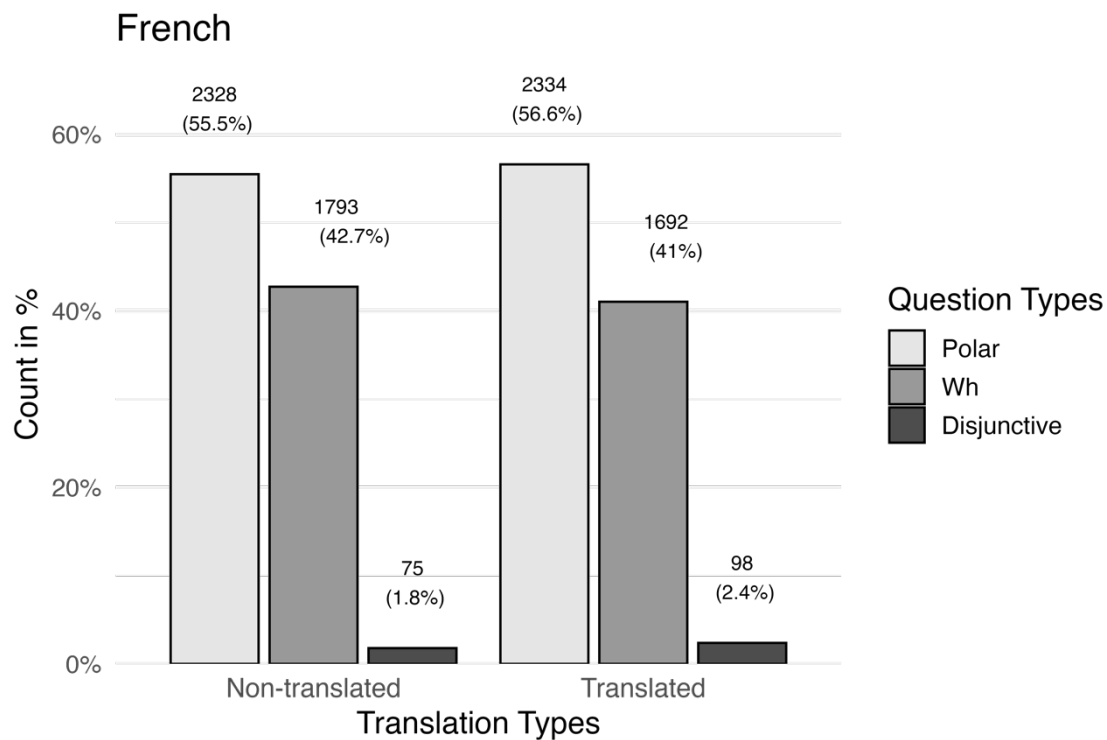


Figure 38 Question types in non-translated and translated French

Overall, the distribution across question types is highly similar between the two varieties, although disjunctive questions occur slightly more frequently in translated French. The chi-square test shows **no significant difference** between non-translated and translated French ($\chi^2 (2) = 5.37, p = 0.06822$). Therefore, **no clear translation effect** is observed.

5.7.3 Discussion of question types in non-translated and translated languages

Overall, the distribution of question types is largely similar between non-translated and translated languages. In other words, **translation does not appear to substantially alter the use of question types**. However, polar and disjunctive questions in translated Taiwan Mandarin constitute a notable exception. Compared with non-translated Taiwan Mandarin, translated Taiwan Mandarin contains significantly more polar questions but significantly fewer disjunctive questions. This pattern suggests **source language interference from French into Taiwan Mandarin**. By contrast, no translation effect is observed when comparing non-translated French and translated French. Accordingly, when examining question types across non-translated and translated languages, **only the first hypothesis is partially supported**.

5.8 Monolingual comparable analysis: *Wh*-in-situ

The topic of *wh*-in-situ mainly focuses on French, given that this syntactic structure is optional and can be readily observed and extracted from the French sub-corpus. In contrast, since *wh*-in-situ is the default structure in Mandarin, it is difficult to investigate this phenomenon in the corpus. Accordingly, **the following discussion primarily centres on the comparison of *wh*-in-situ constructions between non-translated and translated French.** Two hypotheses related to translation effects are formulated and presented in Table 115. Given that *wh*-in-situ is the default structure in Mandarin, its use in translated French is expected to increase, reflecting source language interference. By contrast, since *wh*-in-situ is only an optional structure and occurs rarely in original French, its frequency in translated French is expected to decrease, demonstrating normalisation toward non-translated French.

Table 115 Two hypotheses based on two translation effects for *wh*-in-situ

Pair of comparison	Hypothesis	Translation effect	Frequency of <i>wh</i> -in-situ in translated language
Non-translated French vs. Translated French	I	Interference of source language	↑
	II	Normalisation	↓

In addition to examining the overall translation effect, the following analyses and discussion address both the frequency of *wh*-in-situ occurrences and the types of question words that appear *in-situ*. Furthermore, the syntactic constraints argued to govern the use of *wh*-in-situ in translated French will also be examined.

5.8.1 Comparison of *wh*-in-situ in non-translated and translated French

The frequency of the *wh*-in-situ phenomenon in non-translated and translated French is presented in Table 116 and Figure 39.

Table 116 *Wh*-in-situ in non-translated and translated French

Language	Type	Raw frequencies	Relative frequencies (% of all <i>wh</i> -questions)
French	Non-translated	114	6.4%
	Translated	76	4.5%

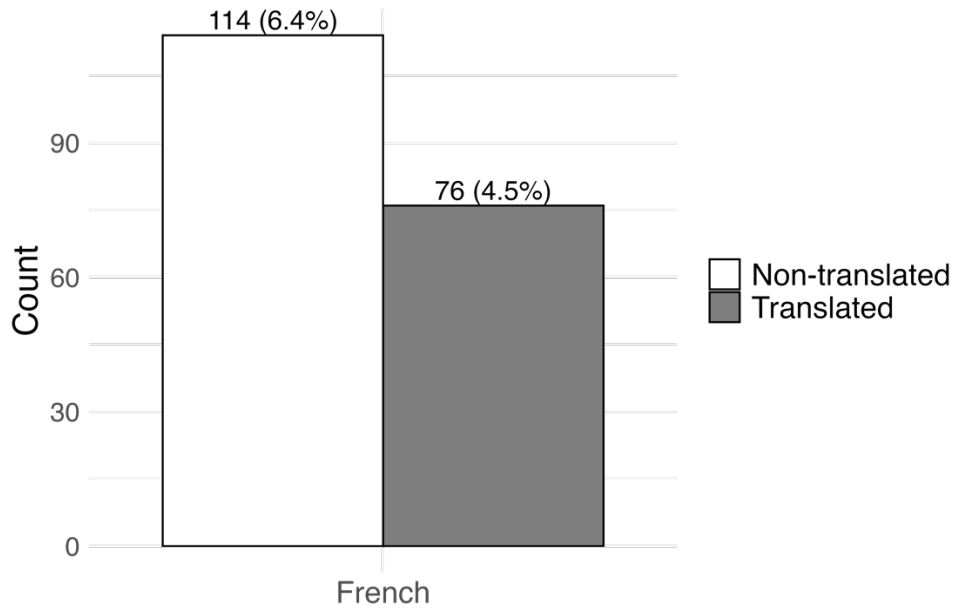


Figure 39 *Wh*-in-situ in non-translated and translated French

Although this syntactic structure is rare in both non-translated and translated French, a decrease in its use is observed in translated French. Only 4.5% of the *wh*-questions in translated French uses *wh*-in-situ, compared to 6.4% in non-translated French. This finding indicates a **normalisation** effect in translation. Given that translated French tends to conform to the linguistic norms of its non-translated counterpart, the following analysis examines whether the distribution of *in-situ* question words also aligns with that observed in non-translated French.

Based on previous studies and the comparable results for original (non-translated) French, the distribution of *in-situ* question words in non-translated French is shown in Figure 40. These results align with the patterns reported in Druetta (2018: 43) (see Section 4.4), namely that the question words *quoi* ‘what’ and *où* ‘where’ are more likely to occur *in-situ*.

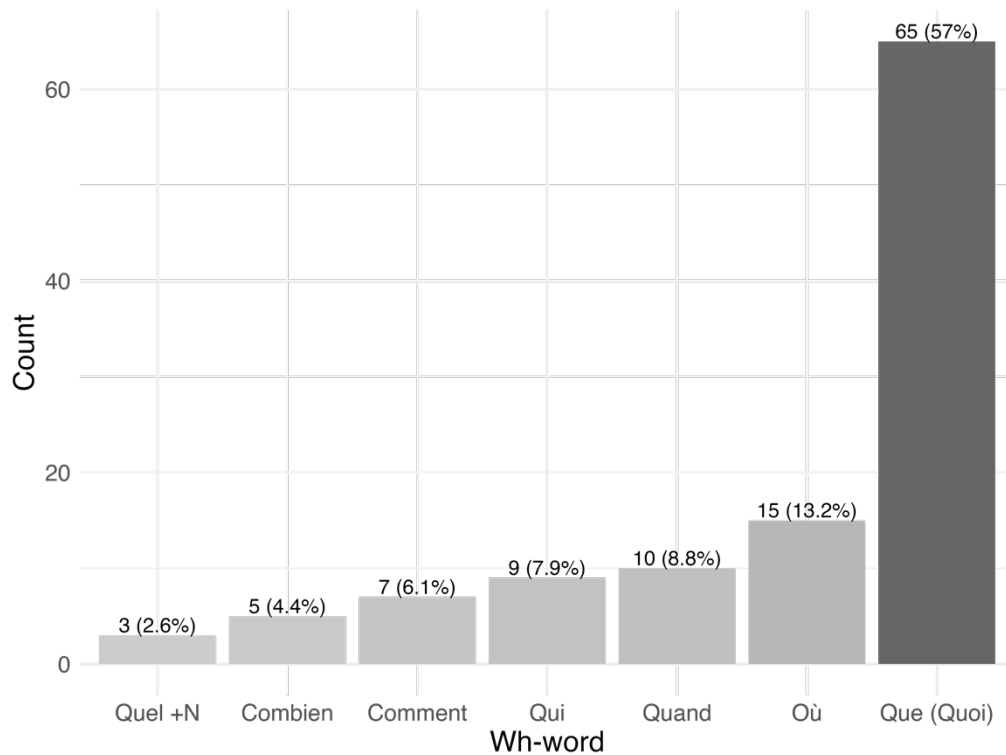


Figure 40 Distribution of *wh*-words in-situ in non-translated French

When examining translated French, *quoi* ‘what’ and *où* ‘where’ are likewise the most frequently attested *in-situ* question words, as shown in Figure 41. Although **the types of *in-situ* question words found in translated French broadly overlap with that of non-translated French**, their relative distribution diverges. More specifically, in non-translated French, the proportion of *quoi* ‘what’ is 40% more frequent than that of *où* ‘where’, whereas this gap is reduced to less than 10% in translated French. In other words, the *où* ‘where’ becomes comparatively more frequent *in-situ* in translated French, while *quoi* ‘what’ shows a noticeable decline.

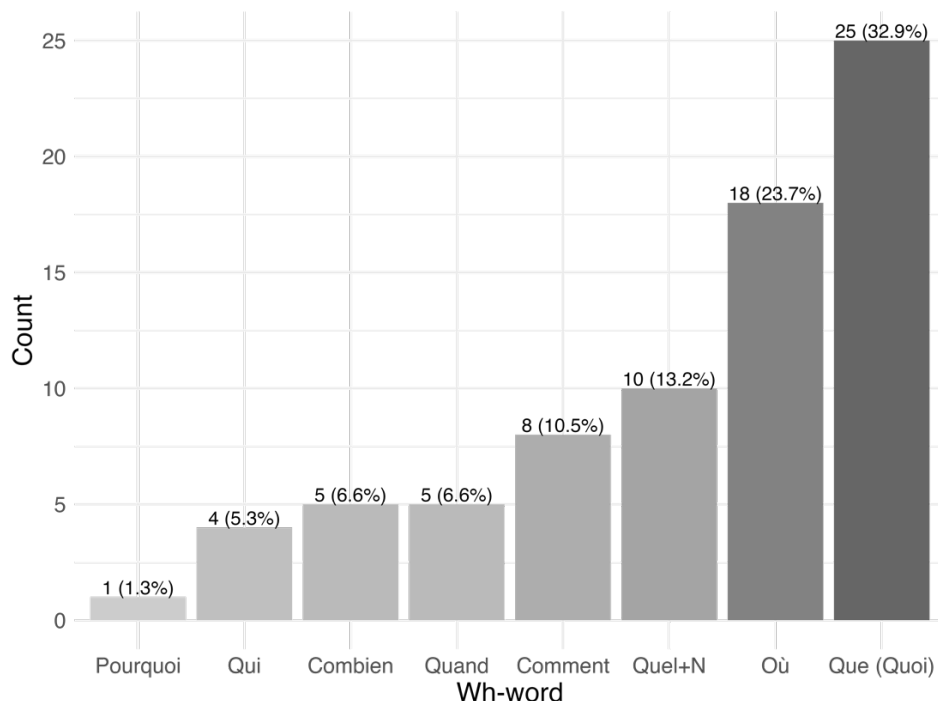


Figure 41 Distribution of *wh*-words in-situ in translated French

Other question words likewise appear *in-situ*, although their proportions within the types of *in-situ* question words show only minor differences between non-translated and translated French. An additional notable observation in translated French is the occurrence of *pourquoi* ‘why’ *in-situ*, as in (5.67).

(5.67) *Tu devines pourquoi ?*
 2SG guess why
 ‘Guess why?’

(*L’homme aux yeux à facettes*)

This finding is consistent with previous studies, since the question word *pourquoi* ‘why’ is either unacceptable or rarely found *in-situ* (Rizzi 1990: 47; Beyssade 2006: 181; Myers 2007: 153; Coveney 2011: 130; Druetta 2018: 41–42). The two major differences in the distribution of *wh*-in-situ between non-translated and translated French may stem from two factors. First, the divergence may be partly due to corpus limitations, as corpus size necessarily affects the range of observable phenomena. Given the already marginal status of *in-situ pourquoi* ‘why’, the present corpus is unlikely to capture its occurrence. Second, a possible explanation, suggested by previous theoretical work, is that the use of *wh*-in-situ can be pragmatically conditioned and strongly grounded in discourse contexts (Boucher 2010b: 130–131). Accordingly, the differences observed between non-translated and translated French may reflect the contextual environments favoured in

translated texts. In other words, the specific contexts provided by translated discourse may have allowed *où* ‘where’, *quel*+N ‘which+N’, and *pourquoi* ‘why’ to be more likely *in-situ*, while *quoi* ‘what’, though typically the most dominant question words being *wh*-*in-situ*, may be situated in such contexts less frequently in translation texts than in non-translated ones.

In addition to the distribution of *in-situ* question words, I examine instances of *in-situ* in translated French with respect to their syntactic environment. Several syntactic constraints are argued to prevent question words from appearing *in-situ*, including negation, modal verbs, universal quantifiers, and relative clauses (see Sections 2.4.4 and 4.4). In non-translated French, two types of counterevidence have been observed. First, *wh*-*in-situ* can co-occur with modal verbs. Second, this syntactic structure is allowed in embedded clauses (see Section 4.4). In translated French, a further type of counterevidence not attested in non-translated French is observed, as shown in (5.68).

- (5.68) *Ils ont trouvé l'identité de qui ?...*
 3PL COP find.PTCP ART identity PREP who
 ‘Who did they identify?’ (Le sniper, son wok et son fusil)

According to Starke (2001: 25), a complement of noun phrase with specific reference is typically inappropriate for *wh*-*in-situ*. However, in the example, the question word *qui* ‘who’ appears *in-situ* as the complement of the noun phrase with specific reference *l'identité* ‘the identity’. Nevertheless, this instance may be an echo-question, for which *wh*-*in-situ* is known to be less syntactically constrained (Chang 1997: 17–18; Cheng & Rooryck 2000: 13). The context, however, is not sufficiently clear to determine definitively whether this is indeed an echo question.

5.8.2 Discussion of *wh*-*in-situ* in non-translated and translated French

The overall results indicate a decrease in the frequency of *wh*-*in-situ* in translated French, confirming the second hypothesis concerning translation-driven **normalisation**. This pattern is consistent with the fact that *wh*-*in-situ* in French is an optional and relatively rare syntactic structure. Moreover, **the distribution of *in-situ* question words is broadly similar between non-translated and translated French**. Specifically, *quoi* ‘what’ and *où* ‘where’ are the two most frequently attested *in-situ* question words, followed by others, such as *comment* ‘how’ and *combien* ‘how much/many’. A rare instance of *pourquoi* ‘why’ occurring *in-situ* is also found in translated French. However,

a difference emerges when comparing the relative proportions of the two most frequent *in-situ* question words across the two datasets. In translated French, the difference between *quoi* ‘what’ and *où* ‘where’ narrows to around 10%, in contrast to 40% in non-translated French. A possible explanation, suggested by previous theoretical work rather than directly derived from the corpus, is that the availability of discourse environments favouring *wh*-in-situ may differ between translated and non-translated texts. Such a contextual variation may therefore account for both the reduced gap between *quoi* ‘what’ and *où* ‘where’. Finally, a possible counterevidence to the syntactic constraints of *wh*-in-situ is observed in translated French: the *in-situ* question word appears as the complement of a noun phrase with specific reference. This instance could potentially be accounted for as an echo question, although the context is insufficiently clear to determine this definitively.

5.9 Monolingual comparable analysis: Question particles

The review and tentative conclusions presented in Section 2.5 indicate that question particles share similarities across Mandarin and French, stemming from their syntactic distribution and pragmatic functions. These shared properties serve as the *tertium comparationis* for comparing question particles in the two languages. Despite these similarities, the languages differ in the variety of particles they have at their disposal. Mandarin employs four main particles (i.e., 嗎 *ma*, 吧 *ba*, 啊 *a*, 呢 *ne*) in questions, whereas French primarily uses one (i.e., *est-ce que*). Based on both bilingual comparable and bidirectional parallel analyses, the results show that question particles occur more frequently in original Taiwan Mandarin ($n = 957$) than in French ($n = 254$). Moreover, the mutual correspondence value of particles is very low (13.3%), particularly from Taiwan Mandarin to French. Therefore, it can be concluded that the two languages not only exhibit different preferences in the use of question particles, but their cross-linguistic correspondence is also very limited. Consequently, **two potential translation effects are hypothesised**. First, under source language interference, the frequency of question particles is expected to decrease in translated Taiwan Mandarin but increase in translated French. Second, under normalisation, the number of question particles is expected to increase in translated Taiwan Mandarin but decrease in translated French. Two hypotheses are summarised in Table 117.

Table 117 Two hypotheses based on two translation effects for question particles

Hypothesis	Pair of comparison	Translation effect	Frequency of question particles in translated language
I	Non-translated Taiwan Mandarin vs. Translated Taiwan Mandarin	Interference of source language	↓
	Non-translated French vs. Translated French		↑
II	Non-translated Taiwan Mandarin vs. Translated Taiwan Mandarin	Normalisation	↑
	Non-translated French vs. Translated French		↓

In what follows, Section 5.9.1 examines the comparison of question particles in non-translated and translated Taiwan Mandarin. Section 5.9.2 presents the results of particle comparison in non-translated and translated French. Finally, Section 5.9.3 provides an overall discussion of question particles across non-translated and translated languages.

5.9.1 Comparison of question particles in non-translated and translated Taiwan Mandarin

The distribution of question particles in non-translated and translated Taiwan Mandarin is shown in Table 118 and Figure 42.

Table 118 Overall distribution of question particles in non-translated and translated Taiwan Mandarin

Language	Type	Raw frequencies of particle (%)	Total frequencies of questions
Taiwan Mandarin	Non-translated	957 (25%)	3,839 (100%)
	Translated	1,502 (34%)	4,410 (100%)

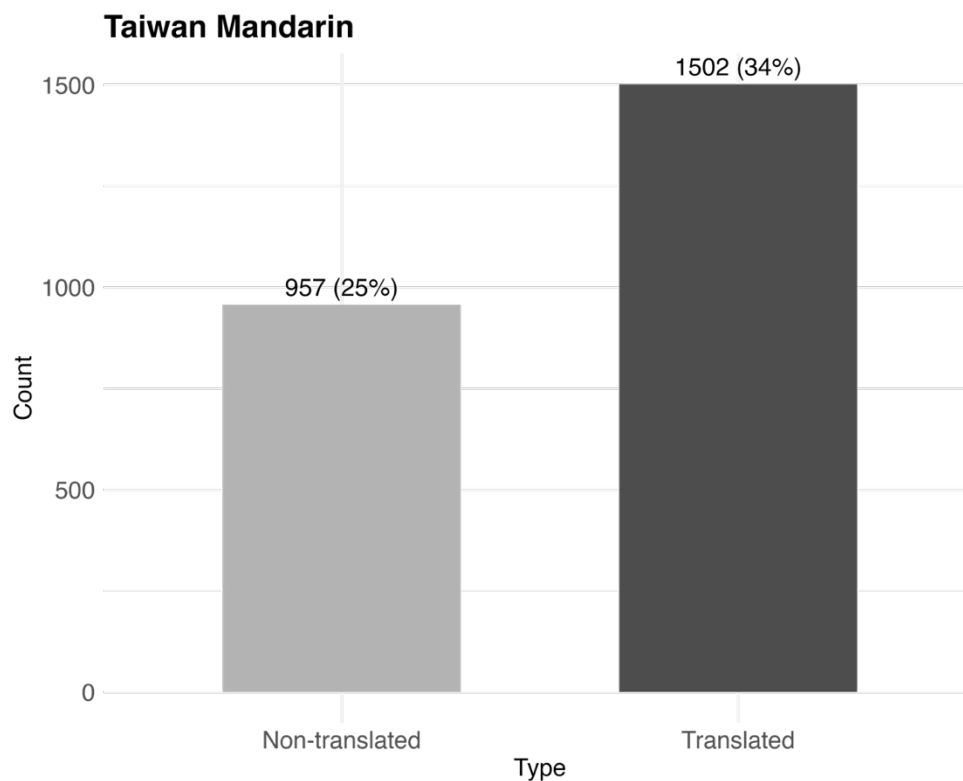


Figure 42 Overall distribution of question particles in non-translated and translated Taiwan Mandarin

Overall, the total number of question particles increases by nearly 10% in translated Taiwan Mandarin and differs **significantly** from that in non-translated Taiwan Mandarin ($\chi^2 (1) = 81.3, p < .0001$). In other words, translated Taiwan Mandarin exhibits a significant overuse of particles. This outcome supports the second hypothesis and indicates a **normalisation** effect in translation. Since the use of particles may vary across question types, I further examine their distribution in detail, as shown in Table 119 and Figure 43.

Table 119 Question particles in non-translated and translated Taiwan Mandarin across question types (raw frequencies)

Language	Type	Polar question	<i>Wh</i> -question	Disjunctive question	Total
Taiwan Mandarin	Non-translated	828	115	14	957
		86.5%	12%	1.5%	100 %
	Translated	1,202	256	44	1,502
		80%	17%	3%	100%

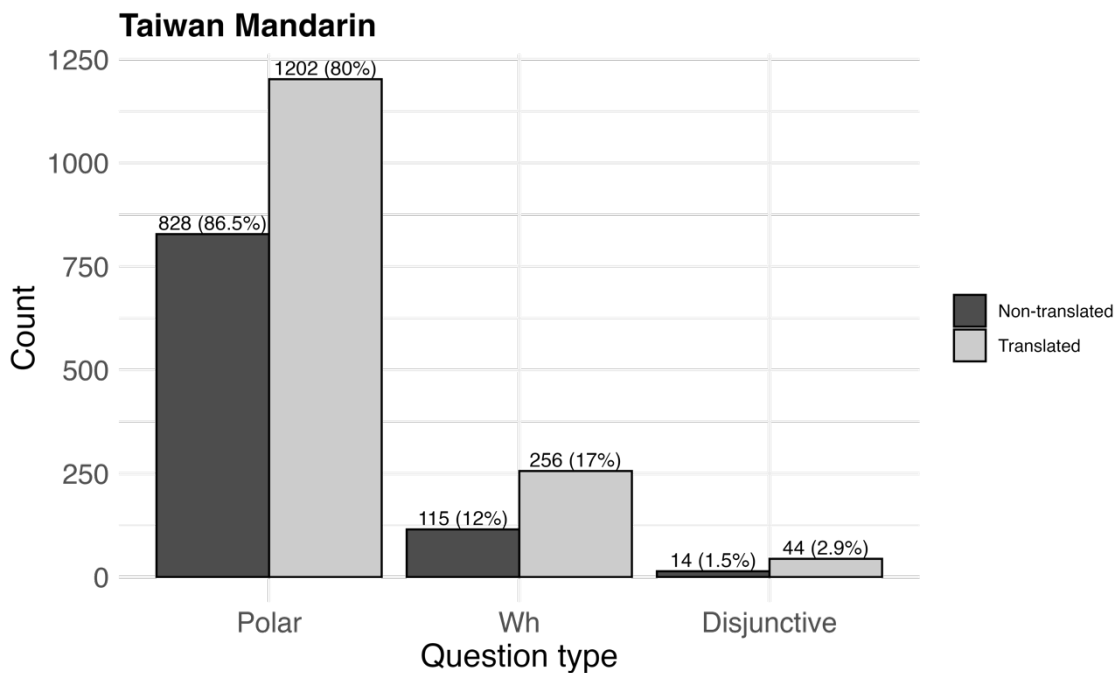


Figure 43 Question particles in non-translated and translated Taiwan Mandarin

In both non-translated and translated Taiwan Mandarin, most particles occur in polar questions, followed by *wh*-questions and disjunctive questions. Moreover, the number of question particles differs **significantly** across the three question types ($\chi^2(2) = 18.506, p < .0001$). However, notable shifts appear in translated data. Specifically, the overall proportion of particle use in polar questions decreases in translated Taiwan Mandarin, while the proportions in *wh*- and disjunctive questions increase correspondingly. This pattern is further supported by the adjusted residuals in Table 120.

Table 120 Adjusted residuals of question particles by translation status

Taiwan Mandarin		
Type	Non-translated	Translated
Polar	4.137	-4.137
<i>Wh</i>	-3.396	3.396
Disjunctive	-2.336	2.336

The residual analysis shows that the particle use in polar questions is significantly higher than expected in non-translated Taiwan Mandarin, but significantly lower than expected in the translated Taiwan Mandarin. By contrast, for *wh*- and disjunctive questions, particle frequencies are significantly lower than expected in non-translated Taiwan Mandarin yet significantly higher in translated data. These findings suggest that, **despite the overall normalisation trend, source language interference may have contributed to a reduced use of particles in polar questions in translated Taiwan Mandarin.**

In addition, I examine the detailed distribution of individual question particles and compared them with those in non-translated Taiwan Mandarin, as shown in Table 121 and Figure 44.

Table 121 Five particles in non-translated and translated Taiwan Mandarin (raw frequencies)

	嗎 <i>ma</i>	吧 <i>ba</i>	哦 <i>o / lo</i>	啊 <i>a/la/ya</i>	呢 <i>ne</i>	Total
Non-translated	666 (69.6%)	112 (11.7%)	29 (3%)	41 (4.3%)	109 (11.4%)	957 (100%)
Translated	1,037 (69%)	125 (8.3%)	21 (1.4%)	62 (4.1%)	257 (17.1%)	1,502 (100%)

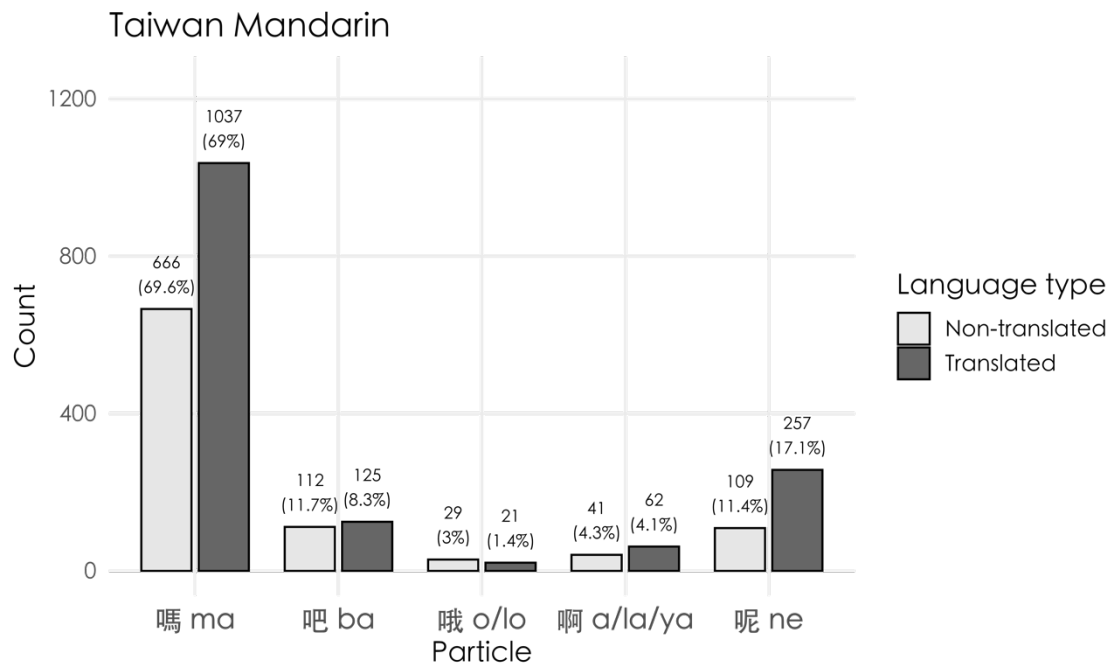


Figure 44 Five particles in non-translated and translated Taiwan Mandarin

The distribution of particles differs **significantly** between non-translated and translated Taiwan Mandarin ($\chi^2(4) = 27.5, p < .0001$). Moreover, the adjusted residuals in Table 122 indicate that certain particles contribute disproportionately to this significant difference.

Table 122 Adjusted residuals of five question particles by translation status

Type	Non-translated	Translated
嗎 <i>ma</i>	0.2888	-0.2888
吧 <i>ba</i>	2.7699	-2.7699
哦 <i>o / lo</i>	2.796	-2.796
啊 <i>a/la/ya</i>	0.1887	-0.1887
呢 <i>ne</i>	-3.886	3.886

Specifically, the particles 吧 *ba* and 哦 *o / lo* occur less frequently than expected in translated texts, while the use of 呢 *ne* exceeds the expected frequency. **This pattern suggests that normalisation manifests primarily through the increased use of particle 呢 *ne*.** Given that the use of particle in Mandarin varies across question types, I further examine the interaction between the five particles and question types, as shown in Table 123 and Figure 45.

Table 123 Five particles in non-translated and translated Taiwan Mandarin across question types (raw frequencies)

Particle	Language type	Polar	<i>Wh</i>	Disjunctive
嗎 <i>ma</i>	Non-translated	80.4% (666/828)	0	0
	Translated	86.3% (1,037/1,202)	0	0
吧 <i>ba</i>	Non-translated	13.5% (112/828)	0	0
	Translated	10.4% (125/1,202)	0	0
哦 <i>o / lo</i>	Non-translated	3.5% (29/828)	0	0
	Translated	1.7% (21/1,202)	0	0
啊 <i>a/la/ya</i>	Non-translated	2.5% (21/828)	17.4% (20/115)	0
	Translated	1.6% (19/1,202)	12.9% (33/256)	22.7% (10/44)
呢 <i>ne</i>	Non-translated	0	82.6% (95/115)	100% (14/14)
	Translated	0	87.1% (223/256)	77.3% (34/44)

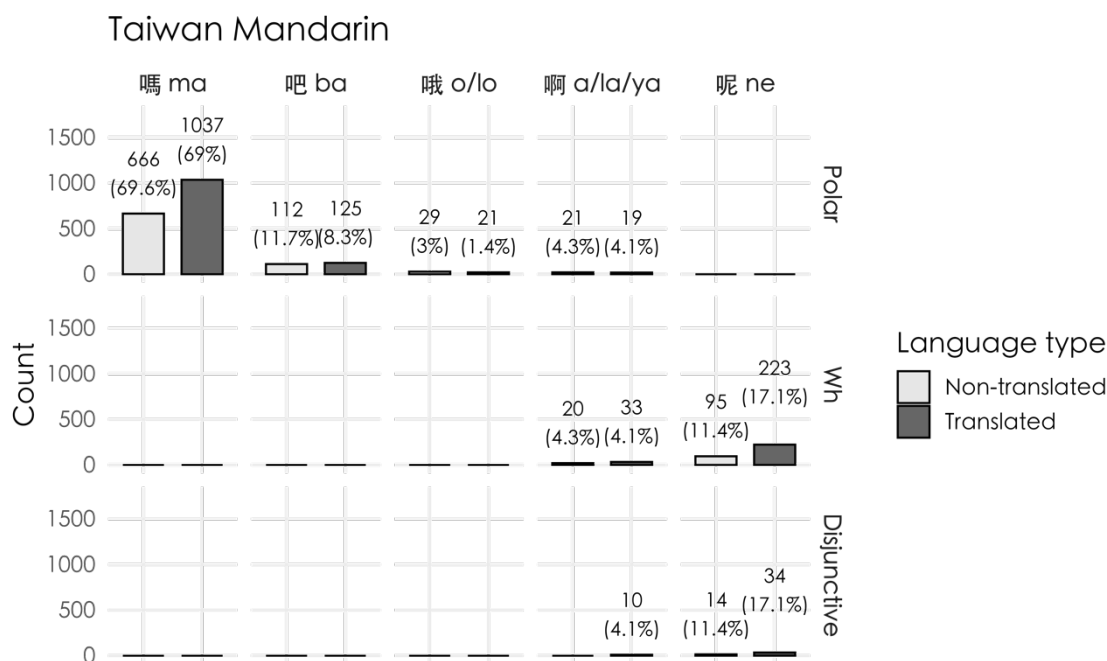


Figure 45 Five particles in non-translated and translated Taiwan Mandarin across question types

No statistical test is conducted here, as the distinctions in particle distribution are already clear. The particles 嗎 *ma*, 吧 *ba* and 哦 *o/lo* occur predominantly in polar questions, whereas 呢 *ne* appears mainly in *wh*- and disjunctive questions. The proportional results also show that each question type exhibits distinct behaviour. For polar questions, all particles tend to decrease in translated Taiwan Mandarin, consistent with the overall reduction of particle use observed in polar questions, particularly 吧 *ba* and 哦 *o / lo*. In contrast, for *wh* and disjunctive questions, particle use increases in translated Taiwan Mandarin, with the 呢 *ne* particle showing the most pronounced increase. This finding aligns with the earlier analyses.

Therefore, although particles appear to be overused in translated Taiwan Mandarin, indicating a **normalisation** effect, more fine-grained differences emerge when examining their distribution across question types and across the five individual particles. Specifically, **normalisation seems more pronounced in the use of particle 呢 *ne* in *wh*- and disjunctive questions, while source language interference may reduce the use of particle 吧 *ba* and 哦 *o / lo*, as well as the overall particle use in polar questions.** On the one hand, certain linguistic features specific to Mandarin may be highlighted and

reinforced through the normalisation, such as the preference for disjunctive questions and the distinctive functions of the particle 呢 *ne*. On the other hand, source language interference affects the overall particle use in polar questions, potentially reflecting the French preference for polar questions with declarative word order. Interestingly, the decrease in the use of particle 嗎 *ma* in translated Taiwan Mandarin is not significant compared to its use in non-translated Taiwan Mandarin. This may suggest that the particle 嗎 *ma* remains highly dominant regardless of translation status.

5.9.2 Comparison of question particles in non-translated and translated French

The overall distribution of question particles in non-translated and translated French is presented in Table 124 and Figure 46. In both datasets, the frequencies of the *ça* and *donc* particles are categorised under *wh*-questions.

Table 124 Overall distribution of question particles in non-translated and translated French

Languages	Type	Raw frequencies of particle (%)	Total frequencies of questions
French	Non-translated	254 (6.1%)	4,196 (100%)
	Translated	466 (11.3%)	4,124 (100%)

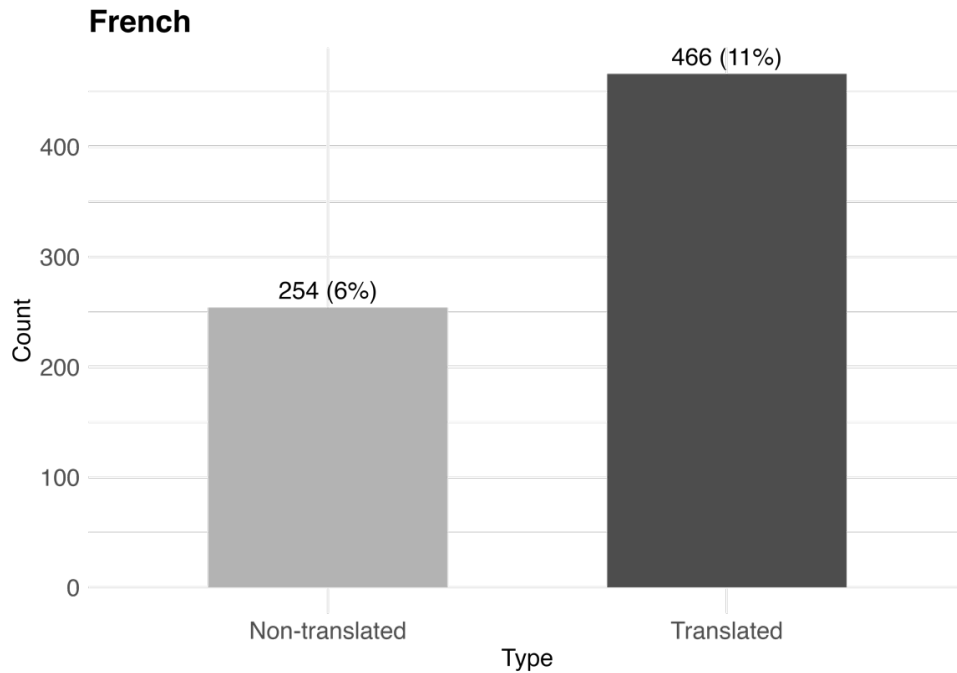


Figure 46 Overall distribution of question particles in non-translated and translated French

An increase of 5% in particle use is observed in translated French compared with non-translated French, and this difference is statistically **significant** ($\chi^2(1) = 71.76, p < .0001$). This outcome supports the first hypothesis, indicating **source language interference** effect from Taiwan Mandarin. I further examine the distribution across question types, as shown in Table 125 and Figure 47.

Table 125 Question particles in non-translated and translated French across question types (raw frequencies)

Languages	Type	Polar question	Wh-question	Disjunctive question	Total
French	Non-translated	60	194	0	254
		23.6%	76.4%	0%	100%
	Translated	149	316	1	466
		32%	67.8%	0.2%	100%

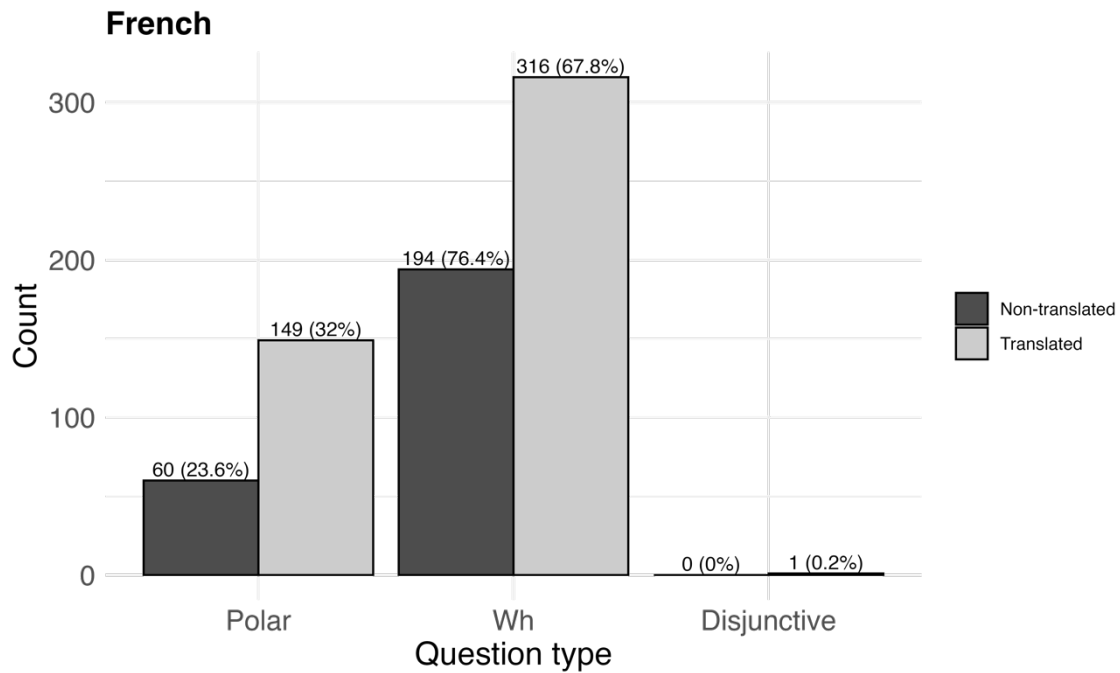


Figure 47 Question particles in non-translated and translated French

The results show that the majority of occurrences are found in *wh*-questions, followed by polar questions, in both non-translated and translated French. However, the number of question particles differs **significantly** across question types ($\chi^2 (1) = 5.2486, p < .0001$).

²¹ In addition, the adjusted residuals reveal that polar questions employ more particles than expected, while *wh*-questions use fewer particles than expected, as shown in Table 126. In other words, **the increase in particle use in polar questions may be attributed to source language interference, while the decrease in *wh*-questions reflects a normalisation effect.**

Table 126 Adjusted residuals of question particles by translation status

Type	French	
	Non-translated	Translated
Polar	-2.377	2.377
<i>Wh</i>	2.377	-2.377

In addition, I examined the detailed distribution of question particles, as shown in Table 127 and Figure 48.

²¹ Disjunctive questions are excluded from this statistical analysis, since their frequency is too low.

Table 127 Three particles in non-translated and translated French (raw frequencies)

	<i>Est-ce que</i>	<i>ça</i>	<i>donc</i>	Total
Non-translated	236 (92.9%)	17 (6.7%)	1 (0.4%)	254 (100%)
Translated	437 (93.8%)	24 (5.2%)	5 (1.1%)	466 (100%)

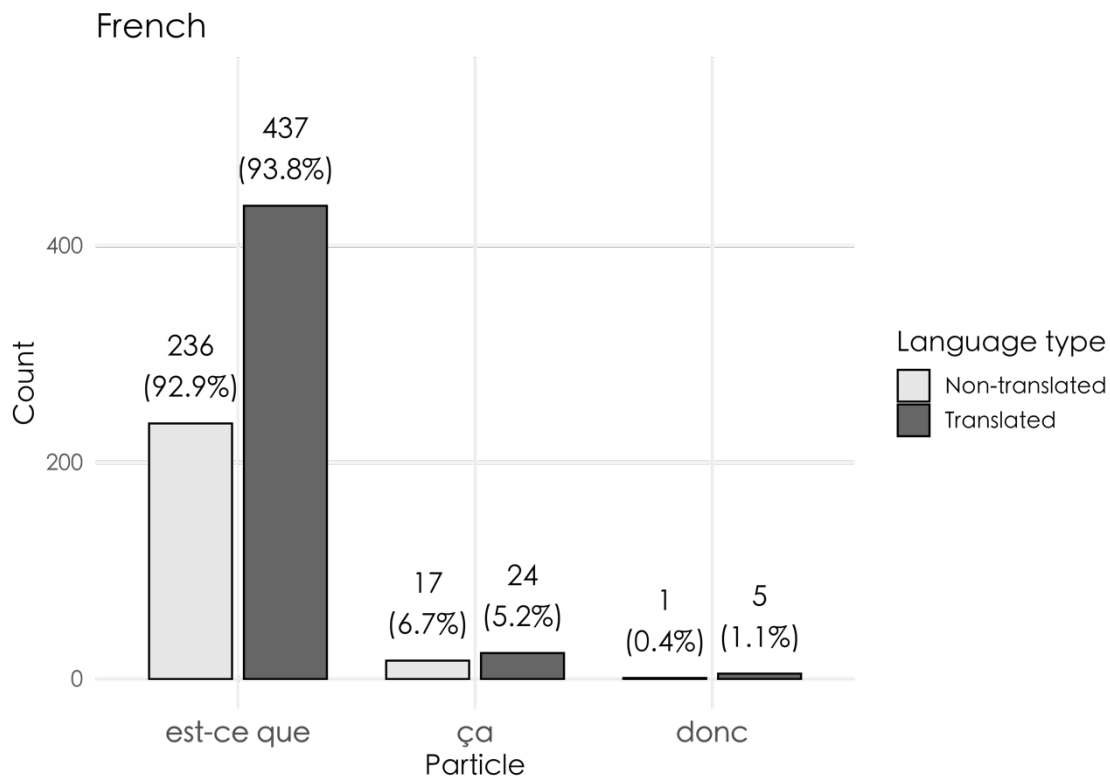


Figure 48 Three particles in non-translated and translated French

The results indicate a slight increase in the use of *est-ce que* and *donc* in translated French, while the use of *ça* particle decreases. Since *ça* and *donc* are more restricted in usage and occur infrequently in the corpus, the subsequent analysis focuses on the distribution of *est-ce que* across question types between non-translated and translated French, as presented in Table 128 and Figure 49.

Table 128 Particle *est-ce que* in non-translated and translated French across question types (raw frequencies)

Particle	Language type	Polar	Wh	Disjunctive	Total
<i>est-ce que</i>	Non-translated	60 (25.4%)	176 (74.6%)	0	236 (100%)
	Translated	149 (34.1%)	287 (65.7%)	1 (0.2%)	437 (100%)

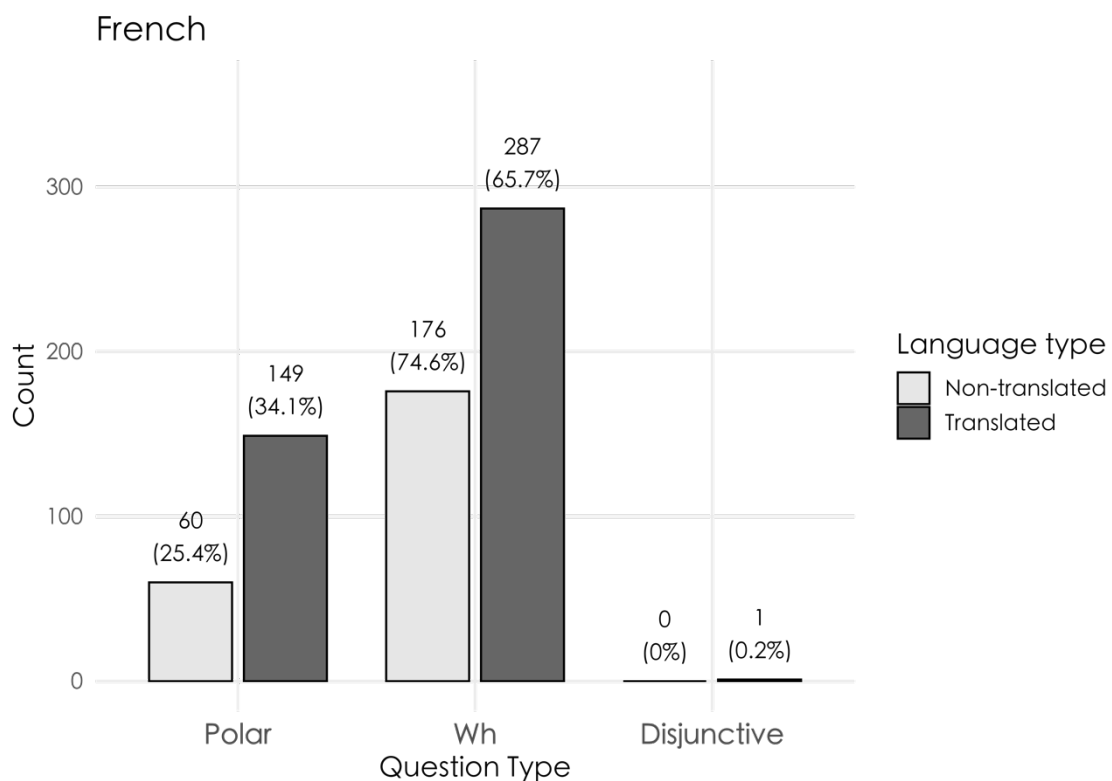


Figure 49 Particle *est-ce que* in non-translated and translated French across question types

The overall distribution of particles in translated French is largely comparable to that in non-translated French (see Section 4.5.2). In both datasets, the *est-ce que* particle occurs mostly in *wh*-questions, followed by polar questions. Noted that *est-ce que* particle does not appear in disjunctive questions in non-translated French, whereas one instance is attested in translated French, as illustrated in (5.69).

- (5.69) *Est-ce qu' il parlait d' humanité ou de*
 PTCL 3SG speak.IMPARF ART humanity or PREP
symboles ?
 symboles
 ‘Was he speaking about humanity or symbols?’ (Rue du dragon couché)

However, a distinct pattern emerges when comparing the proportion of *est-ce que* across question types between non-translated and translated French. In translated French, the use of *est-ce que* increases in polar questions but decreases in *wh*-questions. These differences are statistically **significant** ($\chi^2 (1) = 5.4716, p < .0001$).²² Furthermore, this tendency is supported by the adjusted residuals shown in Table 129.

Table 129 Adjusted residuals of *est-ce que* particle across question types by translation status

Type	French	
	Non-translated	Translated
Polar	-2.339	2.339
<i>Wh</i>	2.339	-2.339

The *est-ce que* particle is used more frequently than expected in polar questions in translated French, while its frequency is lower than expected in *wh*-questions. This finding aligns with earlier conclusion that the increase in polar questions results from source language interference, while the decrease in *wh*-questions reflects a normalisation effect.

Although the use of particle in *wh*-questions decreases in translated French, this question type remains the dominant environment for particle use in both non-translated and translated French. Accordingly, I examine the distribution of the *est-ce que* particle across different question words in both translation status. Among the *wh*-questions containing the *est-ce que* particle in translated French, the question word *que* ‘what’ accounts for the majority of occurrences, as shown in Table 130. This pattern is consistent with the distribution found in non-translated French, as shown in Table 131.

²² Disjunctive questions are excluded from this statistical analysis, since their frequency is too low.

Table 130 Distribution of the *est-ce que* across question words in translated French

Question word	<i>Que</i> ‘what’	<i>Pourquoi</i> ‘why’	<i>Où</i> ‘where’	<i>Comment</i> ‘how’
Raw frequency	233	21	12	11
Question word	<i>Quand</i> ‘when’	<i>Qui</i> ‘who’	<i>Combien</i> ‘how much/ many’	<i>Quel+N</i> ‘which+N’
Raw frequency	6	2	1	1

Table 131 Distribution of the *est-ce que* across question words in non-translated French

Raw frequency	<i>Que</i> ‘what’	<i>Où</i> ‘where’	<i>Pourquoi</i> ‘why’	Total
	172	3	1	176

Although *que* ‘what’ is the most frequently used question word with the *est-ce que* particle, a wider variety of question words co-occurs with this particle in translated French than in non-translated French, as illustrated in Table 130. This expanded range can be attributed to source language interference from Taiwan Mandarin, which may trigger the use of particles with a broader set of question words in translated French. In addition, this increased diversity may also reflect the interference of socio-cultural communication norms in Mandarin, which often favour indirectness, and consequently politeness. Since the particle *est-ce que* carries pragmatic functions related to politeness, such socio-pragmatic transfer from the source language may further contribute to its use with a more varied set of question words in translated French.

5.9.3 Discussion of question particles in non-translated and translated languages

Although question particles are more frequent in both translated Taiwan Mandarin and French, these findings reflect two distinct translation effects. In translated Taiwan Mandarin, the increase in question particles indicates a **normalisation** effect (i.e., adhere to Taiwan Mandarin norms). By contrast, the increase in question particles in translated French reflects an **interference** effect from the source language. However, more nuanced patterns emerge when examining particle distribution across question types and, in the case of Taiwan Mandarin, across individual particles.

First, the analysis across question types reveals patterns that differ from the overall trends. In translated Taiwan Mandarin, despite the general increase in particles suggesting normalisation, source language interference leads to a reduced use of particles in polar question. Conversely, in translated French, although the overall pattern indicates source language interference, a decrease in particle use in *wh*-questions reflects normalisation.

Second, the detailed distribution of individual particles shows that each particle may undergo different translation effects. In translated Taiwan Mandarin, the proportions of particles in polar questions decreases, particularly 吧 *ba* and 哦 *o/lo*, suggesting source language interference, while particles used in *wh*- and disjunctive questions increase, indicating normalisation. In translated French, although there is basically one particle, the *est-ce que* particle displays divergent tendencies across question types. Its increased use in polar questions aligns with source language interference, while its decreased use in *wh*-questions signals normalisation. Yet, when examining the question words that co-occur with *est-ce que*, a different picture emerges: translated French shows a wider variety of question words accompanying this particle than non-translated French, suggesting source language interference.

To sum up, **the overall tendencies in the two languages are consistent with their typological characteristics**, as particles are dominant in Taiwan Mandarin and but relatively restricted in French. Nevertheless, the two translation effects co-exist within a single translated set, as their interaction is complex and sometimes competitive (Lefer & Vogeleer 2013: 5–7).

5.10 Summary of monolingual comparable analysis of questions in Taiwan Mandarin and French

From a monolingual comparable perspective, three main topics (i.e., question types, *wh*-in-situ, and question particles) were examined through comparisons between non-translated and translated languages (i.e., Taiwan Mandarin and French). These analyses revealed two primary translation effects (i.e., interference from the source language and normalisation towards the target language), summarised in Table 132.

Table 132 Overall translation effects in translated languages

Language type	Topic	Translation effect
Non-translated vs. translated Taiwan Mandarin	Question types	Interference
	Question particles	Normalisation
Non-translated vs. translated French	Question types	No effect
	<i>Wh</i> -in-situ	Normalisation
	Question particles	Interference

Interestingly, both translation effects manifested in the two languages, but in different ways. In **Taiwan Mandarin**, source language interference is observed in the distribution of **question types**: translated Taiwan Mandarin shows a slight increase in polar questions and a decrease in disjunctive questions. By contrast, normalisation is evident in **question particles**, as their overall frequency in translated Taiwan Mandarin is substantially higher than in non-translated texts, suggesting potential overuse. A more detailed examination of particles distribution across question types and individual particles revealed a nuanced picture: while the general trend indicates normalisation, source language interference is also observed in polar questions, particularly for the 吧 *ba* and 哦 *o/lo* particles. In **French**, a normalisation effect is observed in use of ***wh*-in-situ**. Given the low frequency of *wh*-in-situ in non-translated French, its occurrence decreases further in translated French. Conversely, interference from the source language is evident in the use of **question particles** and in the diversity of question words accompanying *est-ce que*. This reflects the dominant role of particles in Mandarin questions, resulting in increased particle use and a broader range of question words in translated French. Nevertheless, a normalisation effect is also seen in the use of *est-ce que* particles in *wh*-questions.

Finally, although the overall trends in translation effect may appear categorical, fine-grained analyses revealed more nuanced picture. **The two types of translation effects may compete, with one effect predominating in most cases, or they may co-exist simultaneously, as observed in the translation of question particles.** Translated Taiwan Mandarin tends to adhere to the norms of its original language, as reflected in the general effect of normalisation. However, the decrease in particle use in polar questions

suggests the presence of source language interference. Conversely, translated French shows clear influence from source language interference, while the decrease in particle use in *wh*-questions indicates a normalisation effect. In some cases, the two effects may balance each other, as exemplified by question types in French translation, which exhibit no significant translation effect. These findings highlighted the complexity of translation effects and suggested that interference tolerance operates at different levels, potentially influenced by socio-cultural conditions (Toury 2012: 311; Munday, Pinto & Blakesley 2022: 155). Toury (2012: 314) suggests that “tolerance of interference [...] tends to increase when the translation is carried out from a ‘major’ or highly prestige language/culture, especially if the target language/culture is ‘minor’, or ‘weak’ in some other sense.” Given that Taiwan literature occupies a relatively minor position in World literature (see Section 3.2.1), translated Taiwan Mandarin novels may allow greater interference from French. By contrast, translated French may resist such interference, showing lower tolerance.

As shown in Chapter 4 and previous sections in Chapter 5, the results of the contrastive study of questions between Taiwan Mandarin and French are influenced by the analytical perspective adopted. The final part of the analysis aims to integrate pertinent factors from three different perspectives (i.e., bilingual comparable, parallel, and monolingual comparable) and to examine whether these factors collectively influence the overall picture of questions used in the two languages.

5.11 Exploring the relationship between questions, genre, authorship, and translation status

While examining the data from three distinct perspectives, namely, bilingual comparable, parallel, and monolingual comparable, I initially analysed the results independently from each perspective, without considering potential interactions among them. The final part of the analysis, however, aims to provide a more comprehensive understanding by integrating all relevant factors, including genre, register, language, translation status (i.e., whether the text is translated or not), author and translator, in accounting for the distribution of two dependent variables: question types and the morphosyntactic structures of questions. Given that both dependent variables consist of more than two categorical outcomes, a Bayesian multinomial mixed-effects model was employed using the *brm* function (Bürkner 2021). In the model, the aforementioned

factors were included as **fixed effects** (independent variables), such as genre, language, register, translation status, while author, translator, and work were treated as **random effects**. Note that only “work” was explicitly modelled, as each work corresponds essentially to a specific author and translator. Thus, their effects are **indirectly accounted for via the “work” as random effect**. Due to the complexity of the data, the two dependent variables were analysed in separate models. Both models included the same random effect and largely the same fixed effects. However, for the morphosyntactic structure model, language was not included as a fixed effect because the analyses were conducted separately for each language.

5.11.1 Question types in Taiwan Mandarin and French

The analysis of the regression model begins with the dependent variable: question types (Type), which comprises a total 16,569 observations and is a categorical variable within three levels (i.e., polar, *wh*-, and disjunctive). Disjunctive questions were automatically set as the reference level. The model structure included random effect for the variety of works and fixed effects (i.e., register, language, translation status, and genre) that potentially influence the distribution of the question types. The *R-hat* statistic was 1.00, indicating that the four Markov Chain Monte Carlo (MCMC) chains had successfully converged and mixed well.

Two main indicators are used to interpret the results of the Bayesian regression model. The first is the posterior mean, which represents the estimated effect of a specific variable: larger value indicate stronger effect, with positive values indicating a positive effect and negative values indicating a negative effect. The second is the 95% credible interval (CI), with emphasis placed on effects whose CI does not include zero, as this provides strong evidence for an effect. The overall statistical results are presented in Table 133 for the random effect (i.e., work) and Table 134 for the fixed effects (i.e., register, genre, language, and translation status).

Table 133 Multinomial mixed-effects model output for question types (Random effect: work)

Question type	Posterior mean of standard deviation	95% CI (lower, upper)
Polar	0.58	[0.34, 1.04]
<i>Wh</i>	0.54	[0.31, 0.95]

Formula= brm(Type ~ Register + Language + Translation_Status + Genre + (1 | Work), data = data_all, family = categorical(), chains = 4, cores = 4, iter = 2000, control = list(adapt_delta = 0.9), refresh = 50.

The 95% credible intervals for the random effect do not include zero, indicating that the effects of work are **statistically significant**. The posterior means of the standard deviation suggests a mild variation in the distribution of question types across works, with mean values between 0.5 and 0.6, a moderate range. These results indicate that while work contributes to differences in question type usage, the magnitude of this variation is relatively **modest**. The results for the fixed effects are presented in Table 134.

Table 134 Multinomial mixed-effects model output for question types (Fixed effects: register, language, translation status, genre)

Question type	Predictor	Posterior mean	95% CI (lower, upper)
Polar	Intercept	3.44	[2.80, 4.06]
	Register: Narrative	-0.89	[-1.04, -0.75]
	Register: Other	-0.05	[-0.77, 0.76]
	Genre: Fantasy	0.30	[-0.45, 1.05]
	Language: Taiwan Mandarin	-1.79	[-1.96, -1.61]
	Translation: Translated	0.11	[-0.06, 0.29]
<i>Wh</i>	Intercept	3.01	[2.42, 3.60]
	Register: Narrative	-0.16	[-0.30, -0.01]
	Register: Other	0.12	[-0.62, 0.96]
	Genre: Fantasy	0.24	[-0.45, 0.99]
	Language: Taiwan Mandarin	-1.59	[-1.77, -1.42]
	Translation: Translated	0.06	[-0.12, 0.23]

Formula= brm(Type ~ Register + Language + Translation_Status + Genre + (1 | Work), data = data_all, family = categorical(), chains = 4, cores = 4, iter = 2000, control = list(adapt_delta = 0.9), refresh = 50.

In general, **both polar and *wh*- questions are predicted to occur more frequently in dialogue and in French**, relative to disjunctive questions. This pattern aligns with observations from bilingual comparable analyses. **Register and language are the two primary factors influencing the distribution of question types**, as evidenced by 95% credible intervals that do not include zero. Accordingly, Figure 50 illustrates that only register and language (i.e., the two upper-left and upper-right graphs) show clear differences in predicted distribution, whereas translation status and genre display minimal variation. Specifically, regarding **register**, polar questions are predicted to occur more frequently in dialogue, while *wh*-questions are more frequent in narrative. Disjunctive questions show no substantial variation across registers. As for **language**, French is predicted to use more polar questions than in Taiwan Mandarin, whereas Taiwan Mandarin is predicted to have higher proportion of disjunctive questions. By contrast, translation status (non-translated vs. translated) and the genre exhibit negligible effects in predicting the distribution of question types.

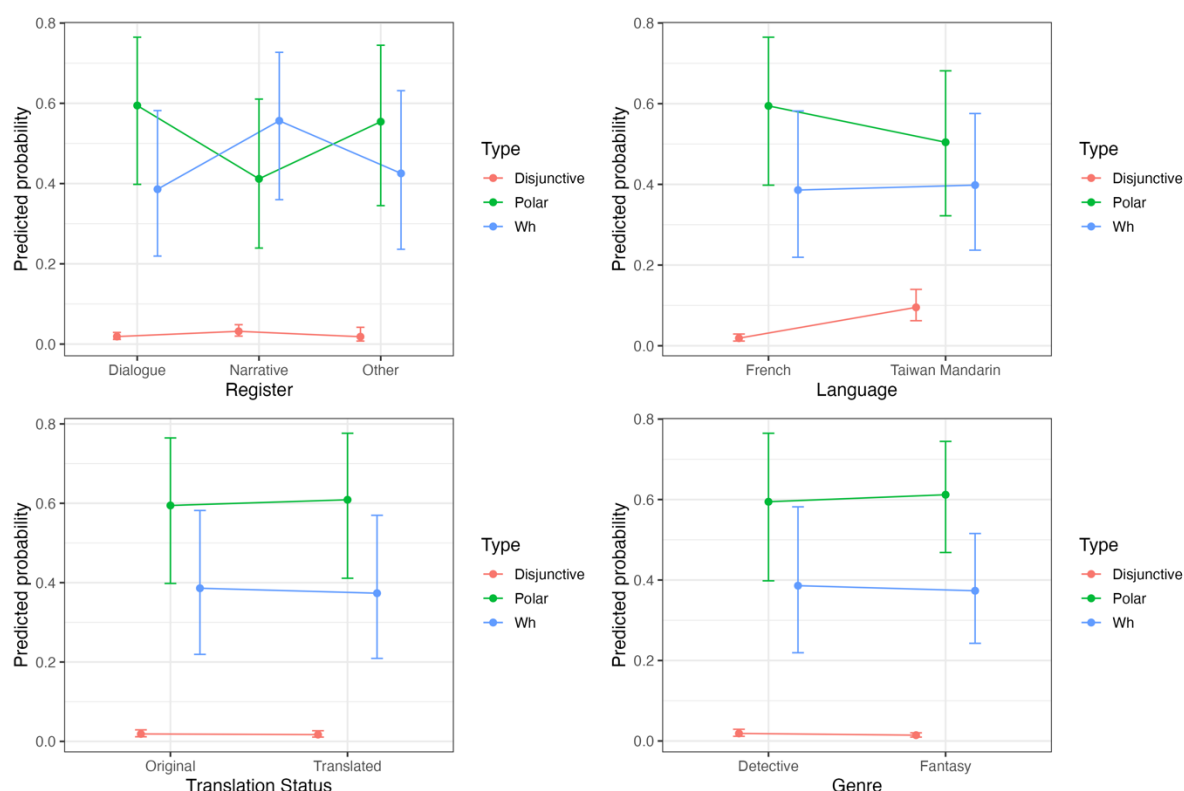


Figure 50 Predicted probability of question types by independent variables

The analyses presented in this section examine a cross-linguistic comparison of question types while accounting for four variables: register, language, translation status,

and genre. Both statistical estimates and predicted probability suggest that **register** (dialogue vs. narrative) and **language** (Taiwan Mandarin vs. French) have stronger effects on the distribution of question types, and that these predictors effectively capture the predicted pattern. Although individual works, including authors and translators, also affect the distribution of question types in the two languages, this effect is relatively minor in relation to the other predictors included in the model, according to the regression results. More importantly, these results further indicate that the distribution of question types is both language-specific and register-related. Using the same regression framework, the next analysis focuses on another dependent variable: morphosyntactic structure in Taiwan Mandarin and French, analysed separately.

5.11.2 Morphosyntactic structures in Taiwan Mandarin and French

To ensure model stability, the analysis focuses only on the typical morphosyntactic structures, including inversion, particles, and declarative word order. Because Taiwan Mandarin and French employ basically different morphosyntactic structures, two Bayesian Multinomial regression models were fitted separately for each language. The aim is to estimate the effects of register, translation status, and genre on the tendency of morphosyntactic structures, while accounting for dependencies among observations nested within individual works. The following discussion first presents the results for Taiwan Mandarin, followed by those for French.

For **Taiwan Mandarin**, the regression model includes a total of 6,727 observations. The dependent variable, morphosyntactic structures (Structure_Simple), is a categorical variable with eight levels (i.e., A-not-A, A-not-A+ particle, No inversion (declarative), No particle, Or, Or + particle, Particle, and VP-Neg). The A-not-A structure was automatically set as the reference level. The model includes fixed effects for register, translation status, and genre, and a varying intercept across the level of work. All *R-hat* values were 1.02 or below, indicating that the four Markov Chain Monte Carlo (MCMC) chains converged and mixed adequately. The model results for Taiwan Mandarin are presented in Table 135 for the random effects and Table 136 for the fixed effects.

Table 135 Multinomial mixed-effects model output for morphosyntactic structures in Taiwan Mandarin (Random effect: work)

Structure	Posterior mean of standard deviation	95% CI (lower, upper)
A-not-A + particle	1.34	[0.46, 2.87]
Declarative	0.81	[0.47, 1.37]
No particle	0.61	[0.34, 1.09]
Or	0.55	[0.18, 1.13]
Or + particle	2.58	[0.68, 6.27]
Particle	0.73	[0.41, 1.29]
VP-Neg	0.94	[0.23, 2.39]

Formula= brm(Structure_Simple ~ Register + Translation_Status + Genre + (1 | Work),
Data data = data_all_TM_STC_TYP, family = categorical(), chains = 4, cores = 4, iter = 2000, control = list(adapt_delta = 0.9), refresh = 50)

Based on the results for the random effect, all morphosyntactic structures show non-zero variation across works, indicating that the use of structure is indeed influenced by differences among works. Most of them, however, exhibit only **modest levels of variation**, with standard deviations generally falling between 0.5 and 0.8. The greatest variation is observed for the two morphosyntactic structures for disjunctive questions, *A-not-A + particle* (SD = 1.34) and *or + particle* (SD = 2.58), suggesting **substantial cross-work differences** in their frequency of use. However, the two structures are among the least frequent in the corpus, which may partially explain their high variability. By contrast, the lowest variation is found for the *Or* structure (SD = 0.55), indicating that its usage is relatively stable and evenly distributed across works.

For the fixed effects, as shown in Table 136, the discussion focuses on predictors whose 95% credible intervals (CI) exclude zero, indicating robust evidence that register, translation status, and genre affect the distribution of morphosyntactic structures.

Table 136 Multinomial mixed-effects model output for morphosyntactic structures in Taiwan Mandarin (fixed effects: register, translation status, genre)

Structure	Predictor	Estimate	95% CI (lower, upper)
ANA + particle	Register: Narrative	1.07	[0.38, 1.79]
	Register: Other	-332.36	[-1543.67, -5.43]
	Translation_Status: Translated	2.31	[0.57, 4.63]
	Genre: Fantasy	1.54	[-0.56, 3.98]
Declarative	Register: Narrative	0.20	[-0.04, 0.46]
	Register: Other	0.95	[-0.28, 2.53]
	Translation_Status: Translated	0.72	[-0.31, 1.72]
	Genre: Fantasy	-0.35	[-1.40, 0.71]
No particle	Register: Narrative	0.64	[0.41, 0.87]
	Register: Other	1.23	[0.08, 2.73]
	Translation_Status: Translated	0.90	[0.12, 1.70]
	Genre: Fantasy	-0.29	[-1.10, 0.53]
or	Register: Narrative	1.63	[1.27, 1.99]
	Register: Other	1.11	[-1.06, 3.09]
	Translation_Status: Translated	1.70	[0.88, 2.50]
	Genre: Fantasy	-0.55	[-1.44, 0.31]
Or + particle	Register: Narrative	3.32	[1.97, 5.01]
	Register: Other	-394.44	[-1879.15, -5.08]
	Translation_Status: Translated	2.51	[-1.39, 7.53]
	Genre: Fantasy	-0.71	[-5.64, 3.84]
Particle	Register: Narrative	0.06	[-0.17, 0.29]
	Register: Other	-0.23	[-1.52, 1.33]
	Translation_Status: Translated	1.21	[0.24, 2.16]
	Genre: Fantasy	0.46	[-0.51, 1.51]
VP-Neg	Register: Narrative	-0.86	[-1.93, 0.09]
	Register: Other	1.00	[-2.21, 3.34]
	Translation_Status: Translated	-0.44	[-2.19, 1.22]
	Genre: Fantasy	-2.45	[-4.70, -0.92]

Formula= brm(Structure_Simple ~ Register + Translation_Status + Genre + (1 | Work),
Data data = data_all_TM_STC_TYP, family = categorical(), chains = 4, cores = 4, iter =
2000, control = list(adapt_delta = 0.9), refresh = 50)

The three independent variables (fixed effects) exert **significant effects** on specific morphosyntactic structures. In narrative contexts, the structures *A-not-A + particle*, *No particle*, *Or*, and *Or + particle* occur significantly more frequently. Regarding translation status, *A-not-A + particle*, *No particle*, *Or*, and *Particle* are significantly more frequent in translated Taiwan Mandarin. Finally, genre show a significant effect for the structure *VP-Neg*, which is less frequent in fantasy novels. Since these results are based on comparisons with the reference level (i.e., *A-not-A*), Figure 51 presents the predicted probabilities of each morphosyntactic structure across three independent variables, using marginal effects to provide a clearer illustration of how these factors influence structure distribution.

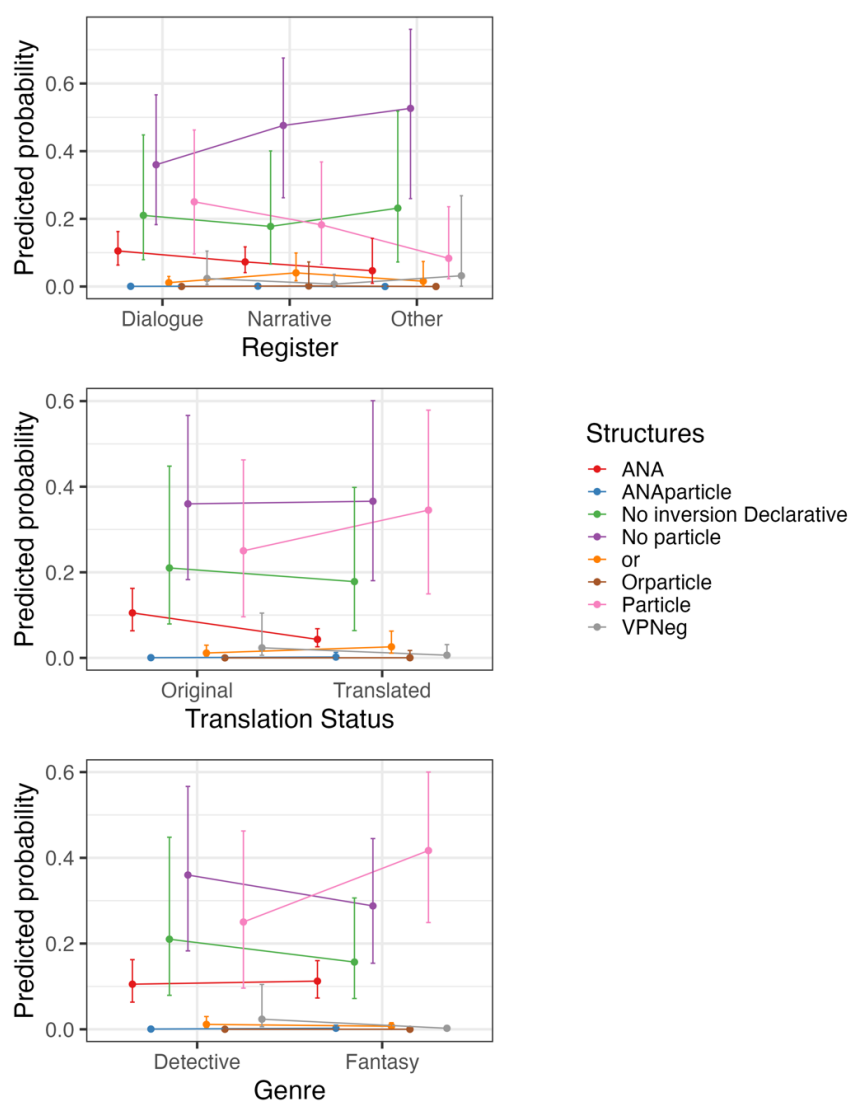


Figure 51 Predicted probability of morphosyntactic structures in Taiwan Mandarin by independent variables

As shown in Figure 51, several observations can be drawn. First, comparing **registers**, *No particle* (purple line and dots) is predicted to occur more frequently in narrative, while *A-not-A* (red line and dots), *No inversion (declarative)* (green line and dots), and *Particle* (pink line and dots) show slightly lower predicted probabilities in narrative. Second, regarding **translation status**, *Particle* (pink line and dots) shows a notable increase in translated Taiwan Mandarin, whereas *A-not-A* (red line and dots) and *No inversion (declarative)* (green line and dots) display slightly lower predicted probabilities. Finally, considering the two **genres**, *Particle* (pink line and dots) is predicted to occur more frequently in fantasy, while *No particle* (purple line and dots) and *No inversion (declarative)* (green line and dots) show a decrease. Table 137 summarises the main findings based on Figure 51.

Table 137 Summary of salient directional changes of morphosyntactic structures in Taiwan Mandarin

	Structure	Predicted probability
Register (Narrative)	No particle	↑
	Particle	
	Declarative word order	↓
	A-not-A	
Translation status (Translated)	Particle	↑
	Declarative word order	↓
	A-not-A	
	Particle	↑
Genre (Fantasy)	No particle	↓
	Declarative word order	
	Particle	

For **French**, a total of 5,828 observations were analysed in the regression model. The dependent variable, morphosyntactic structures (Structure_Simple), is a categorical variable with 8 levels (i.e., *est-ce que*, *wh*-in-situ, Inversion (polar), Inversion + *ou*, Inversion (*wh*), No inversion, Declarative word order, No inversion + *ou*), with *est-ce que* automatically set as the reference level. The model included fixed effects for register, translation status, and genre. In addition, variety of works was the random effect. Overall, the *R-hat* value was close to 1 (mostly between 1.01 to 1.05), indicating good convergence of the model. A few cases showed slightly higher *R-hat* values (up to 1.14-1.22), likely

due to sparse data for certain combinations of morphosyntactic structures and predictors levels (e.g., *Inversion* + *ou* and *No inversion* + *Ou* under *Register* = *Other*), suggesting less stable sampling for these effects. The results for the French morphosyntactic structures are first presented in the Table 138 (random effects) and Table 139 (fixed effects).

Table 138 Multinomial mixed-effects model output for morphosyntactic structures in French (Random effect: work)

Structure	Posterior mean of standard deviation	95% CI (lower, upper)
In-situ	1.09	[0.57, 2.02]
Inversion (polar)	0.86	[0.49, 1.49]
Inversion + <i>ou</i>	1.32	[0.12, 3.97]
Inversion (<i>wh</i>)	0.56	[0.29, 1.00]
No inversion (<i>wh</i>)	0.52	[0.23, 1.02]
declarative word order	0.48	[0.27, 0.84]
No inversion + <i>ou</i>	2.27	[0.52, 6.07]

Formula= `brm(bf_struct_FR_STC_TYP, data = data_all_FR_STC_TYP, family = categorical(), chains = 4, warmup = 1000, cores = 4, iter = 4000, control = list(adapt_delta = 0.9), refresh = 50)`

The results for the random effect indicate that most structures show **only modest variation across works**, as reflected by their relatively low posterior standard deviation. This suggests that these structures are generally evenly distributed across works. By contrast, the distribution of certain structures associated with disjunctive questions varies more considerably across works: *Inversion* + *ou* (SD = 1.32), and *No inversion* + *ou* (SD = 2.27). This greater variation is likely due to their lower frequencies in the dataset.

Regarding the fixed effects, the overall results are shown in Table 139. Following the same approach as before, I focus primarily on effects whose 95% credible intervals (CI) do not include zero, providing strong evidence that the corresponding predictor influences the distribution of morphosyntactic structures.

Table 139 Multinomial mixed-effects model output for morphosyntactic structures in French (fixed effects: register, translation status, genre)

Structure	Predictor	Estimate	95% CI (lower, upper)
In-situ	Register: Narrative	-1.03	[-1.57, -0.53]
	Register: Other	-482.44	[-1942.88, -13.01]
	Translation_Status: Translated	-0.92	[-2.36, 0.36]
	Genre: Fantasy	-1.01	[-2.43, 0.23]
Inversion	Register: Narrative	1.53	[1.30, 1.78]
	Register: Other	-0.44	[-2.14, 1.22]
	Translation_Status: Translated	-0.95	[-2.05, 0.24]
	Genre: Fantasy	-0.66	[-1.97, 0.36]
Inversion + <i>ou</i>	Register: Narrative	2.16	[1.40, 2.95]
	Register: Other	-1698.37	[-9963.74, -20.29]
	Translation_Status: Translated	-34.50	[-197.49, -3.57]
	Genre: Fantasy	-1.73	[-6.50, 1.10]
Inversion (<i>wh-</i>)	Register: Narrative	1.17	[0.96, 1.40]
	Register: Other	0.46	[-0.59, 1.71]
	Translation_Status: Translated	-0.79	[-1.47, -0.17]
	Genre: Fantasy	-0.27	[-1.01, 0.45]
No inversion (<i>wh-</i>)	Register: Narrative	0.41	[0.13, 0.70]
	Register: Other	0.29	[-1.05, 1.75]
	Translation_Status: Translated	-0.63	[-1.40, 0.05]
	Genre: Fantasy	-0.60	[-1.36, 0.22]
Declarative word order	Register: Narrative	-1.31	[-1.54, -1.05]
	Register: Other	-1.20	[-2.40, 0.05]
	Translation_Status: Translated	-0.35	[-0.98, 0.29]
	Genre: Fantasy	-0.44	[-1.04, 0.24]
No inversion + <i>ou</i>	Register: Narrative	-31.82	[-153.73, -1.92]
	Register: Other	-1835.18	[-8160.32, -24.20]
	Translation_Status: Translated	-21.84	[-82.40, -2.44]
	Genre: Fantasy	-3.26	[-9.98, 1.05]

Formula= brm(bf_struct_FR_STC_TYP, data = data_all_FR_STC_TYP, family = categorical(), chains = 4, warmup = 1000, cores = 4, iter = 4000, control = list(adapt_delta = 0.9), refresh = 50)

The results show that **register** has the most consistent influence across morphosyntactic structures in French. In narrative, *Inversion*, *No inversion* in *wh*-questions tend to occur more frequently, whereas *In-situ* and *Declarative word order* tend to decrease. **Translation status** also influences the use of *wh*-questions with inversion, as translated French shows fewer occurrences of these structures. By contrast, the effect of genre is limited. Figure 52, drawn using marginal effects, illustrates the interaction between the distribution of morphosyntactic structures and the three predictors (i.e., register, translation status, and genre), providing a clear view of these patterns.

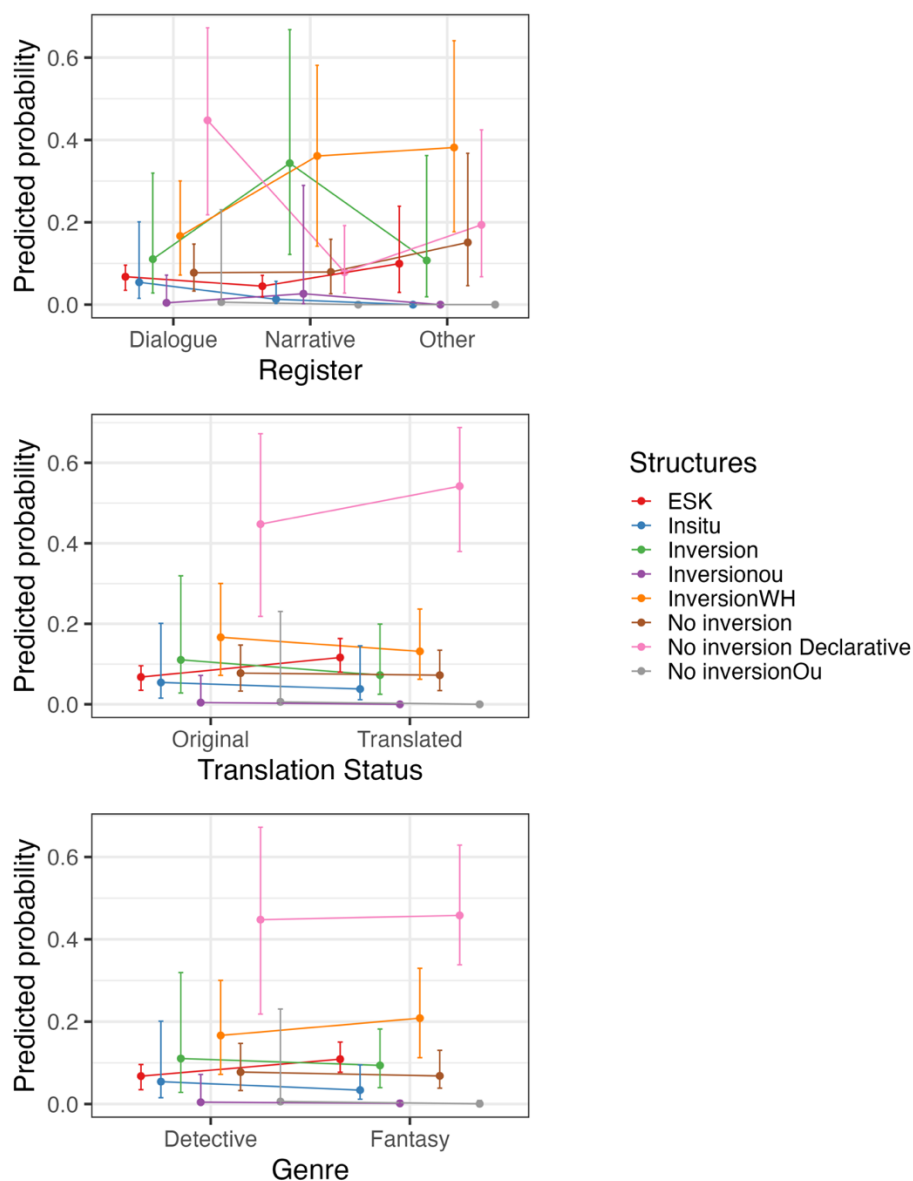


Figure 52 Predicted probability of morphosyntactic structures in French by independent variables

The three plots illustrate the predicted probabilities of different structures, conditional on register, translation status, and genre, while accounting for the effects of the predictors in the model. Several observations can be made based on the plots. For **register**, the structure *No inversion (declarative)* (pink lines and dots) tends to decrease in narrative, while *Inversion* (green and orange lines and dots) tend to increase. Regarding **translation status**, translated French tends to show higher probabilities of *No inversion (declarative)* (pink lines and dots) and *Est-ce que* (red lines and dots), while *Inversion* (green and orange lines and dots) tends to decrease. Finally, the probability of most structures appears relatively stable across **genres**. Some exceptions include *Est-ce que* particle (red lines and dots) and inversion in *wh*-questions (orange lines and dots), which show a slight increase in fantasy novels, whereas inversion in polar questions (green lines and dots) and *wh*-in-situ (blue lines and dots) tend to decrease slightly. Table 140 summarises the main findings based on Figure 52.

Table 140 Summary of salient directional changes of morphosyntactic structures in French

	Structure	Predicted probability
Register (Narrative)	Inversion	↑
	No inversion (declarative)	↓
Translation status (Translated)	<i>Est-ce que</i>	↑
	No inversion (declarative)	↑
	Inversion	↓
Genre (Fantasy)	<i>Est-ce que</i>	↑
	Inversion (<i>wh</i> -questions)	↑
	Inversion (polar questions)	↓
	<i>In-situ</i>	↓

Based on the results of the Bayesian regression models for each language analysed separately, the random effect appears to have little influence on the distribution of morphosyntactic structures in both Taiwan Mandarin and French, except for certain structures with lower frequencies in the corpus. Furthermore, the models for each language indicate that **register** generally has the stronger effect on the distribution of structures. Examining the predicted probabilities across the three predictors, several morphosyntactic structures show salient directional changes within each predictor, such

as the use of particle and A-not-A structure in Taiwan Mandarin, and the use of inversion in French. The following subsection will provide an integrated summary that compares patterns across both languages with respect to the three main factors considered in this study: register, translation status, and genre.

5.11.3 Cross-linguistic summary of question types and morphosyntactic structures with respect to register, translation status, and genre

The analyses based on the Bayesian regression model provided an integrated view of how various factors influence the distribution of question types across Taiwan Mandarin and French, as well as language-specific morphosyntactic structures. Overall, the distribution of question types is strongly affected by **language** and **register**. The bilingual comparable analysis revealed a clear cross-linguistic difference: Taiwan Mandarin favours disjunctive questions, while French shows a higher frequency of polar questions. Regarding register, although most questions in the corpus appear in dialogue, examining the effects of predictors reveals contrasting tendencies: polar questions are more frequent in dialogue, *wh*-questions are more likely to occur in narrative.

Regarding morphosyntactic structures in each language, their distribution is strongly influenced by **register** and, to a lesser extent, by translation status. In **Taiwan Mandarin**, the use of particles and the connective in disjunctive questions is particularly sensitive to these predictors. Examining the predicted probabilities across register, translation status, and genre reveals notable directional changes. For example, questions without particles are predicted to occur more frequently in narrative, whereas questions with particles are more likely in translated language. A-not-A structures, in contrast, are less probable in translated language. With respect to genre, a complementary pattern is observed: particles are more likely to appear in fantasy texts, while questions without particles are more likely to occur in detective texts.

In **French**, register has a considerable effect on the distribution of morphosyntactic structure, with inversion being particularly sensitive and more frequent in narrative. By contrast, declarative word order and *wh*-in-situ tend to decrease in narrative. Examining the predicted probabilities across registers, translation status, and genres further highlights the role of inversion. This latter is more likely to occur in narrative, but shows a decreasing tendency in translated French. Regarding genre, the inversion of polar questions tends to decrease in fantasy novels, whereas the inversion of *wh*-questions is

more likely to increase. Other structures also show notable probabilities. The particle *est-ce que* is predicted to occur more frequently in translated French and in fantasy novels, while *wh*-in-situ structure is predicted to be less frequent in fantasy texts.

To summarise, although the variation across works (authors and translators) has a slight effect on the distribution of question types in the two languages, this effect arises from the limitations of the corpus and is generally not desirable. Nevertheless, regression results indicated that its influence remains **relatively moderate** compared with other predictors. Accordingly, register and language appear to primarily influence the distribution of question types. The comparable analysis in Chapter 4 not only highlighted language-specific preferences for question types, but also showed that questions in both languages are generally more frequent in dialogue. Furthermore, the parallel analysis revealed a low correspondence value when translating disjunctive questions, indicating cross-linguistic divergence in question translation.

Regarding morphosyntactic structures, the Bayesian regression models for Taiwan Mandarin and French indicated that register has stronger effect, while translation status has a more modest effect, particularly in Taiwan Mandarin, and only minimally in French. This discrepancy in the influence of translation may be attributed to the concept of tolerance of interference, which suggests that translations from a highly prestige language/culture are more likely to exhibit interference, as discussed in the monolingual comparable analyses in Section 5.10. Additionally, these tendencies were further reflected in predicted probabilities derived from the models. Notably, these predicted tendencies align with observations from both comparable and parallel analyses. For example, particles in Taiwan Mandarin are predicted to occur more frequently in dialogue and in translated Taiwan Mandarin, while A-not-A structures are predicted to be frequent in dialogue but less common in translated Taiwan Mandarin. These predictions are consistent with the fact that particles often serve pragmatic functions in dialogue and that A-not-A structures are less frequent in translated Taiwan Mandarin (47% vs. 74% in non-translated Taiwan Mandarin), likely due to source language interference from French, which lacks this syntactic construction. Similarly, in French, inversion is predicted to appear more frequently in dialogue but less often in translated French (24% vs. 29% in non-translated French), reflecting the tendency of inversion to be used in formal context

and written language, and the potential interference of Taiwan Mandarin as a source language, which does not use subject-verb inversion to form questions.

Overall, register emerges as the most influential predictor for both distribution of question types and their morphosyntactic structures. This finding echoes the importance of register distinctions within fiction in contrastive studies. As argued by Egbert & Mahlberg (2020), novels comprise multiple discourse levels, and this observation is further supported by Ebeling & Ebeling (2020), who demonstrated the cross-linguistic and cross-register differences both in word lists comparisons of originals and in detailed case studies of *there* and *see*. In contrast, genre differences appear to play only a marginal role in shaping the distribution of questions in the two languages. This result aligns with insights from other contrastive studies across written genres (Neumann 2013; Kunz & Lapshinova-Koltunski 2015; Dupont 2021). Although these studies examined a broader range of written text types, such as editorials, scientific articles, popular-scientific texts, political essays, and instruction manuals, the underlying notion of genre distinction is comparable to the differentiation between detective and fantasy fiction adopted in the present study. Kunz & Lapshinova-Koltunski (2015), and Dupont (2021) showed that cross-genre comparison often do not yield differences in the use of certain linguistic phenomena (e.g., discourse markers), when such differences are already present within each language. In other words, cross-linguistic genre differences largely stem from the inherent characteristics of individual language systems. This insight parallels the distribution of question types in the present study, where statistical analyses reveal no significant genre effects. By contrast, Neumann (2013) showed that differences observed between comparable genres cannot be attributed solely to typological differences between language systems. Genre distinctions may still contribute to cross-linguistic divergence in ways that extend beyond system-level contrasts, underscoring the relevance of genre as an analytical dimension in contrastive studies. Consistent with this view, some morphosyntactic tendencies in the present study appear to differ between detective and fantasy fictions.

To sum up, this statistical exploration provided a more comprehensive account of how register, language, translation status, and genre influence the distribution of question types and their morphosyntactic structures. It not only consolidates and refines the findings from the bilingual comparable, parallel, and monolingual comparable analyses,

thereby offering a coherent cross-linguistic perspective, but also substantiates the relevance of cross-register and cross-genre distinctions in contrastive linguistic research.

Chapter 6 General conclusion

The final chapter synthesises the major findings of the study and discusses their implications for corpus linguistics and contrastive linguistics. The overarching objective of this dissertation has been to explore and compare the distribution of questions in Taiwan Mandarin and French novels, adopting a predominantly syntactic perspective. Three main topics were examined in detail, namely, question types, *wh*-in-situ, and question particles. While the first topic draws on the theoretical dichotomy of question types, the second and third focus on shared syntactic properties between Taiwan Mandarin and French. Given the aim of integrating theoretical accounts and corpus analysis, this study began with a theoretical overview of these three topics. This overview provided a comprehensive understanding of how each phenomenon functions across Mandarin and French, and laid the groundwork for the subsequent corpus analysis. Building upon this theoretical foundation, the empirical part of the study employed a self-built bidirectional parallel corpus of Taiwan Mandarin and French contemporary novels. The corpus analysis was conducted from three complementary perspectives (i) a bilingual comparable approach (i.e., comparing original Taiwan Mandarin and French); (ii) a bidirectional parallel approach (i.e., comparing original texts and their translations in the two translation directions); and (iii) a monolingual comparable approach (i.e., comparing non-translated and translated texts within each language). This general conclusion summarises the main findings of these analyses, highlights their contributions to contrastive linguistics, corpus linguistics, and theoretical research on question taxonomy, and presents a discussion of limitations, along with several promising directions for future research and implications for related fields.

6.1 Summary of the main findings

As stated in the introduction, the theoretical discussion and corpus analyses conducted in this dissertation were guided by two central research questions concerning the three topics under investigation:

(i) Theoretically, what are the similarities and differences between Taiwan Mandarin and French with respect to the three topics examined?

(ii) Empirically, how are these linguistic phenomena distributed in the comparable corpus and how are question types, *wh*-in-situ, and question particles translated across the two languages?

This section summarises the major findings that have emerged from both theoretical discussions and the corpus analyses. It is divided into three subsections, each presenting the results for the three topics from three analytical perspectives: bilingual comparable, parallel, and monolingual comparable, in addition to a statistical analysis. The following sections will first address question types, followed by *wh*-in-situ, and will conclude with question particles.

6.1.1 Question types

From a typological perspective, questions can generally be classified into two major categories. Nevertheless, the categorization into three question types (i.e., polar, *wh*-, and disjunctive questions) remains the most widely recognised taxonomy in the literature. Despite this popularity, a dichotomy of question types is theoretically more compelling. Drawing on pragmatic, semantic, and morphosyntactic evidence, Her, Che & Bodomo (2022) propose a dichotomy in which questions are divided into confirmation-seeking (i.e., polar questions) and information-seeking (i.e., *wh*- and disjunctive questions), on the basis of their number of propositional sets. Confirmation-seeking questions involve a singleton set of propositions and aim to test the truth value of a single proposition, whereas information-seeking questions contain multiple sets of propositions, forcing a choice among two or more propositions.

In addition to pragmatic and semantic distinctions, several morphosyntactic features (e.g., adverbs, particles, direct and indirect question forms) serve as diagnostic tests that reinforce this dichotomy. The theoretical validity of this distinction has been supported by evidence from various languages, including Taiwan Mandarin, Xiang and English (Her, Che & Bodomo 2022), Taiwan Southern Min (Hsiao & Her 2021), and Paiwan (Austronesian) (Huang & Her 2024). However, the empirical findings from our experiments on Taiwan Mandarin and French reveal an interesting contrast. In Taiwan Mandarin, the dichotomy of question types is clearly supported by morphosyntactic evidence, such as adverbs, particles, and (in)direct question forms. In contrast, the French data, tested through adverbs and (in)direct question forms, show the opposite pattern.

These findings suggests that, unlike Mandarin, French morphosyntax does not systematically distinguish polar questions from other question types; instead, the distinction in French appears to rely primarily on prosodic cues.

In addition to typological differences observed at the level of question type, Taiwan Mandarin and French also diverge in their morphosyntactic marking of questions. A closer examination of their respective morphosyntactic structures shows that Taiwan Mandarin primarily employs particles, which effectively differentiate polar questions from the other type. Furthermore, the distinctive A-not-A structure is exclusively used to form disjunctive questions. By contrast, the morphosyntactic marking strategies in French (e.g., inversion, the particle *est-ce que*) are shared across all question types. In addition, because the connective *ou* ‘or’ retains an identical form, it becomes difficult to distinguish the inclusive interpretation of *ou* ‘or’ in polar questions from its exclusive interpretation in disjunctive questions. On the basis of the theoretical discussion and experimental findings, it appears clearly that the distribution of question types differs between Taiwan Mandarin and French. These observations have therefore led to the formulation of several working hypotheses for the subsequent corpus analyses.

Based on the preceding theoretical accounts, it was hypothesised that the proportion of French polar questions would approximate the combined proportion of polar and disjunctive questions in Taiwan Mandarin. Using a **bilingual comparable approach**, the comparison between original Taiwan Mandarin and original French revealed that polar questions occur most frequently in Taiwan Mandarin and French, which is in line with the theory of social economics of questions (Levinson 2012). However, a significant cross-linguistic difference was observed across the three question types, particularly between polar and disjunctive questions. French polar questions and Mandarin disjunctive questions occur more frequently than expected, whereas Mandarin polar questions and French disjunctive questions occur less frequently than expected. This divergence may be explained by two factors. First, it may reflect socio-cultural communication differences, for example, directness in interaction. Second, it may result from morphosyntactic differences between the two languages, such as the presence of the A-not-A construction in Taiwan Mandarin and the lack of morphosyntactic distinction for polar questions in French.

Following the theoretical reasoning and the observations from the bilingual comparable analysis, two major contrasts emerged regarding question types. First, disjunctive questions are relatively frequent in Taiwan Mandarin, largely due to the widespread use of the A-not-A structure. Conversely, French exhibits a lower frequency of disjunctive questions, but a higher proportion of polar questions, reflecting its preference for direct forms of interrogation and the fact that French polar questions may implicitly contain disjunctive questions. Based on these findings, the **bidirectional parallel** was designed to test the hypothesis that disjunctive questions in Mandarin would often be translated with polar questions in French, and conversely, that French polar questions would frequently be translated with disjunctive questions in Mandarin. Only the first part of this hypothesis was confirmed. Most question types tend to be translated with their corresponding types in both directions, although a difference is observed between the two translation directions. This discrepancy is primarily due to the translation of disjunctive questions, whose mutual correspondence value is notably low. In particular, a translation bias was evident when disjunctive questions in Taiwan Mandarin are translated into French, further confirming an asymmetry in the translation of disjunctive questions between the two languages.

Finally, when comparing the non-translated and translated languages from a **monolingual comparable perspective**, it was expected that the translated language may be influenced either by the source language (i.e., interference) or by adherence to target language norms (i.e., normalisation). In the case of question types, no substantial differences were found between non-translated and translated French. However, in translated Taiwan Mandarin, the proportion of polar questions is significantly higher, while that of disjunctive questions is significantly lower. This pattern suggests a clear interference of the source language when French question types are translated into Taiwan Mandarin.

Overall, the results from both the theoretical discussion and the corpus analyses indicate that Taiwan Mandarin and French diverge systematically in their use of question types, particularly with respect to polar and disjunctive questions. While French exhibits a strong preference for polar questions, Taiwan Mandarin favours disjunctive questions.

6.1.2 *Wh*-in-situ

The second topic addressed in this dissertation concerned *wh*-in-situ, which represents the default structure in Mandarin, but serves as an optional alternative among other syntactic structures in French, such as inversion and the *est-ce que* particle. Despite this typological difference, it has been argued in the literature that *wh*-in-situ in both languages involves either covert LF movement or an unselective binding relation. Although the covert LF movement offers a theoretical account for the licensing of *wh*-in-situ in the two languages, various syntactic constraints affect whether a question word can remain *in-situ*, particularly in relation to syntactic islands, such as negation as a weak island and relative clause as a strong island. A cross-linguistic comparison revealed that these constraints operate differently in the two languages. Question words in Mandarin are primarily sensitive to strong islands, while question words in French tend to be constrained by weak islands. However, empirical evidence from French studies (Adli 2006; Zimmermann & Kaiser 2019) challenges these restrictions, showing that *wh*-in-situ can still occur within such syntactic environments. Given the complexity of covert LF movement and inconsistencies regarding island sensitivity, the unselective binding theory provides an appealing alternative, positing a relationship between an *in-situ* question word and an overt or covert operator, such as a question particle in Mandarin or a null element with a strong [+*Wh*] feature in French. In addition, other accounts have approached *wh*-in-situ from prosody in Cheng & Rooryck (2000) or pragmatic perspectives (Boucher 2010b: 104). Nevertheless, given the syntactic focus of this dissertation, the corpus analyses were restricted to the syntactic distribution of *wh*-in-situ. Moreover, the bilingual comparable and monolingual comparable analyses were limited to French, since *wh*-in-situ is a genuine structure in Mandarin that cannot be reliably identified through surface comparison.

Based on the corpus of original French novels, the distribution of *wh*-in-situ was examined. Although this structure occurs infrequently, the relative frequencies of *in-situ* question words align with the pattern reported by Druetta (2018: 43). The most frequent *in-situ* question word is *que* ‘what’, followed by *où* ‘where’ and *quand* ‘when’. Notably, we found that *wh*-in-situ with *pourquoi* ‘why’ is not attested, conforming previous observations that this question word is rarely, if ever, licensed *in-situ* (Rizzi 1990: 47; Beyssade 2006: 181; Myers 2007: 153; Coveney 2011: 130; Druetta 2018: 41–42). The

majority of *wh*-in-situ occurrences were found in dialogue in fiction, reflecting the tendency of this structure to be associated with spoken French. Finally, the corpus data provided further counter evidence to the movement-based syntactic constraints: some *in-situ* question words do co-occur with negation, and others appear in embedded contexts.

Given that *wh*-in-situ is the default structure in Mandarin, conducting a **bidirectional parallel analysis** proved challenging. Accordingly, this part of the analysis remained **exploratory**. The main goal was to investigate how French *wh*-in-situ is translated from Taiwan Mandarin, and conversely, how it is translated into Taiwan Mandarin. On the one hand, most *wh*-in-situ in French translations originated from Mandarin *wh*-questions without particle, while none were translated from *wh*-questions with particle. On the other hand, original French *wh*-in-situ was predominantly translated into Taiwan Mandarin without particle. This pattern suggests a potential association between French *wh*-in-situ and Mandarin *wh*-question without particle. However, since the majority of the *wh*-questions in Taiwan Mandarin in the corpus do not contain particles, establishing a clear and systematic parallel correspondence remains difficult.

In the **monolingual comparable analysis**, the comparison focused on *wh*-in-situ between non-translated and translated French. The results indicate a decrease in *wh*-in-situ usage in translated French, compared with non-translated French. This suggests that translated French may undergo a normalisation effect during translation from Taiwan Mandarin, whereby translators tend to reduce the use of *wh*-in-situ to align with the low frequency observed in non-translated French.

In summary, although *wh*-in-situ is a syntactic feature shared by Mandarin and French, several differences emerged from the analyses. Theoretically, various accounts can explain *wh*-in-situ in both languages, but the specific syntactic constraints differ. Empirically, the distribution of *wh*-in-situ in both original and translated French showed similar tendencies regarding which question words appear *in-situ*. Notably, the normalisation effect in translated French results in a reduction of *wh*-in-situ occurrences, reflecting the influence of target language norms during translation.

6.1.3 Question particles

Question particles represent another point of similarity between Mandarin and French. In Mandarin, particles are a dominant marking strategy for questions, with four

main particles, i.e., 嗎 *ma*, 吧 *ba*, 啊 *a*, 呢 *ne*, frequently appearing. By contrast, French primarily uses a single particle *est-ce que*, while two other particles *ça* and *donc*, are restricted to *wh*-questions. This difference in variety reflects how Mandarin relies on particles to express discourse functions, whereas French encodes these functions mainly through prosody (Hancil, Post & Haselow 2015: 18). Despite these differences, a tentative contrastive analysis revealed that particles in both languages share similar syntactic distribution and pragmatic functions. Syntactically, the Mandarin particle 啊 *a* can appear in all question types, analogous to *est-ce que* in French. Pragmatically, 嗎 *ma* and the *est-ce que* both reinforce interrogative force, eliciting responses from the addressees, while 呢 *ne* and *ça* enhance contextual coherence, and 呢 *ne* and *donc* can mark continuation in discourse. Therefore, the shared syntactic and pragmatic properties of particles provided a strong *tertium comparationis* for the corpus analyses conducted in this study.

Examining the corpus from a **bilingual comparable perspective** revealed a notable difference in particle usage: question particles are frequently used in Taiwan Mandarin than in French. This aligns with the typological distinction that particles are a main marking strategy for questions in Mandarin but only an alternative in French. In Taiwan Mandarin, the particle 嗎 *ma* dominates polar questions, whereas the *est-ce que* particle appears predominantly in French *wh*-questions. Regarding registers, both 嗎 *ma* in Mandarin and *est-ce que* in French occur frequently in dialogue of fiction, reflecting their discourse-related functions. By contrast, particles in Mandarin *wh*- and disjunctive questions (e.g., 呢 *ne*) are mostly found in narrative, consistent with their role in enhancing contextual relevance, which is more prominent in written language.

In the **parallel** data, I examined how particles are translated in both directions. Based on the literature, it was hypothesised that particles in Taiwan Mandarin would be translated into French with declarative word order and without subject-verb inversion, reflecting the tendency of French to encode discourse functions prosodically. Conversely, it was expected that French particles would be translated into Taiwan Mandarin with particles, given their similar syntactic distribution and pragmatic functions. The hypothesis was partially confirmed. From Taiwan Mandarin to French, most particles in Taiwan Mandarin are translated without a particle. Among these, polar questions typically adopt declarative word order, whereas *wh*- and disjunctive questions employ

inversion. In the reverse direction, French particles are mostly translated without particle, which contradicts the hypothesis. Note that when particles are translated with particles in either direction, it is usually the 嗎 *ma* in Mandarin polar questions corresponding to *est-ce que* in French polar questions, and vice versa. Additionally, French particles *ça* and *donc* are rarely translated with particles in Taiwan Mandarin; they are mostly translated with a single *wh*-word or without particle. Overall, the low mutual correspondence value of particles further indicates a limited direct translatability of question particles between the two languages.

The final analysis investigated the translation effect by **comparing the non-translated and translated languages** to explore translation effects. In translated Taiwan Mandarin, a significant increase in particle use is observed compared with the non-translated Taiwan Mandarin. This increase may reflect a normalisation effect in translations into Taiwan Mandarin. In addition, despite the overall higher frequency, the distribution of particle remains similar between non-translated and translated Taiwan Mandarin. Namely, most particles cluster in polar questions; specifically, the 嗎 *ma* particles remains the most frequent. However, a more fine-grained analysis across question types and individual particles revealed a different translation effect. Evidence of source language interference emerges in the reduced use of particles in polar questions, particularly 吧 *ba* and 哦 *o/lo* in translated Taiwan Mandarin. In translated French, a smaller but significant increase in particle use, along with a wider range of question words co-occurring with the *est-ce que* particle, compared with non-translated French. Both tendencies likely reflect interference from Taiwan Mandarin. Nevertheless, a closer examination of particle distribution across question types indicates a significant decrease in *wh*-questions in translated French, suggesting a counteracting normalisation effect. Consequently, although the dominant trend in translation appears to involve the transfer or retention of salient linguistic features, such as particle use, the findings also demonstrate the possible coexistence and even competition of the translation effects within a single translated text.

6.1.4 Integrated analysis based on Bayesian regression models

To provide a clearer understanding of how language phenomena interact with multiple variables, this study concluded with a comprehensive statistical analysis using

Bayesian regression models. The goal was to examine how question types and their morphosyntactic structures are shaped by factors, such as register (dialogue vs. narrative), language (Taiwan Mandarin vs. French), translation status (original vs. translated), and literary genre (detective vs. fantasy). The results show that **language and register play an important role for the distribution of question types**. This finding is consistent with the patterns revealed in the bilingual comparable analysis: French shows a preference for polar questions, while Taiwan Mandarin tends to favour disjunctive questions. The bilingual comparable analysis also indicates that questions occur more often in dialogue than in narrative. However, the Bayesian regression model reveals a more nuanced pattern: while polar questions tend to be found in dialogue, *wh*-questions show a higher probability of being used in narrative. **Regarding the morphosyntactic structures, their distributions are likewise influenced by register, translation status, and genre.** In Taiwan Mandarin, the use of particle, A-not-A structures, and declarative word order are predicted to be more frequent in dialogue. In translated Taiwan Mandarin, however, only the use of particle shows an increase, whereas A-not-A structures and declarative word order do not increase accordingly. Genre also plays a role: the use of particle is predicted to increase in fantasy texts but decrease in detective novels. In French, the predicted probabilities across the three predictors highlight a notable divergence in the use of inversion. Inversions are more likely to appear in narrative and in translated French, while non-inversion questions tend to occur in dialogue. Furthermore, genre produces opposite probability trends for inversion in polar vs. *wh*-questions: inversion in *wh*-questions is predicted to increase in fantasy novels, while inversion in polar questions is predicted to decrease in the same genre. In addition, the particle *est-ce que* is expected to increase in translated French and in fantasy novels, whereas *wh*-in-situ structures tend to be more common in detective novels.

Overall, the predicted distributions of question types and their morphosyntactic structures are shaped by register, translation status, and genre. These statistical findings revealed how different predictors interact and provide a more explicit quantification of the tendencies observed in the corpus analyses.

6.1.5 Summary of the cross-linguistic findings in Taiwan Mandarin and French

Based on this comprehensive cross-linguistic exploration of questions in Taiwan Mandarin and French, the two languages demonstrate both typological divergence and structural convergence at the morphosyntactic levels. First, although polar questions are frequently used in both languages, **the bilingual comparable analysis** revealed distinct tendencies. The preference patterns can be attributed to both socio-cultural communication styles and language-specific morphosyntactic properties, such as the A-not-A structure in Taiwan Mandarin. Second, *wh*-in-situ emerges as a syntactic feature shared by the two languages, yet realised differently. It is default and syntactically implicit in Mandarin but optional and syntactically explicit in French. While several syntactic accounts can explain the *in-situ* phenomenon cross-linguistically, the licensing mechanisms appear to operate independently within each language. Finally, question particles represent a major divergence in marking strategies. Mandarin relies heavily on a rich system of particles, whereas French uses them only marginally. Nevertheless, both languages display parallel patterns in the syntactic distribution and pragmatic functions of these particles.

Building on these observations across the three topics of questions in Taiwan Mandarin and French, several linguistic-specific features are also reflected in translation data. **The bidirectional parallel analyses** revealed a low mutual correspondence value for disjunctive questions, particularly from Taiwan Mandarin into French. Similarly, the overall mutual correspondence value for particles is also low, suggesting a mismatch between the two languages in both question types and particle use. However, the results also indicate that translation effects can shape these patterns. **The monolingual comparable analyses** demonstrated that translated language can be influenced by either source or target language. Specifically, an interference from the source language is observed when question types are translated from French into Taiwan Mandarin and when particles are translated from Taiwan Mandarin into French. Conversely, a normalisation effect occurs when particles are translated from French into Taiwan Mandarin and when *wh*-in-situ structures are translated from Taiwan Mandarin into French.

Finally, **statistical analyses** employing Bayesian regression models incorporated multiple factors, including register, language, translation status, and genre. Cross-linguistically, these factors influence the distribution of both question types and their

morphosyntactic structures. For question types, register and language exert strong effect on its distribution. With respect to morphosyntactic structures, although the two languages use different structures, their distributions are all similarly shaped by register, translation status, and genre. In each language, certain morphosyntactic structures are particularly influenced by these factors, for example, the use of particle and A-not-A structures in Taiwan Mandarin, and the use of inversion in French. A cross-linguistic comparison further showed that the use of particles is sensitive to translation status in the two languages, with an overall increase predicted in translated Taiwan Mandarin as well as in translated French. Overall, the statistical findings not only complemented the corpus analyses, but also offered a broader and more integrated perspective on how the distribution of question types and morphosyntactic structure is shaped by linguistic and textual factors.

To summarise, the findings of this cross-linguistic study highlighted that each language exhibits its own linguistic specificity in the use of question types, *wh*-in-situ, and question particles, even though initial observations suggested structural similarities. Moreover, the stronger a given linguistic specificity is, the more likely it is to be mirrored in translations, thereby shaping correspondence patterns between the two languages. Finally, the statistical analyses further reinforced and expanded upon the conclusions drawn from the corpus-based investigations, offering a more comprehensive account of how questions in both languages interact with multiple factors.

6.2 Main contributions of the study

As the dissertation lies at the intersection of two major fields: **contrastive studies** and **corpus linguistics**, the findings presented above offer both **theoretical** and **methodological** contributions to these fields. In particular, the study advances the understanding of cross-linguistic comparison of questions from a corpus-based perspective, while also proposing a more integrated approach to contrastive studies.

6.2.1 Contribution to contrastive linguistics

In contrastive linguistics, the primary objective is to conduct systematic cross-linguistic comparisons with the aim of identifying both similarities and differences across languages (Bugarski 1991: 77). Notably, this field places particular emphasis on the detailed description of linguistic differences (Gómez González & Doval-Suárez 2005: 8;

Gast 2012; Ke 2019: 8; Ramón García 2021: 394). Originating from its ‘old style’ pedagogical orientation, contrastive linguistics now encompasses every aspect of linguistic analysis, including theoretical (e.g., phonology, morphology, syntax, semantics, etc.) and applied disciplines (e.g., language learning/teaching and translation) (Gómez González & Doval-Suárez 2005: 20). With the advent of corpora, the field has further expanded to new linguistic areas, such as the study of connectors and prepositions, allowing for more empirically grounded comparisons (Aijmer 2024: 11). However, certain topics remain underexplored, particularly the study of questions (Curry & Chambers 2017: 327; Axelsson 2020: 238).

Following the study of Axelsson (2020), which appears to be the first comprehensive contrastive study on the entire system of questions of two related languages (i.e., English and Swedish), **the present dissertation extends this line of research in four major ways. First, it provides a systematic contrastive investigation of the question systems, including both question types and morphosyntactic structures. Second, it focuses on a pair of typologically distant languages (i.e., Taiwan Mandarin and French), thereby broadening the scope of contrastive study. Third, this study incorporates register variation within fiction (dialogue vs. narrative), as suggested by Axelsson (2020: 260) and Ebeling & Ebeling (2020), to capture finer distinctions in question-related phenomena, such as *wh*-in-situ and question particles. Fourth, the study introduces the Bayesian regression models, which have gained increasing popularity across different fields of linguistics (Levshina 2025), to examine complex language phenomena involving more than two categories, such as the three question types and their morphosyntactic structures, while simultaneously accounting for multiple predictors and random effects. Overall, this study contributes to contrastive linguistics by (1) offering a comprehensive cross-linguistic comparison of question systems, (2) addressing a relatively under-investigated language pair, (3) integrating register variation into contrastive framework, and (4) demonstrating the use Bayesian regression models to reinforce cross-linguistic findings and provide a robust and efficient statistical approach for contrastive research.** Within this broader framework, the detailed analyses of each question-related phenomenon further enhance our understanding of how languages encode interrogativity from a contrastive perspective.

First, the contrastive analysis of **question types** in Taiwan Mandarin and French confirmed that polar questions are the most frequently used type across languages. This finding is consistent with cross-linguistic tendencies reported in Enfield, Stivers & Levinson (2010) and with the theory of social economics of questions proposed by Levinson (2012). Moreover, the corpus results showed that this tendency also hold in translated language, suggesting that translation does not disrupt the cross-linguistic preference for polar questions. In this sense, the present study not only reinforces the existing empirical and theoretical accounts established through contrastive research on original languages, but also extends them by incorporating translation as a new linguistic dimension that reflects the same interrogative norms.

In addition, in order to examine the other two topics (i.e., *wh*-in-situ and question particles), the morphosyntactic structures of questions in both Mandarin and French were systematically coded. This coding enabled a more fine-grained understanding of how the two languages encode interrogative force, particularly given their divergent marking strategies. By moving beyond broad classifications, such as question types, the detailed structural coding offered deeper insights into the subtle morphosyntactic differences between the two languages. In turns, this methodological refinement facilitated the investigation of *wh*-in-situ and question particles, allowing for more precise cross-linguistic comparisons within the broader framework of question use.

The topic of ***wh*-in-situ** has received limited attention within contrastive linguistics, despite its prominence in generative syntax. Extensive research has been conducted on this phenomenon in individual languages, such as Mandarin (Huang 1982: 19; Watanabe 1992; 2001), and French (Shlonsky 2012; Poletto & Pollock 2004), as well as from a cross-linguistic perspective (Cheng & Rooryck 2000; Wu 2019). However, most of these studies focus primarily on the syntactic mechanisms of licensing *wh*-in-situ, rather than on its distribution and usage in authentic data. The main contribution of this dissertation is therefore to provide a corpus-based and translation exploration of *wh*-in-situ. This study not only investigated the use of *wh*-in-situ in original French but also compared it with translated French, thereby revealing how the structure behaves across different linguistic varieties. Moreover, by drawing on parallel data, this dissertation represents the first attempt to examine how *wh*-in-situ is translated from French into Mandarin, a language where *wh*-in-situ is the default strategy. Conversely, it also explored what types of

morphosyntactic structures in Taiwan Mandarin correspond to *wh*-in-situ in French, where this structure is only one of several syntactic options. Finally, the corpus analysis provided further empirical evidence regarding syntactic constraints on *wh*-in-situ in French. Several constraints are found to be violated in the corpus, a finding that aligns with previous experimental and descriptive studies (Mathieu 1999; 2004; Adli 2006; Zimmermann & Kaiser 2019). Collectively, these results not only expand the empirical foundation of *wh*-in-situ research, but also demonstrate how contrastive and translation-based corpus analyses can refine theoretical assumptions about this syntactic phenomenon.

Finally, **question particles** have predominantly been investigated from a monolingual perspective, particularly in studies on Mandarin, while they remain largely overlooked in French. Several cross-linguistic works on particles have been carried out in Hancil, Post & Haselow (2015), including comparisons such as English vs. German and Northern Italian vs. Mandarin Chinese. Moreover, detailed analyses of particles in polar questions have been presented by Bailey (2013; 2015). However, a systematic comparison involving French has been absent to date, most likely due to the limited attention paid to particles in French, given their low frequency and their primary occurrence in *wh*-questions. Furthermore, the very notion of question particle has not yet reached consensus among French linguists. On the basis of the shared syntactic distribution and pragmatic functions of particles in the two languages, the present dissertation proposes, to the best of my knowledge, the first tentative cross-linguistic comparison of question particles between Mandarin and French. The corpus findings, however, revealed that two languages employ their particles in different ways, and that the cross-linguistic correspondence between their respective particle systems is much weaker than might be expected.

Overall, this contrastive investigation began with the observation of apparent similarities between the two languages but ultimately uncovered deeper structural differences that emerge beneath this surface resemblance. These findings, in turn, offers a more substantial contribution to the theory of question.

6.2.2 Contribution to the theory of questions

Following the discussion of the overall contributions to contrastive linguistics, it is important to highlight the **theoretical implications for the theory of questions**, particularly the dichotomy of question types.

In Chapter 2, question types were examined from a typological perspective, often employing a three-way distinction. However, building on the dichotomy of question types proposed by Her, Che & Bodomo (2022), supported by theoretical evidence from other languages and two experimental studies to date, their framework appears more empirically and theoretically plausible. As demonstrated in both theoretical accounts and experimental outcomes, Mandarin can successfully distinguish two question types (i.e., confirmation vs. information-seeking questions) through specific morphosyntactic features (e.g., adverbs, particles, and question forms). By contrast, in French, the language compared with Mandarin in the present study, these morphosyntactic cues are largely absent and distinctions between question types may instead rely on prosodic features. In other words, the corpus data are expected to reflect this absence of a dichotomy in French questions relative to Taiwan Mandarin. Accordingly, the observed morphosyntactic similarities between polar and disjunctive questions in French can be primarily attributed to this lack of differentiation.

In examining the **bilingual comparable data**, the combined proportion of polar and disjunctive questions in Taiwan Mandarin is roughly equivalent to the proportion of polar questions in French. This initial observation provided a theoretical basis for the claim that French polar questions do not distinguish disjunctive questions based on their morphosyntactic structures, which leads to the assumption that French polar questions may include a subset of disjunctive questions. This overlap arises from morphosyntactic similarities (i.e., the same particle *est-ce que*, and the same connective *ou* ‘or’) and potential ambiguity between inclusive and exclusive disjunctive readings. Regarding the **parallel analysis** of question types, the results revealed low translatability of disjunctive questions, particularly from Taiwan Mandarin into French. This finding further supports the distinction in the taxonomy of question types between the two languages: translating a language with two question types into a language with only one single question type poses inherent difficulties, leading to mismatches in translation. From a theoretical perspective, the same finding may also be interpreted as suggesting that French polar

questions can assume a dual role, encompassing both polar and disjunctive questions. Finally, the **monolingual comparable analysis** of question types demonstrated source language interference, particularly comparing non-translated and translated Taiwan Mandarin. Specifically, translated Taiwan Mandarin shows a decrease in disjunctive questions and a corresponding increase in polar questions. This pattern suggests that when translating from a language with a single question type or with an ambiguous distinction between polar and disjunctive questions, the translation tends to favour the question type characteristic of the source language, while suppressing the use of question types specific to the target language.

These corpus analyses on question types not only provided contrastive evidence between the two languages but also made a clear contribution by reinforcing and extending the theoretical dichotomy of question types. They offered concrete evidence that some languages, particularly French, lack clear morphosyntactic means to distinguish polar questions from other question types. As a result, French may exhibit a blurred delineation between polar questions and other question types. This ambiguous boundary supports the view that French polar questions may also encompass disjunctive questions. Crucially, such theoretical insights have only become accessible through the integration of comparable and parallel corpus analyses, which together highlight the nuanced interplay between typological similarity and language-specific realization. Beyond these theoretical implications, the dissertation also makes several methodological contributions to the field of corpus linguistics.

6.2.3 Contribution to corpus linguistics

In addition to its contribution to contrastive linguistics, the methodological framework adopted in this dissertation constitutes an indispensable element in obtaining and presenting the findings. Since technological development in the 1990s, large-scale text collections have become digitally accessible (Gómez González & Doval-Suárez 2005: 22; Ramón García 2021: 395). This advancement has introduced the corpora as a powerful tool to the field of contrastive linguistics, broadening the possibilities for cross-linguistic comparison and offering new empirical insights based on authentic language use. Consequently, contrastive studies began to employ corpus-based methodologies to examine linguistic phenomena in greater detail. In particular, comparable and parallel

corpora have become essential resources for contrastive research (Aijmer 2024: 12). Each provides unique advantages: parallel corpora enable efficient identification of less straightforward equivalents across languages through the direct comparison of source and target texts, although they are inevitably influenced by translation effects. Comparable corpora, by contrast, are free from such translation effects because they consist of original texts in each language; however, they pose challenges in defining a clear the *tertium comparationis* between languages. Building on the respective strengths and weaknesses of both corpus types, Johansson (1998; 2007) proposed combining them within a unified framework, resulting in the construction of a multilingual corpus model that integrates parallel and comparable corpora. This design allows researchers to control translation effects while still identifying cross-linguistic correspondences with precision. Accordingly, **the first contribution of this dissertation to corpus linguistics lies in the compilation of a bidirectional parallel corpus specifically designed for Taiwan Mandarin-French comparison.** Furthermore, the analyses conducted in this study demonstrated that Johansson's model (1998; 2007) is both empirically useful and methodologically efficient for carrying out contrastive research through three different approaches: bilingual comparable, bidirectional parallel, and monolingual comparable.

The second contribution to corpus linguistics concerns the language pair involved in this corpus, namely Taiwan Mandarin and French. To the best of my knowledge, this bidirectional parallel corpus of contemporary novels represents the first of its kind. While several comparable or parallel corpora of Chinese have been employed in contrastive studies, such as *The Lancaster Corpus of Mandarin Chinese*, *The Zhejiang University Corpus of Translational Chinese*, and *The Babel English-Chinese Parallel Corpus*, these corpora primarily focus Mandarin Chinese from Mainland China to ensure the textual homogeneity. However, given the linguistic variation in Mandarin Chinese from Taiwan or Singapore, it is essential to explore the varieties of Chinese languages beyond Mainland China (Xiao & Hu 2015: 38, 43). Moreover, as frequently noted in corpus-based contrastive studies, research has been heavily clustered around European languages, with English being particularly dominant (Xiao & Hu 2015: 173; Hasselgård 2020; Čermáková et al. 2021: 90). There is thus a notable scarcity of corpora that include genetically distant languages, such as French and Mandarin Chinese. Even though corpora for other language pairs, such as Japanese-Chinese or Russian-Chinese, are

gradually being developed (Shei & Gao 2017: 533), most studies remain limited to unidirectional translation. In this study, following the standards for building a balanced comparable corpus (e.g., controlling for genre and publication time), a selection of contemporary novels from Taiwan and France were carefully made, taking into account local recognition of the works and authors as well as the availability of translations in both directions. Accordingly, the creation of this bidirectional parallel corpus from scratch constitutes a significant contribution to corpus linguistics by combining the advantages of both parallel and comparable corpora, while also providing access to a language pair that has been underrepresented in the field. However, this corpus is not publicly accessible, as copyright permissions for the included novels were not obtained from publishers.

To summarise, the contributions to corpus linguistics are twofold. First, **this study demonstrates the efficiency of the model integrating comparable and parallel corpora**. Using this model, the study systematically investigated questions from three different approaches, an analysis that can only be achieved with a bidirectional parallel corpus. Second, **the self-compiled multilingual corpus of contemporary novels in Taiwan Mandarin and French addresses the long-standing gap in language pair diversity in contrastive studies**. Furthermore, while the corpus is not publicly accessible, it opens up new opportunities for my own future research across multiple linguistic levels between the two languages. These contributions pave the way for further investigations. However, the present study is subject to certain limitations, particularly in its methodology, which will be addressed in the following section.

6.3 Limitations of the study

This dissertation has provided several new insights into the systems of questions in Taiwan Mandarin and French. Nevertheless, it does not claim to have exhausted all possible perspectives on this topic, and there remains room for methodological and empirical refinement in future research. Accordingly, this section aims to outline the main limitations of the present study, particularly those of a methodological nature.

The first limitation concerns the **composition of the corpus**, which is arguably the most critical issue in ensuring comparability and representativeness in a contrastive study. When examining the distribution of questions across individual works, variation appears

to have a minor effect on the overall patterns observed in the study, suggesting that corpus composition indeed remains a potential source of concern. Despite efforts to maintain comparability in terms of genre and publication time, limitations in the availability of novels made the task very challenging. In particular, the pool of eligible Taiwan Mandarin novels for inclusion was relatively small, largely due to the considerable imbalance in the translation direction from Taiwan Mandarin to French, as discussed in Chapter 3. Furthermore, it should be acknowledged that other Taiwan Mandarin novels not included in the corpus were published and translated into French after 2010 (e.g., *Les Lignes de navigation du sommeil*, *La Salle d'attente*, *Encore plus loin que Pluton*).²³ These could have been included in the corpus. However, **practical constraints, such as limited institutional resources for obtaining and processing texts**, as well as the need to maintain comparability between genres across languages, **guided the final selection of the twelve novels included in the corpus**. Given these representativeness caveats, variables such as authorship and translators were taken into account when interpreting the findings. Nonetheless, if a larger number of novels were to be included, the corpus would naturally achieve greater balance and better representativeness, thereby providing a more robust foundation for contrastive and translation analysis.

In addition, **the restriction to a single literary genre** presents another limitation in the composition of the corpus. Since the present study examines only novels, its findings may be confined to this specific type of written texts. Expanding the corpus to include a broader range of literary genre (e.g., poems, prose, or essays) would not only yield potentially different patterns, but also enhance the generalization of the findings.

Another limitation concerns the **extraction of the data**. The present research focused on questions marked with question marks, following the most common and efficient approach adopted in corpus-based studies of questions (Wikberg 1996; Curry & Chambers 2017; Axelsson 2020). However, beyond the direct use of questions that genuinely seek information, those ending with a question mark may also serve indirect

²³ See the following novels translated from Taiwan Mandarin into French after 2010. *Les Lignes de navigation du sommeil* (2012), translated by Gwennaël Gaffric (Éditions You Feng). Translation of 睡眠的航線 (*Shuìmián de hángxiàn*, 2007), by 吳明益 (Wu Ming-yi). *La Salle d'attente* (2016), translated by Marie-Louise Orsini (Piranha). Translation of 等候室 (*Děnghòushì*, 2013) by 鄒永珊 (Tsou Yung-Shan). *Encore plus loin que Pluton* (2018), translated by Lucie Modde (L'Asiathèque). Translation of 比冥王星更遠的地方 (*Bǐ Míngwángxīng gèng yuǎn de dìfāng*, 2012), by 黃崇凱 (Huang Chong-kai).

functions, such as expressing permission, suggestion, surprise, or invitation (Siemund 2017: 20). Conversely, there are also instances in which questions lack a question mark for some reasons (Axelsson 2020: 241). Therefore, if one aims to provide a more exhaustive description of a language's question system, it would be necessary to manually identify both the questions without question marks and those with questions that express indirect functions.

Finally, another limitation concerns **mode of language**. Despite efforts to contact relevant institutions to acquire official movie subtitles as a source of the spoken data, the attempts were unsuccessful due to copyright restrictions. Consequently, only written language is considered. To compensate for the absence of spoken data, this study distinguished between two registers in fiction (dialogue vs. narrative), since dialogue in fiction is not only a reproduction that closely imitates colloquial, spontaneously conceived speech, but also differs from the rest of the text in certain aspects, such as linguistic features (Guillot-Barbance, Pincemin & Lavrentiev 2018: 54–55; Ebeling & Ebeling 2020). Furthermore, dialogue in fiction is situated on a continuum that lies closer to the pole of communicative proximity and orality, in contrast to genres such as journalistic writing (Koch & Oesterreicher 2001). Nevertheless, dialogue in fiction still differs from natural, spontaneous speech, as it is premeditated and written in advance (Lefeuvre 2020: 23). Thus, although questions are often found to be infrequent in spontaneous spoken language (Li 2007: 116; Zimmermann & Kaiser 2019: 133), this aspect remains worthy of investigation, given the considerable differences in linguistic features between written and spoken language modes. In other words, the distribution of question types and morphosyntactic structures may vary across the two modes (written and spoken).

Overall, the limitations of this study are primarily methodological. In addition, these limitations can in turn suggest promising directions for future investigation into questions in Taiwan Mandarin and French.

6.4 Promising avenues for future research

These limitations, in fact, open up several promising avenues for future research. **Methodologically**, the first improvement concerns the addition of another level of coding, particularly for pragmatic functions of questions and question particles. Incorporating this

layer of coding would make it easier to manually identify questions that contain question marks but function as indirect speech acts. Moreover, the analysis of *wh*-in-situ and question particles would become richer and more comprehensive when their pragmatic functions are taken into account. Second, the number and diversity novels could be expanded by extending the time slot of the texts included in the corpus. In addition, incorporating other literary genres would also be worth considering. Such expansion would allow the bidirectional parallel corpus to become more balanced and diverse in terms of content and the range of authors/ translators involved. Additionally, such an expansion could shed light on translator style, a phenomenon not explored in the present contrastive study but increasingly recognised in translation research (Shei & Gao 2017: 525). Finally, the construction of a bidirectional parallel corpus for spoken language (e.g., films, interpreted conferences) or a bilingual comparable corpus of spoken language (e.g., everyday conversations, radio broadcasts, lectures) appears particularly promising. A comparison between written and spoken language would provide a more thorough and comprehensive picture of language systems. Beyond the written-spoken distinction within one language, cross-varietal comparisons, such as the Mandarin between China and Taiwan or among different varieties of French, would also offer valuable insights for future contrastive and corpus-based research.

In addition to methodological directions outlined above, the present findings also offer **theoretical avenues** for future research. The taxonomy of question types developed in this study provides a foundation for a more comprehensive (re)examination of question typologies in other languages. As more linguistic varieties are incorporated into future analyses, the increased diversity of data will further reinforce and validate the dichotomy of question types.

Finally, with regard to **the topic investigated**, the present study focuses exclusively on typical question forms for the reasons of methodological symmetry in the contrastive analysis. Further research may therefore extend the scope of investigation to elliptical question forms in the two languages. Such an extension is feasible, as all questions in the corpus have already been coded for question types and morphosyntactic structures, together with information, such as the types of question words and particles involved. Ultimately, a contrastive study incorporating both complete and elliptical question forms

would provide a more comprehensive corpus-based account of the question systems in the two languages.

6.5 Implications for related fields

This general conclusion ends with a brief discussion of possible applications in other fields. In particular, any discipline that involves the analysis of two or more languages can benefit from corpus-based cross-linguistic research, such as natural language processing and lexicography (Granger 2010). Among these potential areas, the most pertinent and direct applications of the present study lie in **foreign language teaching** and **translation education**. In these fields, the corpus compiled for this research can serve as a valuable resource for data-driven learning (Gilquin & Granger 2022). Using bilingual corpora, learners discover the regularities of patterning in the target language and teachers develop exercises based on concordance outputs (Johns & King 1991: iii). Furthermore, data-driven learning can assist trainees in translation by drawing their attention to typical and atypical translation solutions for typical problems found by expert translators (Bernardini 2004: 20). Beyond data-driven learning, the present findings also lend themselves to material design in language teaching. Based on the contrastive patterns uncovered in this study, teaching units dedicated to questions in Mandarin and French could be developed. For example, conversational materials may highlight the specific preference for certain question types, enabling learners to acquire these distinctions at an early stage. In addition, structural similarities, such as *wh*-in-situ or the use of particles in questions, can be explicitly pointed out at appropriate moments in instruction to facilitate the learning process.

To summarise, future research should begin with several methodological improvements, such as adding an additional layer of coding for pragmatic functions, diversifying the corpus in terms of content and text types, and constructing a corpus for spoken language. By addressing these current limitations, new avenues of exploration can be made available, for instance, manually identifying the questions that function as indirect speech acts, examining *wh*-in-situ and question particles from perspectives beyond the morphosyntactic level, and comparing questions across written and spoken modes. In addition to these methodological refinements and research extensions, the

corpus and analyses developed in the present study also hold potential for practical applications, particularly in foreign language pedagogy and translation training.

Appendix

This appendix shows the structure of the code using for the regression models in both Chapter 4 and Chapter 5: Poisson regression model and Bayesian regression model.

*****Chapter4_Comparable analysis*****

##Pacakges preparation

```
knitr::opts_chunk$set(echo = TRUE)
# set seed for randomness
set.seed(123)
# remove scientific notations such as 10000=1e04
options(scipen=999,stringsAsFactors = FALSE)
# reading basic packages
library(tidyverse)
library(stringi)
library(readxl)
# read some more packages for plotting
library(plotly)
library(ggpubr)
library(GGally)
# for mixed models
library(lme4)
#install.packages("car")
library(car)
#install.packages(c("lme4","afex","emmeans","lsmeans","Matrix"))
library(afex)
library(lmerTest)
library(emmeans) # library(lsmeans)
# for dependency parsing
# library(udpipe) ``

options(repos = c(CRAN = "https://cran.rstudio.com/"))
if(!require(vcd)) install.packages("vcd")
library(vcd)
if(!require(brms)) install.packages("brms")
library(brms)
```

##Read data

```
#Import data
data_MD <- read_csv("Mandarin_original.csv")
data_FR <- read_csv("French_original.csv")
```

```

#Add genre to the data set
data_MD <- data_MD %>% mutate(Genre = case_when(Work %in% c("炒飯","私家")~ "Detective", Work %in% c("天橋","泡沫","複眼人","鬼地方")~ "Fantasy"))

data_FR <- data_FR %>% mutate(Genre = case_when(Work %in% c("Gataca","Trois jours","Avion")~ "Detective", Work %in% c("Anomalie","Elfes","Voleur")~ "Fantasy"))

#Add language to the data set
data_MD <- data_MD %>% mutate(Language = "Taiwan Mandarin")
data_FR <- data_FR %>% mutate(Language = "French")

## Poisson regression for the overall distribution of question marks

#Import the file for the meta data of author, work, genre, and text size to run the Poisson regression
Meta <- read_csv("Meta.csv")
Meta <- Meta %>% na.omit(Meta)

#Poisson regression for question marks in Mandarin (Table48)
GLM_poisson_QMark_MD <- Meta %>% filter(Language == "Taiwan Mandarin") %>%
  glmer(Count ~ Genre+ (1 | Work),
    family = poisson,
    offset = log(Total),
    control = glmerControl(optimizer = "bobyqa"),
    data = .)

summary(GLM_poisson_QMark_MD)
# Overdispersion check for poisson regression in Mandarin
overdisp_fun <- function(GLM_poisson_QMark_MD) {
  rdf <- df.residual(GLM_poisson_QMark_MD)
  rp <- residuals(GLM_poisson_QMark_MD, type = "pearson")
  Pearson.chisq <- sum(rp^2)
  ratio <- Pearson.chisq / rdf
  pval <- pchisq(Pearson.chisq, df = rdf, lower.tail = FALSE)
  data.frame(
    Pearson_chisq = Pearson.chisq,
    df = rdf,
    dispersion_ratio = ratio,
    p_value = pval
  )
}

```

```
overdisp_fun(GLM_poisson_QMark_MD)
```

```
#Poisson regression for question marks in French (Table49)
```

```
GLM_poisson_QMark_FR <- Meta %>% filter(Language == "French") %>%  
glmer(Count ~ Genre + (1 | Work),  
      family = poisson,  
      offset = log(Total),  
      control = glmerControl(optimizer = "bobyqa"),  
      data = .)
```

```
summary(GLM_poisson_QMark_FR)
```

```
# Overdispersion check for poisson regression of French
```

```
overdisp_fun <- function(GLM_poisson_QMark_FR) {  
  rdf <- df.residual(GLM_poisson_QMark_FR)  
  rp <- residuals(GLM_poisson_QMark_FR, type = "pearson")  
  Pearson.chisq <- sum(rp^2)  
  ratio <- Pearson.chisq / rdf  
  pval <- pchisq(Pearson.chisq, df = rdf, lower.tail = FALSE)  
  data.frame(  
    Pearson_chisq = Pearson.chisq,  
    df = rdf,  
    dispersion_ratio = ratio,  
    p_value = pval  
  )  
}
```

```
overdisp_fun(GLM_poisson_QMark_FR)
```

```
## Possion regression for the distribution of question marks by genres (Detective)
```

```
library(MASS)
```

```
# Create dataset
```

```
Q_Mark_Detective <- data.frame(  
  Language = c("Mandarin", "Mandarin", "French", "French", "French"),  
  TextSize = c(70395, 113488, 173688, 153598, 67852),  
  QuestionMarks = c(868, 1117, 1450, 1204, 304))
```

```
Q_Mark_Detective$Language <- factor(  
  Q_Mark_Detective$Language,  
  levels = c("Mandarin", "French"))
```

```

#offset (log TextSize )
Q_Mark_Detective$offset <- log(Q_Mark_Detective$TextSize)

# Poisson regression

model_poisson <- glm(
  QuestionMarks ~ Language,
  family = poisson(),
  offset = offset,
  data = Q_Mark_Detective
)

cat("=== Poisson Regression Summary ===\n")
summary(model_poisson)

#Check overdispersion (Poisson )
pearson_chi2 <- sum(
  residuals(model_poisson, type = "pearson")^2)
dispersion <- pearson_chi2 / df.residual(model_poisson)

cat("\nPoisson Dispersion:", dispersion, "\n")

# Use glm.nb to estimate theta
theta_hat_Detective <- glm.nb(QuestionMarks ~ Language + offset(offset),
  data = Q_Mark_Detective)$theta

# Negative Binomial regression

model_nb_glm_Detective <- glm(QuestionMarks ~ Language, family =
negative.binomial(theta = theta_hat_Detective), offset = offset,
  data = Q_Mark_Detective)

summary(model_nb_glm_Detective)

#Check overdispersion (Negative binomial regression )
pearson_chi2_nb_detective <- sum(
  residuals(model_nb_glm_Detective, type = "pearson")^2)
dispersion_nb_detective <- pearson_chi2_nb_detective /
df.residual(model_nb_glm_Detective)
cat("\nNB Dispersion:", dispersion_nb_detective, "\n")

```

```

## Poisson regression for the distribution of question marks by genres (Fantasy)

# Create dataset (Fantasy genre )
Q_Mark_Fantasy <- data.frame(
  Language = c(
    "Mandarin", "Mandarin", "Mandarin", "Mandarin",
    "French", "French", "French"),
  TextSize = c(99742, 73973, 98816, 56281, 99591, 70879, 81271), QuestionMarks = c(
    685, 519, 483, 249, 657, 373, 236))

Q_Mark_Fantasy$Language <- factor(
  Q_Mark_Fantasy$Language,
  levels = c("Mandarin", "French")
)

# offset (log TextSize )

Q_Mark_Fantasy$offset <- log(Q_Mark_Fantasy$TextSize)

# Poisson regression
model_poisson_fantasy <- glm(
  QuestionMarks ~ Language,
  family = poisson(),
  offset = offset,
  data = Q_Mark_Fantasy)

cat("=== Poisson Regression Summary ===\n")
summary(model_poisson_fantasy)

# Check overdispersion (Poisson regression)
pearson_chi2_fantasy <- sum(
  residuals(model_poisson_fantasy, type = "pearson")^2)
dispersion_fantasy <- pearson_chi2_fantasy / df.residual(model_poisson_fantasy)

cat("\\nPoisson Dispersion:", dispersion_fantasy, "\\n")

# Use glm.nb to estimate theta
theta_hat_Fantasy <- glm.nb(
  QuestionMarks ~ Language + offset(offset),
  data = Q_Mark_Fantasy)$theta

```

```
# Negative Binomial regression
```

```
model_nb_glm_Fantasy <- glm(  
  QuestionMarks ~ Language,  
  family = negative.binomial(theta = theta_hat_Fantasy),  
  offset = offset,  
  data = Q_Mark_Fantasy)
```

```
summary(model_nb_glm_Fantasy)
```

```
#Check overdispersion (Negative binomial regression )
```

```
pearson_chi2_nb_fantasy <- sum(  
  residuals(model_nb_glm_Fantasy, type = "pearson")^2  
)  
dispersion_nb_fantasy <- pearson_chi2_nb_fantasy /  
df.residual(model_nb_glm_Fantasy)
```

```
cat("\\nNB Dispersion:", dispersion_nb_fantasy, "\\n")
```

```
## Multinomial regression for question type
```

```
#I create two data frames by selecting four variables from data_MD and data_FR, also  
I drop the excluded cases
```

```
data_MD1 <- data_MD %>% filter(Type != "X") %>% dplyr::select(Type, Structure,  
Register, Language, Genre, Work)  
data_FR1 <- data_FR %>% filter(Type != "X") %>% dplyr::select(Type, Structure,  
Register, Language, Genre, Work)
```

```
#Multinomial regression for question type
```

```
model_MD_type <- brm(formula = Type ~ Register + Genre + (1 | Work),  
  data = data_MD1,  
  family = categorical())  
summary(model_MD_type)
```

```
plot(model_MD_type)
```

```
model_FR_type <- brm(formula = Type ~ Register + Genre + (1 | Work),  
  data = data_FR1,  
  family = categorical())
```

```
summary(model_FR_type)
```

*****Chapter5_Parallel analysis*****

```
knitr::opts_chunk$set(echo = TRUE)
# set seed for randomness
set.seed(123)
# remove scientific notations such as 10000=1e04
options(scipen=999,stringsAsFactors = FALSE)
# reading basic packages
library(tidyverse)
library(stringi)
library(readxl)
# read some more packages for plotting
# library(plotly)
library(ggpubr)
library(GGally)
# for mixed models
library(lme4)
#install.packages("car")
library(car)
#install.packages(c("lme4","afex","emmeans","lsmeans","Matrix"))
library(afex)
library(lmerTest)
library(emmeans) # library(lsmeans)
# for dependency parsing
# library(udpipe) ``
install.packages("vcd")
library(vcd)
install.packages("brms")
library(brms)

## Poisson regression for the overall distribution of question marks (Chapter 5
monolingual comparable)

#Import the file for the meta data of author, work, genre, and text size to run the
Poisson regression
Meta_para <- read_csv("Meta_para.csv")
Meta_para <- Meta_para %>% na.omit(Meta_para)
```

```
GLM_poisson_QMark_MD_para <- Meta_para %>% filter(Language == "Taiwan
Mandarin") %>%
```

```
  glmer(Count_translation ~ Genre+ (1 | Work),
        family = poisson,
        offset = log(Total_translation),
        control = glmerControl(optimizer = "bobyqa"),
        data = .)
```

```
GLM_poisson_QMark_FR_para <- Meta_para %>% filter(Language == "French")
%>%
```

```
  glmer(Count_translation ~ Genre + (1 | Work),
        family = poisson,
        offset = log(Total_translation),
        control = glmerControl(optimizer = "bobyqa"),
        data = .)
```

```
summary(GLM_poisson_QMark_MD_para)
```

```
summary(GLM_poisson_QMark_FR_para)
```

```
## Bayesian regression model for question types
```

```
data_all <- read_csv("TMFR_ORTR.csv")
```

```
data_all %>% select("Structure_Simple") %>% filter(data_all$Language ==
c("Taiwan Mandarin")) %>% table()
```

```
data_all %>% select("Structure_Simple") %>% filter(data_all$Language ==
c("French")) %>% table()
```

```
data_all <- data_all %>% filter(Type != "X")
```

```
data_all <- data_all %>%
```

```
  dplyr::mutate(
    Type = factor(Type),
    Structure_Simple = factor(Structure_Simple),
    Register = factor(Register),
    Language = factor(Language),
    Translation_Status = factor(Translation_Status),
    Genre = factor(Genre),
    Work = factor(Work)
  )
```

```
fit_final <-brm(
  Type ~ Register + Language + Translation_Status + Genre + (1 | Work),
  data = data_all,
  family = categorical(),
  chains = 4, cores = 4, iter = 2000,
  control = list(adapt_delta = 0.9),
  refresh = 50)
```

```
saveRDS(fit_final, "fit_final_bf_type.rds")
fit_final <- readRDS("fit_final_bf_type.rds")
fit_final
```

Bayesian regression model for Question types (plot)

```
library(patchwork)
```

```
# create marginal effects
```

```
ce <- conditional_effects(fit_final, effects = c("Register", "Language",
"Translation_Status", "Genre"),categorical = TRUE)
```

```
plot_ce <- function(df, x_var, x_label){
  ggplot(df, aes_string(x = x_var, y = "estimate__", color = "cats__", group =
"cats__")) +
  geom_point(position = position_dodge(width = 0.5), size = 2) +
  geom_line(position = position_dodge(width = 0.5)) +
  geom_errorbar(aes(ymin = lower__, ymax = upper__), width = 0.2,
  position = position_dodge(width = 0.5)) +
  labs(x = x_label, y = "Predicted probability", color = "Type") +
  theme_bw(base_size = 13)
}
```

```
# transform cats__ as factor
```

```
ce$Register$cats__ <- as.factor(ce$Register$cats__)
ce$Language$cats__ <- as.factor(ce$Language$cats__)
ce$Translation_Status$cats__ <- as.factor(ce$Translation_Status$cats__)
ce$Genre$cats__ <- as.factor(ce$Genre$cats__)
```

```
# create four plots
```

```
p_register <- plot_ce(ce$Register, "effect1__", "Register")
p_language <- plot_ce(ce$Language, "effect1__", "Language")
p_translation <- plot_ce(ce$Translation_Status, "effect1__", "Translation Status")
p_genre <- plot_ce(ce$Genre, "effect1__", "Genre")
```

```

# combine four plots
Thesis_ch5_Fig50 <- (p_register | p_language) / (p_translation | p_genre)

# save plots
ggsave(plot = Thesis_ch5_Fig50, file = "Thesis_ch5_Fig13.png", width = 12, height =
8, dpi = 300)

## Bayesian regression model for Question structures in Taiwan Mandarin

data_all_TM_STC_TYP <- data_all %>% filter(Language == "Taiwan Mandarin",
S_structure == "Typical")

bf_struct_TM_STC_TYP <- bf(Structure_Simple ~ Register + Translation_Status +
Genre + (1 | Work))

fit_struct <- brm(
  bf_struct_TM_STC_TYP,
  data = data_all_TM_STC_TYP,
  family = categorical(),
  chains = 4,
  cores = 4,
  iter = 2000,
  control = list(adapt_delta = 0.9),
  refresh = 50
)

saveRDS(fit_struct, "fit_final_bf_struct_TM.rds")
fit_struct <- readRDS("fit_final_bf_struct_TM.rds")
fit_struct

## Bayesian regression model for Question Structures_TM (plot)

# create marginal effects
ce_str_tm <- conditional_effects(fit_struct,
  effects = c("Register", "Translation_Status", "Genre"),
  categorical = TRUE)

plot_ce_structures <- function(df, x_var, x_label){
  df$cats__ <- as.factor(df$cats__)
  df[[x_var]] <- as.factor(df[[x_var]])
  # setting dodge width
  dodge_width <- 0.8
  ggplot(df, aes_string(x = x_var, y = "estimate__", color = "cats__", group = "cats__"))
  +

```

```

geom_point(position = position_dodge(width = dodge_width), size = 2) +
geom_line(position = position_dodge(width = dodge_width)) +
geom_errorbar(aes(ymin = lower__, ymax = upper__), width = 0.2, position =
position_dodge(width = dodge_width)) +
scale_color_manual(values = c(
"#E41A1C",
"#377EB8",
"#4DAF4A",
"#984EA3",
"#FF7F00",
"#A65628",
"#F781BF",
"#999999"
))+
labs(x = x_label, y = "Predicted probability", color= "Structures") +
theme_bw(base_size = 20)+
theme(legend.position = "none")
}

# Create three plots
p_register <- plot_ce_structures(ce_str_tm$Register, "effect1__", "Register")
p_translation <- plot_ce_structures(ce_str_tm$Translation_Status, "effect1__",
"Translation Status")
p_genre <- plot_ce_structures(ce_str_tm$Genre, "effect1__", "Genre")

legend_plot <- ggplot(ce_str_tm$Register, aes(x = effect1__, y = estimate__, color =
cats__)) +
geom_point() +
labs(color = "Structure") +
theme(legend.title = element_text(size = 20),
legend.text = element_text(size = 20)
)+ theme_void()

combined <- (p_register / p_translation / p_genre) | guide_area()

Thesis_ch5_Fig51<- combined + plot_layout(guides = "collect") &
theme(legend.position = "right")

Save plot
ggsave(plot = Thesis_ch5_Fig51, file = "Thesis_ch5_Fig51.png", width = 10, height =
12, dpi = 300)

```

Bayesian regression model for Question structures in French

```
data_all_FR_STC_TYP <- data_all %>%  
  filter(Language == "French", S_structure == "Typical") %>%  
  mutate(S_structure = factor(S_structure)) %>%  
  droplevels()  
  
bf_struct_FR_STC_TYP <- bf(Structure_Simple ~ Register + Translation_Status +  
  Genre + (1 | Work))
```

```
fit_struct_FR <- brm(  
  bf_struct_FR_STC_TYP,  
  data = data_all_FR_STC_TYP,  
  family = categorical(),  
  chains = 4,  
  warmup = 1000,  
  cores = 4,  
  iter = 4000,  
  control = list(adapt_delta = 0.9),  
  refresh = 50  
)
```

```
saveRDS(fit_struct_FR, "fit_final_bf_struct_FR.rds")  
fit_struct_FR <- readRDS("fit_final_bf_struct_FR.rds")  
fit_struct_FR
```

Bayesian regression model for Question Structures_FR (plot)

```
ce_str_FR <- conditional_effects(  
  fit_struct_FR,  
  effects = c("Register", "Translation_Status", "Genre"),  
  categorical = TRUE  
)
```

```
plot_ce_structures <- function(df, x_var, x_label){  
  df$cats__ <- as.factor(df$cats__)  
  pd <- position_dodge(width = 0.8)  
  
  ggplot(df, aes_string(x = x_var, y = "estimate__", color = "cats__", group =  
    "cats__")) +  
    geom_point(position = pd, size = 2) +  
    geom_line(position = pd) +  
    geom_errorbar(aes(ymin = lower__, ymax = upper__), width = 0.2, position = pd) +
```

```

scale_color_manual(values = c(
  "#E41A1C", "#377EB8", "#4DAF4A", "#984EA3",
  "#FF7F00", "#A65628", "#F781BF", "#999999"
)) +
labs(x = x_label, y = "Predicted probability", color = "Structures") +
theme_bw(base_size = 20) +
theme(legend.position = "none")
}

p_register_FR <- plot_ce_structures(ce_str_FR$Register, "effect1__", "Register")
p_translation_FR <- plot_ce_structures(ce_str_FR$Translation_Status, "effect1__",
"Translation Status")
p_genre_FR <- plot_ce_structures(ce_str_FR$Genre, "effect1__", "Genre")

legend_plot_FR <- ggplot(ce_str_FR$Register, aes(x = effect1__, y = estimate__, color
= cats__)) +
geom_point() +
labs(color = "Structure") +
theme(legend.title = element_text(size = 20),
  legend.text = element_text(size = 20)) +
theme_void()

combined_FR <- (p_register_FR / p_translation_FR / p_genre_FR) | guide_area()

Thesis_ch5_Fig52_FR <- combined_FR +
plot_layout(guides = "collect") &
theme(legend.position = "right")

ggsave(
  plot = Thesis_ch5_Fig52_FR,
  file = "Thesis_ch5_Fig52_FR.png",
  width = 10, height = 12, dpi = 300
)

```

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

References

- Abeillé, Anne, Benoît Crabbé & Danièle Godard Jean-Marie Marandin. 2012. French questioning declaratives: a corpus study. In *Proceedings of the 16th workshop on the semantics and pragmatics of dialogue*. Paris, France: SEMDIAL.
- Abeillé, Anne & Danièle Godard. 2021. *La grande grammaire du français*. Arles [Paris]: Actes Sud Imprimerie nationale éditions.
- Ädel, Annelie. 2020. Corpus compilation. In Magali Paquot & Stefan Th. Gries (eds.), *A practical handbook of corpus linguistics*, 3–24. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-46216-1_1.
- Adli, Aria. 2006. French *wh*-in-situ questions and syntactic optionality: Evidence from three data types. *Zeitschrift für Sprachwissenschaft* 25(2). 163–203. <https://doi.org/10.1515/ZFS.2006.007>.
- Adli, Aria. 2015. What you like is not what you do: Acceptability and frequency in syntactic variation. In Aria Adli, Marco García García & Göz Kaufmann (eds.), *Variation in language: System- and usage-based approaches*, 173–200. Berlin, München, Boston: De Gruyter. <https://doi.org/10.1515/9783110346855-008>.
- Aijmer, Karin. 2008. Parallel and comparable corpora. In Anke Lüdeling & Merja Kytö (eds.), *Corpus linguistics. An international handbook*, vol. 1, 275–292. New York: De Gruyter Mouton.
- Aijmer, Karin. 2024. The present status and recent trends in corpus-based contrastive linguistics. In Anna Cermáková, Hilde Hasselgård, Markéta Malá & Denisa Šebestová (eds.), *Contrastive corpus linguistics: patterns in lexicogrammar and discourse*, 11–25. 1st edn. London: Bloomsbury Academic. <https://doi.org/10.5040/9781350385962>.
- Aijmer, Karin & Diana Lewis. 2017. Introduction. In Karin Aijmer & Diana Lewis (eds.), *Contrastive analysis of discourse-pragmatic aspects of linguistic genres* (Yearbook of Corpus Linguistics and Pragmatics), vol. 5, 1–9. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-54556-1_1.
- Alleton, Viviane. 1981. Final particles and expression of modality in Modern Chinese. *Journal of Chinese Linguistics* 9(1). 91–115.
- Alleton, Viviane. 1988. The so-called “Rhetorical Interrogation” in Mandarin Chinese. *Journal of Chinese Linguistics* 16(2). 278–297.
- Altenberg, Bengt. 1999. Adverbial connectors in English and Swedish: Semantic and lexical correspondences. In Hilde Hasselgård & Signe Oksefjell (eds.), *Out of corpora: studies in honour of Stig Johansson*, 249–268. Leiden: Brill. https://doi.org/10.1163/9789004653689_022.

- Altenberg, Bengt. 2001. Review of Ebeling (2000): presentative constructions in English and Norwegian. A Corpus-based Contrastive Study. *Languages in Contrast* 3(1). 149–156. <https://doi.org/10.1075/lic.3.1.08alt>.
- Altenberg, Bengt & Sylviane Granger. 2002. Recent trends in cross-linguistic lexical studies. In *Studies in Corpus Linguistics*, 3–48. Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/scl.7.04alt>.
- Aoun, Joseph. 1985. *Generalized binding: The syntax and logical form of wh-interrogatives*. Berlin, Boston: De Gruyter Mouton. <https://doi.org/10.1515/9783111392851>.
- Aoun, Joseph & Yen-hui Audrey Li. 1993. *Wh*-Elements in situ: syntax or LF? *Linguistic Inquiry* 24(2). 199–238.
- Artstein, Ron & Massimo Poesio. 2008. Inter-coder agreement for computational linguistics. *Computational Linguistics*. 34(4). 555–596.
- Auger, Julie & Anne-José Villeneuve. 2021. Étude comparative des particules interrogatives en picard et dans deux variétés de français parlées au Canada. *Langue française* 212(4). 57–74. <https://doi.org/10.3917/lf.212.0057>.
- Austin, John Langshaw. 1962. *How to do things with words*. (Ed.) Marina Sbisa & J. O. Urmson. Oxford [Eng.]: Clarendon Press.
- Axelsson, Karin. 2009. Tag Questions in Translations between English and Swedish. In B. J. Epstein (ed.), *Northern lights: translation in the nordic countries*, 81–106. Oxford: Peter Lang.
- Axelsson, Karin. 2011. A cross-linguistic study of grammatically-dependent question tags: Data and theoretical implications. *Studies in Language* 35(4). 793–851. <https://doi.org/10.1075/sl.35.4.02axe>.
- Axelsson, Karin. 2020. Questions in English and Swedish fiction texts: A study based on parallel and comparable corpus data. *Languages in Contrast* 20(2). 235–262. <https://doi.org/10.1075/lic.00017.axe>.
- Bailey, Laura. 2010. Sentential word order and the syntax of question particles. *Newcastle Working Papers in Linguistics* 16. 23–43.
- Bailey, Laura R. 2015. Word order and the syntax of question particles. In Sylvie Hancil, Alexander Haselow & Margje Post (eds.), *Final particles*, 407–426. Berlin, München, Boston: De Gruyter Mouton. <https://doi.org/10.1515/9783110375572-017>.
- Bailey, Laura Rudall. 2013. The syntax of question particles. Newcastle upon Tyne: Newcastle University PhD dissertation.
- Baker, C. L. 1970. Notes on the description of English questions: The role of an abstract question morpheme. *Foundations of Language* 6(2). 197–219.

- Baker, Mona. 1993. Corpus Linguistics and Translation Studies — Implications and Applications. In Mona Baker, Gill Francis & Elena Tognini-Bonelli (eds.), *Text and Technology*, 233–250. Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/z.64.15bak>.
- Baker, Mona. 1996. The challenges that lie ahead: Corpus-based translation studies. In Harold Somers (ed.), *Terminology, LSP and translation: studies in language engineering in honour of Juan C. Sager* (Benjamins Translation Library), 175–186. Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/btl.18.17bak>.
- Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using **lme4**. *Journal of Statistical Software* 67(1). <https://doi.org/10.18637/jss.v067.i01>.
- Bayer, Josef & Lisa Lai-Shen Cheng. 2017. *Wh*-in-Situ. In Martin Everaert & Henk C. Riemsdijk (eds.), *The Wiley Blackwell companion to syntax, second edition*, 1–44. Wiley. <https://doi.org/10.1002/9781118358733.wbsyncom106>.
- Becher, Viktor. 2011. When and why do translators add connectives?: A corpus-based study. *Target. International Journal of Translation Studies* 23(1). 26–47. <https://doi.org/10.1075/target.23.1.02bec>.
- Bencini, Giulia. 2003. Toward a diachronic typology of yes/no question constructions with Particles. *Proceedings from the annual meeting of the Chicago linguistic society* 39(1). 604–621.
- Benzitoun, Christophe. 2022. Évolution des interrogatives partielles directes en français : Le cas de *combien*. (Ed.) F. Neveu, S. Prévost, A. Steuckardt, G. Bergounioux & B. Hama. *SHS Web of Conferences* 138. 13003. <https://doi.org/10.1051/shsconf/202213813003>.
- Bergez, Daniel, Christiane Lauvergnat-Gagnière, Anne Paupert, Yves Stalloni & Gilles Vannier. 1995. *Précis de littérature française*. Paris: Dunod.
- Bernardini, Silvia. 2004. In the classroom: Corpora in the classroom: An overview and some reflections on future developments. In John McH. Sinclair (ed.), *Studies in corpus linguistics*, vol. 12, 15–36. Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/scl.12.05ber>.
- Bernardini, Silvia. 2022. How to use corpora for translation. In *The Routledge Handbook of Corpus Linguistics*, 485–498. 2nd edn. London: Routledge. <https://doi.org/10.4324/9780367076399-34>.
- Bessard-Banquy, Olivier, Sylvie Ducas & Alexandre Gefen. 2021. *Best-sellers l'industrie du succès* (HORS COLLECTION). Paris: Armand Colin.

- Beyssade, Claire. 2006. La structure de l'information dans les questions : Quelques remarques sur la diversité des formes interrogatives en français. *Linx. Revue des Linguistes de l'Université Paris X Nanterre* (55). 173–193. <https://doi.org/10.4000/linx.470>.
- Bhatt, Rajesh & Veneeta Dayal. 2020. Polar question particles: Hindi-Urdu kya: *Natural Language & Linguistic Theory* 38(4). 1115–1144. <https://doi.org/10.1007/s11049-020-09464-0>.
- Bianchi, Valentina, Giuliano Bocci & Silvio Cruschina. 2017. Two types of subject inversion in Italian *WH*-Questions. *Revue Roumaine de Linguistique* LXII (3). 233–252.
- Biber, Douglas. 1994. Representativeness in corpus design. In Antonio Zampolli, Nicoletta Calzolari & Martha Palmer (eds.), *Current issues in computational linguistics: in Honour of Don Walker*, 377–407. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-0-585-35958-8_20.
- Biber, Douglas. 1995. *Dimensions of register variation: A cross-linguistic comparison*. Cambridge: Cambridge University Press.
- Biber, Douglas & Susan Conrad. 2019. *Register, genre, and style*. 2nd edn. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781108686136>.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman grammar of spoken and written English*. New York: Longman.
- Bickel, Balthasar. 2007. Typology in the 21st century: Major current developments. *Linguistic Typology* 11(1). <https://doi.org/10.1515/LINGTY.2007.018>.
- Blanche-Benvéniste, Claire. 1997. A propos de : qu'est-ce que c'est et c'est quoi. *Recherches sur le Français Parlé* 14. 127–146.
- Blum-Kulka, Shoshana. 1987. Indirectness and politeness in requests: Same or different? *Journal of Pragmatics* 11(2). 131–146. [https://doi.org/10.1016/0378-2166\(87\)90192-5](https://doi.org/10.1016/0378-2166(87)90192-5).
- Blum-Kulka, Shoshana & Elite Olshtain. 1984. Requests and apologies: A cross-cultural study of speech act realization patterns (CCSARP). *Applied Linguistics* 5(3). 196–213.
- Boas, Franz. 1911. *Handbook of American Indian languages*. Washington, DC: Government Printing Office.
- Boeckx, Cedric. 2012. *Syntactic islands: key topics in syntax*. 1st edn. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139022415>.
- Bošković, Željko. 1998. LF movement and the minimalist program. *Proceedings of NELS* 28, vol. 1, 43–58.

- Boucher, Paul. 2010a. L'interrogation partielle en français : L'interface syntaxe / sémantique. *Syntaxe & Sémantique* 11(1). 55–82. <https://doi.org/10.3917/ss.011.0055>.
- Boucher, Paul. 2010b. *Wh*-questions in French and English: Mapping syntax to information structure. In Carsten Breul & Edward Göbbel (eds.), *Linguistik aktuell/Linguistics today*, vol. 165, 101–138. Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/la.165.05bou>.
- Bowe, Heather & Kylie Martin (eds.). 2007. Speech acts and politeness across cultures. In *Communication across cultures: Mutual understanding in a global world*, 46–60. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511803925.006>.
- Brown, Penelope & Stephen C. Levinson. 1987. *Politeness: some universals in language usage* (Studies in Interactional Sociolinguistics). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511813085>.
- Bruening, Benjamin. 2007. *Wh*-in-situ does not correlate with *wh*-indefinites or question particles. *Linguistic Inquiry* 38(1). 139–166.
- Bugarski, Ranko. 1991. Contrastive analysis of terminology and the terminology of contrastive analysis. In Vladimir Ivir & Damir Kalogjera (eds.), *Essays in contact linguistics*, 73–82. Berlin, New York: De Gruyter Mouton. <https://doi.org/doi:10.1515/9783110869118.73>.
- Bürkner, Paul-Christian. 2021. Bayesian item response modeling in *R* with **brms** and *Stan*. *Journal of Statistical Software* 100(5). <https://doi.org/10.18637/jss.v100.i05>.
- Buyse, Lieven. 2017. Question tags in translation: An investigation into the translatability of English question tags into Dutch. *Languages in Contrast* 17(2). 157–182. <https://doi.org/10.1075/lic.17.2.01buy>.
- Cadiot, Pierre. 1988. De quoi *ça* parle? A propos de la référence de *ça*, pronom-sujet. *Français Moderne* 56(3–4). 174–192.
- Caponigro, Ivano & Jon Sprouse. 2019. Rhetorical questions as questions. *Proceedings of Sinn und Bedeutung*. 121–133. <https://doi.org/10.18148/SUB/2007.V11I0.635>.
- Cappelle, Bert & Rudy Loock. 2013. Is there interference of usage constraints?: A frequency study of existential *there is* and its French equivalent *il y a* in translated vs. non-translated texts. *Target. International Journal of Translation Studies* 25(2). 252–275. <https://doi.org/10.1075/target.25.2.05cap>.
- Castagnoli, Sara & Marta Kajzer-Wietrzny. 2024. Corpus Studies. In. Zenodo. <https://doi.org/10.5281/ZENODO.10938944>.

- Čermáková, Anna, Jarmo Jantunen, Tommi Jauhiainen, John Kirk, Michal Křen, Marc Kupietz & Elaine Uí Dhonnchadha. 2021. The international comparable corpus: Challenges in building multilingual spoken and written comparable corpora. *Research in Corpus Linguistics* 9(1). 89–103. <https://doi.org/10.32714/ricl.09.01.06>.
- Chang, Lisa. 1997. *Wh-in-situ phenomena in French*. Vancouver: University of British Columbia Master thesis. <https://doi.org/10.14288/1.0087736>.
- Chao, Yuen Ren. 1968. *A grammar of spoken Chinese*. Oakland: University of California Press.
- Chappell, Hilary. 1991. Strategies for the assertion of obviousness and disagreement in mandarin: a semantic study of the modal particle *me*. *Australian Journal of Linguistics* 11(1). 39–65. <https://doi.org/10.1080/07268609108599451>.
- Chen, Xu. 2024. *Wh-island effects in Chinese: A formal experimental study*. Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/la.282>.
- Cheng, Lisa L. S. 1991. On the typology of *wh*-questions. Massachusetts Institute of Technology Dissertation.
- Cheng, Lisa Lai-Shen. 2009. *Wh-in-situ, from the 1980s to now*. *Language and Linguistics Compass* 3(3). 767–791. <https://doi.org/10.1111/j.1749-818X.2009.00133.x>.
- Cheng, Lisa Lai-Shen & Cheng-Teh James Huang. 1996. Two types of donkey sentences. *Natural Language Semantics*. 4(2). 121–163.
- Cheng, Lisa Lai-Shen, James Huang & Jane Tang. 1996. Negative particle questions: A dialectal comparison. In James R. Black & Virginia Motapanyane (eds.), *Current issues in linguistic theory*, vol. 139, 41. Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/cilt.139.04che>.
- Cheng, Lisa Lai-Shen & Johan Rooryck. 2000. Licensing *Wh*-in-situ. *Syntax* 3(1). 1–19. <https://doi.org/10.1111/1467-9612.00022>.
- Cheng, Robert L. 1977. Taiwanese question particles. *Journal of Chinese Linguistics* 5(2). 153–185.
- Cheng, Robert L. 1984. Chinese question forms and their meanings. *Journal of Chinese Linguistics* 12(1). 86–147.
- Cheng, Robert L. 1985. A comparison of Taiwanese, Taiwan Mandarin, and Peking Mandarin. *Language*. JSTOR 352–377.
- Chesterman, Andrew. 1998. *Contrastive functional analysis* (Pragmatics & Beyond New Series). Vol. 47. Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/pbns.47>.

- Chesterman, Andrew. 2004. Beyond the particular. In Anna Mauranen & Pekka Kujamäki (eds.), *Translation Universals Do they exist?*, 33–49. Amsterdam: John Benjamins Publishing Company. <https://doi.org/doi:10.1075/btl.48.04che>.
- Cheung, Candice Chi-Hang. 2014. *Wh*-fronting and the left periphery in Mandarin. *Journal of East Asian Linguistics* 23(4). 393–431. <https://doi.org/10.1007/s10831-013-9112-4>.
- Chiu, Kuei-fen. 2018. “Worlding” world literature from the literary periphery. *Modern Chinese Literature and Culture* 30(1). 13–41.
- Chomsky, Noam. 1995. *The minimalist program* (Current Studies in Linguistics 28). Cambridge, Massachusetts: The MIT Press.
- Chomsky, Noam. 2001. Derivation by phase. In Michael Kenstowicz (ed.), *Ken Hale: a life in language*, vol. 36. Cambridge, Massachusetts: The MIT Press. <https://doi.org/10.7551/mitpress/4056.003.0004>.
- Chomsky, Noam. 2015. *Aspects of the theory of syntax*. 50th Anniversary edition. Cambridge, Massachusetts: The MIT Press.
- Chu, Chauncey. 2009. Relevance and the discourse functions of Mandarin utterance-final modality particles. *Language and Linguistics Compass* 3(1). 282–299. <https://doi.org/10.1111/j.1749-818X.2008.00105.x>.
- Chu, Chauncey C. 1998. *A discourse grammar of Mandarin Chinese*. New York, Berne: Peter Lang. <https://www.peterlang.com/document/1087506>.
- Chu, Chauncey C. 2002. Relevance theory, discourse markers and the Mandarin utterance-final particle *a/ya*. *Journal-Chinese Language Teachers Association* 37(1). 1–42.
- Chu, Chauncey C. 2006. A contrastive approach to discourse particles —— a case study of the Mandarin UFP Ne 呢. *Wàiguóyǔ (Shànghǎi Wàiguóyǔ Dàxué Xuébào)* [*Journal of Foreign Languages = Journal of the Shanghai Foreign Language University*] (03). 7–29
- Ciardelli, Ivano, Jeroen Groenendijk & Floris Roelofsen. 2018. *Inquisitive semantics*. 1st edn. Oxford, UK: Oxford University Press. <https://doi.org/10.1093/oso/9780198814788.001.0001>.
- Cinque, Guglielmo. 1999. *Adverbs and functional heads: A cross-linguistic perspective* (Oxford Studies in Comparative Syntax). New York: Oxford University Press.
- Cinque, Guglielmo & Richard S. Kayne. 2008. *The Oxford handbook of comparative syntax*. Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780195136517.001.0001>.
- CKIP group. 1998. *Introduction to Sinica Corpus*. Taipei: Academia Sinica.

- Claridge, Claudia. 2008. Historical Corpora. In Anke Lüdeling & Merja Kytö (eds.), *Corpus linguistics: An international handbook*. vol. 1. 242-259. Berlin: De Gruyter.
- Comrie, B., M. Haspelmath, B. Bickel & Max Planck Institute for Evolutional Anthropology. 2008. *The Leipzig glossing rules: Conventions for interlinear morpheme-by-morpheme glosses*. Max Planck Institute for Evolutionary Anthropology.
- Connor, Ulla M. & Ana I. Moreno. 2005. Chapter 10. Tertium comparationis: A vital component in contrastive rhetoric research. In Paul Bruthiaux, Dwight Atkinson, William Eggington, William Grabe & Vaidehi Ramanathan (eds.), *Directions in applied linguistics*, 153–164. Bristol, Blue Ridge Summit: Multilingual Matters. <https://doi.org/10.21832/9781853598500-015>.
- Cornulier, Benoît de. 1974. “Pourquoi” et l’inversion du sujet non clitique. In *Teil I études de syntaxe*, 139–163. Berlin, Boston: Max Niemeyer Verlag. <https://doi.org/doi:10.1515/9783111611761-010>.
- Cosme, Christelle. 2007. Clause linking across languages: A corpus-based study of coordination and subordination in English, French and Dutch. Louvain-la-Neuve: Université Catholique de Louvain PhD dissertation.
- Coveney, Aidan. 1996. *Variability in spoken French: A sociolinguistic study of interrogation and negation* (Elm Bank Modern Language Studies). Exeter: Elm Bank Publications.
- Coveney, Aidan. 2011. L’interrogation directe. *Travaux de linguistique* 63(2). 112–145. <https://doi.org/10.3917/tl.063.0112>.
- Coveney, Aidan Benedict. 1989. Variability in interrogation and negation in spoken French. Newcastle University Dissertation.
- Creissels, Denis. 2006. *Syntaxe générale: Une introduction typologique*. Paris: Lavoisier.
- Croft, William. 2002. *Typology and universals*. 2nd edn. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511840579>.
- Culicover, Peter W. & Ray Jackendoff. 2005. *Simpler syntax*. 1st edn. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199271092.001.0001>.
- Curry, Niall & Angela Chambers. 2017. Questions in English and French research articles in linguistics: A corpus-based contrastive analysis. *Corpus Pragmatics* 1(4). 327–350. <https://doi.org/10.1007/s41701-017-0012-0>.
- Damrosch, David. 2003. *What is world literature?* (Translation/Transnation). Princeton: Princeton University Press.

- De Baets, Pauline, Lore Vandevoorde & Gert De Sutter. 2020. On the usefulness of comparable and parallel corpora for contrastive linguistics. Testing the semantic stability hypothesis. In Renata Enghels, Bart Defrancq & Marlies Jansegers (eds.), *New approaches to contrastive linguistics*, 85–126. Berlin, Boston: De Gruyter Mouton. <https://doi.org/10.1515/9783110682588-004>.
- De Boer, Cees. 1926. L'évolution des formes de l'interrogation en français. *Romania* 52(207). 307–327. <https://doi.org/10.3406/roma.1926.4258>.
- Debaisieux, Jeanne-Marie & Alex Boulton. 2007. *Alors la question c'est...?* Questions pragmatiques et annotation pédagogique des corpus. *Cahiers de l'AFLS* 13(2). 31–59.
- Defrancq, Bart. 2015. Contrasting contrastive approaches. *Languages in Contrast* 15(1). 1–3. <https://doi.org/10.1075/lic.15.1.01def>.
- Dekhissi, Laurie. 2013. Variation syntaxique dans le Français multiculturel du cinéma de banlieue. Exeter: University of Exeter PhD dissertation.
- Del Gobbo, Francesca, Nicola Munaro & Cecilia Poletto. 2015. On sentential particles: A crosslinguistic study. In Sylvie Hancil, Alexander Haselow & Margje Post (eds.), *Final particles*, 359–386. Berlin, München, Boston: De Gruyter Mouton. <https://doi.org/10.1515/9783110375572-015>.
- Deley, Nathalie. 2019. L'indexation amateur des oeuvres littéraires au service des algorithmes de recommandation sur Babelio. In Valérie Croissant (ed.), *L'avis des autres: prescription et recommandation culturelles à l'ère numérique*, 109–130. France: Archives contemporaines.
- De Sutter, Gert & Marie-Aude Lefer. 2020. On the need for a new research agenda for corpus-based translation studies: a multi-methodological, multifactorial and interdisciplinary approach. *Perspectives* 28(1). 1–23. <https://doi.org/10.1080/0907676X.2019.1611891>.
- Diessel, Holger. 2003. The relationship between demonstratives and interrogatives. *Studies in Language* 27(3). 635–655. <https://doi.org/10.1075/sl.27.3.06die>.
- Dixon, R. M. W. 2012. *Basic linguistic theory. vol. 3: Further grammatical topics*. 1. Oxford: Oxford University Press.
- Drijkoningen, Frank. 2010. French inversions in contrast. *Linguisticae Investigationes* 33(2). 165–178. <https://doi.org/10.1075/li.33.2.02dri>.
- Druetta, Ruggero. 2002. *Qu'est-ce tu fais?* État d'avancement de la grammaticalisation de *est-ce que*. Première partie. *Linguae & Rivista di Lingue e Culture Moderne* 1(2). 67–88.
- Druetta, Ruggero. 2011. Les formes interrogatives au début du XXI^e siècle: Évolution ou continuité? *L'Information Grammaticale* 129. 26–34.

- Druetta, Ruggero. 2018. Syntaxe de l'interrogation en français et clivage écrit-oral : Une description impossible ? In Marie-José Béguelin, Aidan Coveney & Alexander Guryev (eds.), *L'interrogative en français* (Sciences Pour La Communication), vol. 124, 19–50. Lausanne, Switzerland: Peter Lang. <https://doi.org/10.3726/b13079>.
- Druetta, Ruggero. 2021. La structure alternative en français: Un type interrogatif à part entière? *Langue Française* (4). 91–105.
- Dryer, Matthew S. 1992. The Greenbergian word order correlations. *Language* 68(1). 81–138. <https://doi.org/10.1353/lan.1992.0028>.
- Dryer, Matthew S. 2013a. Position of polar question particles (v2020.3). (Ed.) Matthew S. Dryer & Martin Haspelmath. *The world atlas of language structures online*. Zenodo. <https://doi.org/10.5281/zenodo.7385533>.
- Dryer, Matthew S. 2013b. Polar questions (v2020.3). (Ed.) Matthew S. Dryer & Martin Haspelmath. *The world atlas of language structures online*. Zenodo. <https://doi.org/10.5281/zenodo.7385533>.
- Dryer, Matthew S. 2013c. Position of interrogative phrases in content questions (v2020.3). (Ed.) Matthew S. Dryer & Martin Haspelmath. *The world atlas of language structures online*. Zenodo. <https://doi.org/10.5281/zenodo.7385533>.
- Dupont, Maïté. 2021. *Conjunctive markers of contrast in English and French: From syntax to lexis and discourse* (Studies in Corpus Linguistics). Vol. 99. Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/scl.99>.
- Ebeling, Jarle. 2000. *Presentative constructions in English and Norwegian. A corpus-based contrastive study*. (Acta Humaniora 68). Oslo: Unipub forlag.
- Ebeling, Jarle & Signe Oksefjell Ebeling. 2015. An English-Norwegian contrastive analysis of downtoners, more or less. *Nordic Journal of English Studies* 14(1). 62–89. <https://doi.org/10.35360/njes.340>.
- Ebeling, Signe Oksefjell & Jarle Ebeling. 2020. Dialogue vs. narrative in fiction: A cross-linguistic comparison. *Languages in Contrast* 20(2). 288–313. <https://doi.org/10.1075/lic.00019.oks>.
- Egan, Thomas. 2013. Tertia comparationis in multilingual corpora. In Karin Aijmer & Bengt Altenberg (eds.), *Studies in Corpus Linguistics*, vol. 54, 7–24. Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/scl.54.03ega>.
- Egbert, Jesse & Michaela Mahlberg. 2020. Fiction – one register or two?: Speech and narration in novels. *Register Studies*. 2(1). 72–101. <https://doi.org/10.1075/rs.19006.egb>.
- Einman, Maria. 2018. « Ce n'est pas un livre, mais une pièce de théâtre... »: La lecture de Harry Potter et l'enfant maudit par les utilisateurs du réseau social littéraire Babelio. *Interlitteraria* 23(1). 189–202.

- Enfield, N.J., Tanya Stivers & Stephen C. Levinson. 2010. Question–response sequences in conversation across ten languages: An introduction. *Journal of Pragmatics* 42(10). 2615–2619. <https://doi.org/10.1016/j.pragma.2010.04.001>.
- Engdahl, Elisabet. 2006. Information packaging in questions. *Empirical issues in syntax and semantics* 6(1). 93–111.
- Erlewine, Michael Yoshitaka. 2017. Low sentence-final particles in Mandarin Chinese and the Final-over-Final Constraint. *Journal of East Asian Linguistics* 26(1). 37–75. <https://doi.org/10.1007/s10831-016-9150-9>.
- Fairon, Cédric, Anne-Catherine Simon & Maurice Grevisse. 2018. *Le petit bon usage de la langue française: Grammaire* (Grevisse langue française). Louvain-la-Neuve: De Boeck supérieur.
- Fang, Hongmei. 2018. Mirativity in Mandarin: The sentence-final particle Le (了). *Open Linguistics* 4(1). 589–607. <https://doi.org/10.1515/opli-2018-0029>.
- Fang, Hongmei & Kees Hengeveld. 2020. A mitigator in Mandarin: The sentence-final particle *ba* (吧). *Open Linguistics* 6(1). 284–306. <https://doi.org/10.1515/opli-2020-0018>.
- Fang, Hongmei & Kees Hengeveld. 2022. Sentence-final particles in Mandarin. *Studia Linguistica* 76(3). 873–913. <https://doi.org/10.1111/stul.12198>.
- Feng, Zhiwei. 2006. Evolution and present situation of corpus research in China. *International Journal of Corpus Linguistics* 11(2). 173–207. <https://doi.org/10.1075/ijcl.11.2.03fen>.
- Fiengo, Robert, C.-T. James Huang, Howard Lasnik & Tanya Reinhart. 1988. The syntax of *wh*-in-situ. *Proceedings of WCCFL* 7, 81–98.
- Fliessbach, Jan & Johanna Rockstroh. 2024. Verb influence on French *wh*-placement: A parallel corpus study. *Corpus Linguistics and Linguistic Theory*. <https://doi.org/10.1515/cllt-2024-0001>.
- Flöck, Ilka & Ronald Geluykens. 2015. Speech acts in corpus pragmatics: A quantitative Contrastive Study of Directives in Spontaneous and Elicited Discourse. In Jesús Romero-Trillo (ed.), *Yearbook of corpus linguistics and pragmatics 2015*, vol. 3, 7–37. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-17948-3_2.
- Foulet, Lucien. 1921. Comment ont évolué les formes de l'interrogation. *Romania* 47(186). 243–348. <https://doi.org/10.3406/roma.1921.4437>.
- François, Thomas & Marie-Aude Lefer. 2022. Revisiting simplification in corpus-based translation studies: Insights from readability research. *Meta* 67(1). 50–70. <https://doi.org/10.7202/1092190ar>.

- Frankenberg-Garcia, Ana. 2009. Are translations longer than source texts?: A corpus-based study of explicitation. In *Corpus use and translating*, 47–58. Amsterdam: John Benjamins Publishing Company.
- Frankenberg-Garcia, Ana & Diana Santos. 2014. Introducing COMPARA, the Portuguese-English parallel corpus. In Ana Frankenberg-Garcia, & Diana Santos (eds.), *Corpora in translator education*, 71–87. London: Routledge.
- Fromaigeat, Emile. 1938. Les formes de l’interrogation en français moderne: Leur emploi, leurs significations et leur valeur stylistique. *Vox Romanica*. 3. 1–47.
- Gasde, Horst-Dieter. 2004. Yes/no questions and A-not-A questions in Chinese revisited. *Linguistics* 42. 293–326.
- Gast, Volker. 2012. Contrastive Linguistics: Theories and Methods. In B. Kortmann (ed.), *Dictionaries of linguistics and communication science: Linguistic theory and methodology*. Berlin: Mouton de Gruyter.
- Ge, Fiona, Jiyoung Park & Paula R. Pietromonaco. 2022. How you talk about it matters: Cultural variation in communication directness in romantic relationships. *Journal of Cross-Cultural Psychology* 53(6). 583–602. <https://doi.org/10.1177/00220221221088934>.
- Geluykens, Ronald. 1988. On the myth of rising intonation in polar questions. *Journal of Pragmatics* 12(4). 467–485. [https://doi.org/10.1016/0378-2166\(88\)90006-9](https://doi.org/10.1016/0378-2166(88)90006-9).
- Gilquin, Gaëtanelle & Sylviane Granger. 2022. Using data-driven learning in language teaching. In Anne O’Keeffe & Michael J. McCarthy (eds.), *The Routledge handbook of corpus linguistics*, 430–442. London: Routledge.
- Gómez González, María De Los Ángeles. 2014. Canonical tag questions in English, Spanish and Portuguese: A discourse-functional study. *Languages in Contrast* 14(1). 93–126. <https://doi.org/10.1075/lic.14.1.06gom>.
- Gómez González, María De Los Ángeles & Susana M. Doval-Suárez. 2005. On contrastive linguistics: Trends, challenges and problems. In Christopher S. Butler, María De Los Ángeles Gómez González & Susana M. Doval-Suárez (eds.), *Pragmatics & beyond new series*, vol. 140, 19–45. Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/pbns.140.05gom>.
- Granger, Sylviane. 2010. Comparable and translation corpora in cross-linguistic research. Design, analysis and applications. *Journal of Shanghai Jiaotong University* 2. 14–21.
- Granger, Sylviane & Marie-Aude Lefer. 2020a. Introduction. In *Translating and comparing languages: corpus-based insights*. (Corpora and Language in Use. Proceedings 6), 11–22. Louvain-la-Neuve: Presses universitaires de Louvain.

- Granger, Sylviane & Marie-Aude Lefer. 2020b. Introduction: A two-pronged approach to corpus-based crosslinguistic studies. *Languages in Contrast* 20(2). 167–183. <https://doi.org/10.1075/lic.00014.int>.
- Granger, Sylviane & Marie-Aude Lefer. 2022. Corpus-based translation and interpreting studies: a forward-looking review. In Sylviane Granger & Marie-Aude Lefer (eds.), *Extending the Scope of Corpus-Based Translation Studies*, 13–41. London: Bloomsbury Publishing.
- Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg (ed.), *Universals of human language*, 58–90. Cambridge, Mass: MIT Press.
- Grevisse, Maurice & André Goosse. 2007. *Le bon usage: grammaire française*. 14th edn. Bruxelles: De Boeck.
- Gries, Stefan Th. 2010. Corpus linguistics and theoretical linguistics: A love–hate relationship? Not necessarily.... *International Journal of Corpus Linguistics* 15. 327–343.
- Groenendijk, J. a. G. & M. J. B. Stokhof. 1984. Studies on the semantics of questions and the pragmatics of answers. Amsterdam: University of Amsterdam PhD dissertation.
- Groenendijk, Jeroen & Martin Stokhof. 1997. Questions. In J. F. A. K. Van Benthem, Johan van Benthem & Alice G. B. Ter Meulen (eds.), *Handbook of logic and language*, 1059–1131. Massachusetts: The MIT Press.
- Gryllia, Stella, Jenny S. Doetjes, Yang Yang & Lisa L. Cheng. 2020. Prosody, clause typing, and *wh*-in-situ: Evidence from Mandarin. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 11(1). 19. <https://doi.org/10.5334/labphon.169>.
- Guillot-Barbance, Céline, Bénédicte Pincemin & Alexei Lavrentiev. 2018. Représentation de l’oral en français médiéval et genres textuels. *Langages* 208(4). 53–68. <https://doi.org/10.3917/lang.208.0053>.
- Gunlogson, Christine. 2008. A question of commitment. *Belgian Journal of Linguistics* 22. 101–136. <https://doi.org/10.1075/bjl.22.06gun>.
- Guryev, A. 2017. La forme des interrogatives dans le corpus suisse de SMS en français: étude multidimensionnelle. Université de la Sorbonne Nouvelle (France) & Université de Neuchâtel (Suisse) PhD dissertation.
- Guryev, Alexander & François Delafontaine. 2022. Étude de trois variantes de l’interrogation totale en français: pour une approche en termes de ‘potentiel interactif.’ *Lingvisticae Investigationes* 45(1). 52–82. <https://doi.org/10.1075/li.00067.gur>.

- Guryev, Alexander & François Delafontaine. 2023. L'interrogative *in situ* à la lumière des principes de 'End-Weight' et 'End-Focus.' *Journal of French Language Studies* 33(3). 299–323. <https://doi.org/10.1017/S0959269523000145>.
- Gussenhoven, C. & A. J. Chen. 2000. Universal and language-specific effects in the perception of question intonation. In *Proceedings of the 6th international conference on spoken language processing (ICSLP)*, 91–94. Beijing.
- Haan, Judith & Vincent J. Van Heuven. 2003. This is a *yes/no* -question? *Linguistics in the Netherlands* 20. 59–70. <https://doi.org/10.1075/avt.20.09haa>.
- Hall, Edward T. 1976. *Beyond culture*. New York: Anchor Books/Doubleday.
- Halliday, M.A.K. & Edward McDonald. 2004. Metafunctional profile of the grammar of Chinese. In Alice Caffarel, J.R. Martin & Christian M.I.M. Matthiessen (eds.), *Current issues in linguistic theory*, vol. 253, 305–396. Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/cilt.253.08hal>.
- Halverson, Sandra. 2017. Gravitational pull in translation. Testing a revised model. In Gert De Sutter, Marie-Aude Lefer & Isabelle Delaere (eds.), *Empirical translation studies*, 9–46. Berlin, Boston: De Gruyter Mouton. <https://doi.org/10.1515/9783110459586-002>.
- Halverson, Sandra. 2024. The gravitational pull hypothesis: status and prospects. In *Towards an empirical verification of the gravitational pull hypothesis. Evidence from the COVALT Corpus*, 21–52. Berlin, Germany: Peter Lang.
- Halverson, Sandra. & Haidee Kotze. 2024. Linguistics. In. Zenodo. <https://doi.org/10.5281/ZENODO.10943349>.
- Hamblin, C. L. 1973. Questions in Montague English. *Foundations of Language*. Springer 10(1). 41–53.
- Hamblin, C.L. 1958. Questions. *Australasian Journal of Philosophy* 36(3). 159–168. <https://doi.org/10.1080/00048405885200211>.
- Han, Yang Saxena. 1995. A pragmatic analysis of the BA particle in Mandarin Chinese. *Journal of Chinese Linguistics* 23(2). 99–128.
- Hancil, Sylvie, Margje Post & Alexander Haselow. 2015. 1. Introduction: Final particles from a typological perspective. In Sylvie Hancil, Alexander Haselow & Margje Post (eds.), *Final particles*, 3–36. Berlin, München, Boston: De Gruyter Mouton. <https://doi.org/10.1515/9783110375572-001>.
- Hansen, Maj-Britt Mosegaard. 1997. Alors and donc in spoken French: A reanalysis. *Journal of Pragmatics* 28(2). 153–187. [https://doi.org/10.1016/S0378-2166\(96\)00086-0](https://doi.org/10.1016/S0378-2166(96)00086-0).

- Hansen, Maj-Britt Mosegaard. 2001. Syntax in interaction: Form and function of yes/no interrogatives in spoken standard French. *Studies in Language* 25(3). 463–520. <https://doi.org/10.1075/sl.25.3.04mos>.
- Harris, Martin. 1998. French. In Martin Harris & Nigel Vincent (eds.), *The Romance languages*. 1st edn. London: Routledge. <https://doi.org/10.4324/9780203426531>.
- Haspelmath, Martin. 2007. Coordination. In Timothy Shopen (ed.), *Language typology and syntactic description: Volume 2: Complex constructions*, vol. 2, 1–51. 2nd edn. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511619434.001>.
- Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in crosslinguistic studies. *Language* 86(3). 663–687. <https://doi.org/10.1353/lan.2010.0021>.
- Haspelmath, Martin. 2014. Comparative syntax. In Andrew Carnie, Yosuke Sato & Dan Siddiqi (eds.), *The Routledge handbook of syntax*, 490–508. New York: Routledge.
- Hasselgård, Hilde. 2010. Parallel corpora and contrastive studies. In Richard Xiao (ed.), *Proceedings of the International symposium on using corpora in contrastive and translation studies (UCCTS2010)*, 1–30. Lancaster University.
- Hasselgård, Hilde. 2020. Corpus-based contrastive studies: beginnings, developments and directions. *Languages in Contrast*. Amsterdam: John Benjamins Publishing Company 20(2). 184–208. <https://doi.org/10.1075/lic.00015.has>.
- Hedberg, Nancy, Juan M. Sosa & Emrah Görgülü. 2017. The meaning of intonation in yes-no questions in American English: A corpus study. *Corpus Linguistics and Linguistic Theory* 13(2). 321–368. <https://doi.org/10.1515/cllt-2014-0020>.
- Heine, Bernd, Ulrike Claudi & Friederike Hünemeyer. 1991. *Grammaticalization: A conceptual framework*. Chicago: University of Chicago Press.
- Her, One-Soon, Dewei Che & Adams Bodomo. 2022. On a dichotomy of question types: The case of Mandarin Chinese and Changsha Xiang. *The Linguistic Review* 39(2). 257–291. <https://doi.org/10.1515/tlr-2022-2087>.
- Hölzl, Andreas. 2016. Towards a new typology of questions. *Bavarian Working Papers in Linguistics* (5). 17–28.
- Hölzl, Andreas. 2018. *A typology of questions in northeast Asia and beyond: An ecological perspective*. Berlin: Language Science Press. <https://doi.org/10.5281/zenodo.1344467>.
- House, Juliane & Dániel Z. Kádár. 2021. *Cross-cultural pragmatics*. 1st edn. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781108954587>.
- Hsiao, Pei-Yi & One-Soon Her. 2021. Taxonomy of questions in Taiwan Southern Min. *Concentric* 47(2). 253–299. <https://doi.org/10.1075/consl.00029.her>.

- Hsieh, Yaolung James. 2011. Cross-cultural communication: East vs. West. In Shaoming Zou & Huifen Fu (eds.), *Advances in international marketing*, vol. 21, 283–307. Bingley: Emerald Group Publishing Limited. [https://doi.org/10.1108/S1474-7979\(2011\)0000021015](https://doi.org/10.1108/S1474-7979(2011)0000021015).
- Hsin, Ai-li C. 2016. An analysis of chinese tag questions with a cross-linguistic comparison to English tags. *Taiwan journal of linguistics* 14. 69–119.
- Hsu, Hung-Hsin, Chung-yu Chen & One-Soon Her. Submitted. Justifying a dichotomy of questions in Mandarin: Evidence from a timed grammaticality judgment task.
- Hsu, Hung-Hsin, Marc Allasonnière-Tang & One-Soon Her. In preparation. Justifying a dichotomy of questions in French: Evidence from a timed grammaticality judgment task
- Hu, Xian Yao & Man Zheng. 2024. Corpora and Cross-Linguistic Research of Translation Universals. In *The Routledge Handbook of Corpus Translation Studies*, 215–238. 1st edn. London: Routledge. <https://doi.org/10.4324/9781003184454-16>.
- Huang, Cheng-Teh James. 1982. Logical relations in Chinese and the theory of grammar. Massachusetts: Massachusetts Institute of Technology PhD dissertation.
- Huang, C.-T. James, Y.-H. Audrey Li & Yafei Li. 2009. *The syntax of Chinese* (Cambridge Syntax Guides). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139166935>.
- Huang, Lillian M., Elisabeth Zeitoun, Yeh Marie M, Chang Anna H & Wu Joy J. 1999. Interrogative constructions in some Formosan languages. In Yin Yuen-mei, Yang I-li & Chan Hui-chen (eds.), *Chinese languages and linguistics, vol. 5: Interactions in Language*, 639–680. Taipei: Institute of Linguistics.
- Huang, Po-Hsuan & One-Soon Her. 2024. On a dichotomy of question types: The case of Paiwan. *Concentric. Studies in Linguistics* 50(1). 20–56. <https://doi.org/10.1075/consl.00033.her>.
- Huang, Rui-heng Ray. 2010. Disjunction, coordination, and question: A comparative study. Taipei: National Taiwan Normal University PhD Dissertation.
- Huang, Xiaoli. 2007. Initialness of sentence-final particles in Mandarin Chinese. In *Proceedings of the 21st Pacific Asia conference on language, information and computation*.
- Huddleston, Rodney. 1994. The contrast between interrogatives and questions. *Journal of Linguistics* 30(2). 411–439.
- Idiatov, Dmitry. 2007. A typology of non-selective interrogative pronominals. Antwerp: Universiteit Antwerpen PhD dissertation.

- Izquierdo, Marlén & Zuriñe Sanz-Villar. 2023. Cross-linguistic research and corpora: paving the way for application(s). In Marlén Izquierdo & Zuriñe Sanz-Villar (eds.), *Studies in Corpus Linguistics*, vol. 113, 1–11. Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/scl.113.intro>.
- Jackendoff, Ray. 1974. *Semantic interpretation in generative grammar* (Studies in Linguistics Series). 6th edn. Cambridge: MIT Press.
- James, Carl. 1980. *Contrastive analysis* (Applied Linguistics and Language Study). 13th edn. London: Longman.
- Jensen, Frede. 1990. *Old French and comparative Gallo-Romance syntax*. Berlin, New York: Max Niemeyer Verlag. <https://doi.org/10.1515/9783110938166>.
- Jespersen, Otto. 1933. *Essentials of English grammar*. London: Routledge.
- Johansson, Stig. 1998. On the role of corpora in cross-linguistic research. In Stig Johansson & Signe Oksefjell (eds.), *Corpora and cross-linguistic research: Theory, method, and case studies*, 3–24. Amsterdam & Atlanta, GA: Rodopi.
- Johansson, Stig. 2007. *Seeing through multilingual corpora: On the use of corpora in contrastive studies* (Studies in Corpus Linguistics). Vol. 26. Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/scl.26>.
- Johansson, Stig. 2012. Cross-linguistic perspectives. In Christian Mair, Charles F. Meyer & Nelleke Oostdijk (eds.), *English Corpus Linguistics: Crossing Paths*, 43–68. Leiden, The Netherlands: Brill. https://doi.org/10.1163/9789401207935_005.
- Johns, T. & P. King. 1991. *Classroom concordancing* (ELR Journal). Centre for English Language Studies, the University of Birmingham.
- Karttunen, Lauri. 1977. Syntax and semantics of questions. *Linguistics and Philosophy* 1(1). 382–420. <https://doi.org/10.1007/BF00351935>.
- Kayne, Richard & Jean Yves Pollock. 2001. New thoughts on stylistic inversion. In A. Hulk & J.-Y. Pollock (eds.), *Subject inversion in Romance and the theory of universal grammar*, 107–162. New York: Oxford University Press.
- Kayne, Richard S. 1998. Overt vs. covert movements. *Syntax* 1(2). 128–191. <https://doi.org/10.1111/1467-9612.00006>.
- Kayne, Richard S. 2013. Comparative syntax. *Lingua* 130. 132–151. <https://doi.org/10.1016/j.lingua.2012.10.008>.
- Ke, Ping. 2019. *Contrastive Linguistics* (Peking University Linguistics Research 1). 1st edn. Singapore: Springer Singapore. <https://doi.org/10.1007/978-981-13-1385-1>.
- Kearsley, Greg P. 1976. Questions and question asking in verbal discourse: A cross-disciplinary review. *Journal of Psycholinguistic Research* 5(4). 355–375. <https://doi.org/10.1007/BF01079934>.

- Kendrick, Robin H. 2018. Adjusting epistemic gradients: The final particle *ba* in Mandarin Chinese conversation. *East Asian Pragmatics* 3(1). 5–26. <https://doi.org/10.1558/eap.36120>.
- Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý & Vít Suchomel. 2014. The Sketch Engine: Ten years on. *Lexicography* 1(1). 7–36. <https://doi.org/10.1007/s40607-014-0009-9>.
- Kim, Bryan S. K. & Yong S. Park. 2015. Communication styles, cultural values, and counseling effectiveness with Asian americans. *Journal of Counseling & Development* 93(3). 269–279. <https://doi.org/10.1002/jcad.12025>.
- Kim, Jungsoo & Jong-Bok Kim. 2023. Aggressively non-D-linked construction and ellipsis: a Direct Interpretation approach. *Journal of Linguistics* 59(2). 257–291. <https://doi.org/10.1017/S0022226722000226>.
- Kiss, Angelika. 2024. Form-meaning relations in non-canonical questions. Canada: University of Toronto PhD dissertation.
- Koch, Peter & Wulf Oesterreicher. 2001. 62. Gesprochene Sprache und geschriebene Sprache / Langage parlé et langage écrit. In Günter Holtus, Michael Metzeltin & Christian Schmitt (eds.), *Methodologie (Sprache in der Gesellschaft / Sprache und Klassifikation / Datensammlung und -verarbeitung)*, 584–627. Berlin, New York: Max Niemeyer Verlag. <https://doi.org/10.1515/9783110938371.584>.
- König, Ekkehard. 2012. Contrastive linguistics and language comparison. *Languages in Contrast* 12(1). 3–26. <https://doi.org/10.1075/lic.12.1.02kon>.
- König, Ekkehard & Peter Siemund. 2007. Speech act distinctions in grammar. In Timothy Shopen (ed.), *Language typology and syntactic description: Volume 1: Clause structure*, 276–324. 2nd edn. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511619427.005>.
- Korzen, Hanne. 1990. Pourquoi *pourquoi* est-il différent? L’adverbial de cause et la classification des adverbiaux en général. *Langue Française* 88(1). 60–79. <https://doi.org/10.3406/lfr.1990.5753>.
- Kotze, Haidee & Bertus Van Rooy. 2024. Introduction: The constrained communication framework for studying contact-influenced varieties. In Bertus Van Rooy & Haidee Kotze (eds.), *Contact Language Library*, vol. 60, 1–28. Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/coll.60.01kot>.
- Krifka, Manfred. 2015. Bias in commitment space semantics: Declarative questions, negated questions, and question tags. *Semantics and Linguistic Theory* 25. 328. <https://doi.org/10.3765/salt.v25i0.3078>.
- Krzyszowski, Tomasz P. 1990. *Contrasting languages: the scope of contrastive linguistics*. Berlin, New York: De Gruyter Mouton. <https://doi.org/10.1515/9783110860146>.

- Kubler, Cornelis C. & George Ho. 1984. *Varieties of spoken standard Chinese: volume 2: a speaker from Taipei*. Berlin, New York: De Gruyter Mouton. <https://doi.org/10.1515/9783110882346>.
- Kubler, Cornelius C. 1985. The influence of Southern Min on the Mandarin of Taiwan. *Anthropological Linguistics*. JSTOR 27(2). 156–176.
- Kung, Carmen, Dorothee J. Chwilla & Herbert Schriefers. 2014. The interaction of lexical tone, intonation and semantic context in on-line spoken word recognition: An ERP study on Cantonese Chinese. *Neuropsychologia* 53. 293–309. <https://doi.org/10.1016/j.neuropsychologia.2013.11.020>.
- Kunz, Kerstin & Ekaterina Lapshinova-Koltunski. 2015. Cross-linguistic analysis of discourse variation across registers. *Nordic Journal of English Studies* 14(1). 258–288. <https://doi.org/10.35360/njes.347>.
- Ladd, D. Robert. 2001. 98. Intonation. In *Language typology and language universals*, 1380–1390. Berlin, Boston: De Gruyter Mouton. <https://doi.org/10.1515/9783110194265-035>.
- Lahousse, Karen. 2003a. La distribution de l'inversion nominale en français dans les principales non interrogatives et les subordonnées circonstan. *Lingvisticae Investigationes* 26(1). <https://doi.org/10.1075/li.26.1.10lah>.
- Lahousse, Karen. 2003b. La complexité de la notion de topique et l'inversion du sujet nominal. *Travaux de Linguistique* 47(2). 111–136. <https://doi.org/10.3917/tl.047.0111>.
- Lahousse, Karen. 2006. NP subject inversion in French: two types, two configurations. *Lingua* 116(4). 424–461. <https://doi.org/10.1016/j.lingua.2004.08.020>.
- Lahousse, Karen. 2014. L'inversion nominale dans les phrases simples : Syntaxe et structure de l'information. (Ed.) Franck Neveu, Peter Blumenthal, Linda Hriba, Annette Gerstenberg, Judith Meinschaefer & Sophie Prévost. *SHS Web of Conferences* 8. 2457–2471. <https://doi.org/10.1051/shsconf/20140801390>.
- Landis, J. Richard & Gary G. Koch. 1977. The Measurement of observer agreement for categorical Data. *Biometrics* 33(1). 159. <https://doi.org/10.2307/2529310>.
- Lange, Rael T. 2011. Inter-rater reliability. In Jeffrey S. Kreutzer, John DeLuca & Bruce Caplan (eds.), *Encyclopedia of clinical neuropsychology*, 1348–1348. New York: Springer New York. https://doi.org/10.1007/978-0-387-79948-3_1203.
- Larrivée, Pierre & Alexander Guryev. 2021. Variantes formelles de l'interrogation. Présentation. *Langue Française* 212(4). 9–24. <https://doi.org/10.3917/lf.212.0009>.
- Lasnik, Howard. 2002. The minimalist program in syntax. *Trends in Cognitive Sciences* 6(10). 432–437. [https://doi.org/10.1016/S1364-6613\(02\)01977-0](https://doi.org/10.1016/S1364-6613(02)01977-0).

- Leech, Geoffrey. 1992. Corpora and theories of linguistic performance. In Jan Svartvik (ed.), *Directions in corpus linguistics*. Berlin, New York: De Gruyter Mouton. <https://doi.org/10.1515/9783110867275.105>.
- Leech, Geoffrey N. & Mick Short. 2007. *Style in fiction: A linguistic introduction to English fictional prose* (English Language Series). 2nd edn. London New York: Routledge, Taylor & Francis Group.
- Lefer, Marie-Aude. 2009. Exploring lexical morphology across languages: A corpus-based study of prefixation in English and French writing. Louvain-la-Neuve: Université Catholique de Louvain PhD dissertation.
- Lefer, Marie-Aude. 2020. Parallel Corpora. In Magali Paquot & Stefan Th. Gries (eds.), *A practical handbook of corpus linguistics*, 257–282. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-46216-1_12.
- Lefer, Marie-Aude. 2025. Corpus-Based Methods in Translation Studies. In *Reference Module in Social Sciences*. Elsevier. <https://doi.org/10.1016/B978-0-323-95504-1.01295-3>.
- Lefer, Marie-Aude & Svetlana Vogeleer. 2013. Interference and normalisation in genre-controlled multilingual corpora: Introduction. *Belgian Journal of Linguistics* 27. 1–21. <https://doi.org/10.1075/bjl.27.01lef>.
- Lefevre, Florence. 1999. *La phrase averbale en français* (Collection Langue & parole). Paris, Montréal (Québec): l'Harmattan.
- Lefevre, Florence. 2006. *Quoi de neuf sur “quoi” ? étude morphosyntaxique du mot “quoi”* (Rivages linguistiques). Rennes: Presses universitaires de Rennes.
- Lefevre, Florence. 2018. Les interrogatives partielles averbales : Le cas de « proforme interrogative + ça / cela ». In Marie-José Béguelin, Aidan Coveney & Alexander Guryev (eds.), *L'interrogative en français*, 183–208. Berne: Peter Lang Verlag. <https://doi.org/10.3726/b13079>.
- Lefevre, Florence. 2020. Les interrogatives partielles dans un corpus de théâtre contemporain. *Langages* 217(1). 23–38. <https://doi.org/10.3917/lang.217.0023>.
- Levelt, Willem JM. 1972. Some psychological aspects of linguistic data. *Linguistische Berichte* 17. 18–30.
- Levinson, Stephen C. 2012. Interrogative intimations: On a possible social economics of interrogatives. In Jan P. de Ruiter (ed.), *Questions: formal, functional and interactional perspectives* (Language Culture and Cognition), 11–32. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139045414.003>.
- Levshina, Natalia. 2025. A Bayesian dawn in linguistics: Trends, benefits and good practices. arXiv preprint. <https://doi.org/10.48550/ARXIV.2509.17161>.

- Li, Boya. 2006. Chinese final particles and the syntax of the periphery. Leiden, Netherland: Leiden University PhD dissertation.
- Li, Charles N. & Sandra A. Thompson. 1981. *Mandarin Chinese: A functional reference grammar*. Berkeley, Los Angeles, London: University of California Press.
- Li, Charles N. & Sandra A. Thompson. 1984. Mandarin. In William Chisholm, Louis T. Milic & John A.C. Greppin (eds.), *Typological studies in language*, vol. 4, 47–62. Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/tsl.4.04li>.
- Li, Eden Sum-hung. 2007. *A systemic functional grammar of Chinese: A text-based analysis*. 1st edn. London: Bloomsbury Publishing.
- Li, Haojie & Guicheng Wang. 2018. A study of the development of generative grammar from the perspective of changes in the movement pattern. *Journal of Language Teaching and Research* 9(6). 1277–1283. <https://doi.org/10.17507/jltr.0906.17>.
- Li, Yen-Hui Audrey. 1992. Indefinite *wh* in Mandarin Chinese. *Journal of East Asian Linguistics* 1(2). 125–155.
- Liao, Shengyu & Lei Lei. 2017. What we talk about when we talk about corpus: A bibliometric analysis of corpus-related research in linguistics (2000–2015). *Glottometrics* 38. 1–20.
- Lin, Jo-Wang. 1992. The syntax of *Zenmeyang* “How” and *Weishenme* “Why” in Mandarin Chinese. *Journal of East Asian Linguistics* 1(3). 293–331.
- Lin, Jo-Wang. 2014. *Wh*-expressions in Mandarin Chinese. In C.-T. James Huang, Y.-H. Audrey Li & Andrew Simpson (eds.), *The handbook of Chinese linguistics*, 180–207. 1st edn. Chichester: John Wiley & Sons. <https://doi.org/10.1002/9781118584552.ch8>.
- Lin, Pei-yin & Wen-chi Li (eds.). 2023. *Taiwanese literature as world literature* (Literatures as World Literature). New York: Bloomsbury Academic.
- Lin, Shih-Yueh. 2015. An inquisitive semantics analysis for the Chinese polar question particle *ma*. In *Proceedings of the 20th Amsterdam Colloquium*, 237–245. Amsterdam: University of Amsterdam.
- Liou, M.L.T. 2020. A new perspective on polar questions: The focus mechanism in generating answering patterns in mandarin Chinese and English. Hsinchu: National Tsing Hua University PhD dissertation.
- Liu, Yaobin. 2019. Explicating *wh*-fronting in Mandarin: A scrambling approach. *Proceedings of the Linguistic Society of America* 4(1). 44. <https://doi.org/10.3765/plsa.v4i1.4553>.

- Lombardi Vallauri, Edoardo. 2016. Insubordinated conditionals in spoken and non-spoken Italian. In Nicholas Evans & Honoré Watanabe (eds.), *Insubordination* (Typological Studies in Language), 145–170. Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/tsl.115.06val>.
- Lozano, Cristóbal. 2020. Generative Approaches. In Nicole Tracy-Ventura & Magali Paquot (eds.), *The Routledge handbook of second language acquisition and corpora*, 213–227. New York: Routledge. <https://doi.org/10.4324/9781351137904-19>.
- Luo, Tianhua. 2013. Interrogative strategies. Konstanz: University of Konstanz PhD dissertation.
- Luo, Tianhua & Yicheng Wu. 2017. Toward a typology of question particles in the languages of China. *Lingua* 191–192. 81–93. <https://doi.org/10.1016/j.lingua.2017.03.002>.
- Ly, Annelise. 2016. Internal e-mail communication in the workplace: Is there an “East-West divide”? *Intercultural Pragmatics* 13(1). <https://doi.org/10.1515/ip-2016-0002>.
- Mackenzie, J. Lachlan. 2009. Content interrogatives in a sample of 50 languages. *Lingua* 119(8). 1131–1163. <https://doi.org/10.1016/j.lingua.2007.12.005>.
- Maiden, P.M., C. Robustelli, M. Maiden & C. Robustelli. 2014. *A reference grammar of Modern Italian* (Routledge Reference Grammars). London: Routledge.
- Malá, Markéta. 2007. Czech negative polar questions and their English counterparts in a parallel corpus of Czech and English. In František Štícha & Miriam Fried (eds.), *Grammar and corpora. Gramatika a korpus*, 261–270. Praha: Academia.
- Malá, Markéta. 2010. English declarative yes/no questions as seen through their Czech counterparts. *Acta Universitatis Carolinae. Philologica: Prague Studies in English* 141–151.
- Marchello-Nizia, Christiane, Bernard Combettes, Sophie Prévost & Tobias Scheer (eds.). 2020. *Grande grammaire historique du Français (GGHF)*. Berlin, Boston: De Gruyter Mouton. <https://doi.org/10.1515/9783110348194>.
- Masini, Francesca & Raffaele Simone. 2014. *Word classes* (Current Issues in Linguistic Theory). Amsterdam: John Benjamins Publishing Company.
- Mathieu, Eric. 1999. *WH* in situ and the intervention effect. *UCL Working Papers in Linguistics* 11. 441–472.
- Mathieu, Eric. 2004. The mapping of form and interpretation: The case of optional *wh*-movement in French. *Lingua* 114(9–10). 1090–1132. <https://doi.org/10.1016/j.lingua.2003.07.002>.

- Mauranen, Anna. 2004. Corpora, universals and interference. In Anna Mauranen & Pekka Kujamäki (eds.), *Translation Universals: Do they exist?*, vol. 48, 65–82. Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/btl.48.07mau>.
- Mauri, Caterina. 2008. The irreality of alternatives: Towards a typology of disjunction. *Studies in Language* 32(1). 22–55. <https://doi.org/10.1075/sl.32.1.03mau>.
- McArthur, Tom, Feri McArthur, Madeline Kripke & Indiana University Madeline Kripke Collection of the History of Lexicography (Lilly Library Bloomington). 1992. *The Oxford companion to the English language*. Oxford: Oxford University Press.
- McCawley, James D. 1994. Remarks on the syntax of Mandarin yes-no questions. *Journal of East Asian Linguistics*. Springer 3(2). 179–194.
- McConachy, Troy & Helen Spencer-Oatey. 2021. Cross-cultural and intercultural pragmatics. In Michael Haugh, Dániel Z. Kádár & Marina Terkourafi (eds.), *The Cambridge handbook of sociopragmatics*, 733–757. 1st edn. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781108954105.037>.
- McEnery, Anthony & Zhonghua Xiao. 2007. Parallel and comparable corpora: What are they up to? In Gunilla M. Anderman & Margaret Rogers (eds.), *Incorporating corpora: The linguist and the translator* (Translating Europe), 18–31. Clevedon; Buffalo: Multilingual Matters.
- McEnery, Tony & Andrew Hardie. 2012. *Corpus linguistics: Method, theory and practice* (Cambridge Textbooks in Linguistics). 1st edn. Cambridge: Cambridge University Press.
- McEnery, Tony & Andrew Wilson. 2001. *Corpus linguistics: An introduction* (Edinburgh Textbooks in Empirical Linguistics). 2nd edn. Edinburgh: Edinburgh University Press.
- McEnery, Tony, Richard Xiao & Yukio Tono. 2006. *Corpus-based language studies: An advanced resource book* (Routledge Applied Linguistics). Reprinted. London: Routledge.
- Meier, Ardith J. 2010. Culture and its effect on speech act performance. In Alicia Martínez-Flor & Esther Usó-Juan (eds.), *Language learning & language teaching*, vol. 26, 75–90. Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/llt.26.05mei>.
- Minga, Jamila, Davida Fromm, ClarLynda Williams-DeVane & Brian MacWhinney. 2020. Question use in adults with right-hemisphere brain damage. *Journal of Speech, Language, and Hearing Research* 63(3). 738–748. https://doi.org/10.1044/2019_JSLHR-19-00063.
- Mithun, Marianne. 2012. Tags: Cross-linguistic diversity and commonality. *Journal of Pragmatics* 44(15). 2165–2182. <https://doi.org/10.1016/j.pragma.2012.09.010>.

- Moravcsik, Edith A. 1971. Some cross-linguistic generalizations about yes-no questions and their answers. *Working Papers on Language Universals*, 7, 45-193.
- Morin, Annick. 2017. Questioning particles: A cross-linguistic approach to Quebec French polar interrogatives. Toronto: University of Toronto PhD dissertation.
- Moyer, Morgan C. & Kristen Syrett. 2021. The study of questions. *WIREs Cognitive Science* 12(1). e1531. <https://doi.org/10.1002/wcs.1531>.
- Mulder, Jean & Sandra A. Thompson. 2008. The grammaticization of *but* as a final particle in English conversation. In Ritva Laury (ed.), *Typological studies in language*, vol. 80, 179–204. Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/tsl.80.09mul>.
- Muller, Claude. 2007. Les inversions du sujet et la structure de la proposition en français. In Louis Begioni, André Rousseau, Claude Muller & Université Lille 3 (eds.), *Problèmes de sémantique et de syntaxe: Hommage à André Rousseau* (Collection UL3 Travaux et recherches), 251–272. Villeneuve d'Ascq: Université Charles-de-Gaulle - Lille 3.
- Munaro, Nicola & Jean-Yves Pollock. 2008. *Qu'est-ce-que (qu)-est-ce-que?: A Case study in comparative Romance interrogative Syntax*. Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780195136517.013.0013>.
- Munday, Jeremy, Sara Ramos Pinto & Jacob Blakesley. 2022. *Introducing translation studies: theories and applications*. 5th edn. London: Routledge. <https://doi.org/10.4324/9780429352461>.
- Munteanu, Andrei & Angelika Kiss. 2025. Form-meaning relations in Russian confirmative and surprise declarative questions. *Language and Speech*. <https://doi.org/10.1177/00238309251314862>.
- Myers, Lindsay Lee. 2007. *Wh-interrogatives in spoken French: A corpus-based analysis of their form and function*. Austin: The University of Texas at Austin PhD dissertation.
- Neumann, Stella. 2013. *Contrastive register variation: A quantitative approach to the comparison of English and German* (Trends in Linguistics 251). Berlin Boston: de Gruyter Mouton.
- Nishigauchi, Taisuke. 1986. Quantification in Syntax (Logical form, Japanese, *Wh*-constructions). Amherst: University of Massachusetts PhD dissertation.
- Noh, Eun-Ju. 1998. Echo Questions: Metarepresentation and pragmatic enrichment. *Linguistics and Philosophy* 21(6). 603–628.
- Obenauer, Hans-Georg. 1976. *Etudes de syntaxe interrogative du français: Quoi, combien et le complémenteur. Etudes de syntaxe interrogative du français*. Berlin, Boston: Max Niemeyer Verlag. <https://doi.org/10.1515/9783111340364>.

- Obenauer, Hans-Georg. 1994. Aspects de la syntaxe A-barre-effets d'intervention et mouvements des quantifieurs. Paris: Université de Paris VIII PhD dissertation.
- O'Keeffe, Anne & Michael J. McCarthy. 2022. 'Of what is past, or passing, or to come': Corpus linguistics, changes and challenges. In *The Routledge handbook of corpus linguistics*, 1–9. 2nd edn. London: Routledge. <https://doi.org/10.4324/9780367076399-1>.
- Ordóñez, Francisco. 1998. The inversion construction in interrogatives in Spanish and Catalan. In José Lema & Esthela Treviño (eds.), *Current issues in linguistic theory*, vol. 157, 329. Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/cilt.157.18ord>.
- Oster, Katharina. 2022. Normalisation. In. Zenodo. <https://doi.org/10.5281/ZENODO.6370284>.
- Pak, Miok D. 2008. Types of clauses and sentence end particles in Korean. *Korean Linguistics* 14. 113–156. <https://doi.org/10.1075/kl.14.06mdp>.
- Pan, Victor Junnan. 2018. Derivation of the apparent narrow scope of sentence-final particles in Chinese: A Reply to Erlewine (2017). *Studies in Chinese Linguistics* 39(2). 99–126. <https://doi.org/10.2478/scl-2018-0004>.
- Pan, Victor Junnan. 2019. System repairing strategy at interface: *Wh*-in-situ in Mandarin Chinese. In Jianhua Hu & Haihua Pan (eds.), *Language faculty and beyond*, vol. 15, 133–166. Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/lfab.15.06pan>.
- Pan, Victor Junnan. 2021. Sentence-final particles in Chinese. In *Oxford research encyclopedia of linguistics*. Oxford: Oxford University Press. <https://doi.org/10.1093/acrefore/9780199384655.013.882>.
- Pan, Victor Junnan. 2023. Quantification and *wh*-in-situ in Mandarin Chinese: An Introduction. In Chiara Romagnoli & Linda Badan (eds.), *Lessons on Chinese linguistics* (Collection Des Cahiers de Linguistique – Asie Orientale), vol. 16, 93–132. Paris: CRLAO, EHESS.
- Pan, Victor Junnan & Waltraud Paul. 2016. Why Chinese SFPs are neither optional nor disjunctors. *Lingua* 170. 23–34. <https://doi.org/10.1016/j.lingua.2015.10.005>.
- Pang, Shuangzi & Kefei Wang. 2020. Language contact through translation: The influence of explicitness in English–Chinese translation on language change in vernacular Chinese. *Target. International Journal of Translation Studies* 32(3). 420–455. <https://doi.org/10.1075/target.19001.pan>.
- Paquot, Magali & Stefan Th. Gries (eds.). 2020. *A practical handbook of corpus linguistics*. Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-030-46216-1>.

- Park, Hee Sun, Timothy R. Levine, Rene Weber, Hye Eun Lee, Lucía I. Terra, Isabel C. Botero, Elena Bessarabova, Xiaowen Guan, Sachiyo M. Shearman & Marc Stewart Wilson. 2012. Individual and cultural variations in direct communication style. *International Journal of Intercultural Relations* 36(2). 179–187. <https://doi.org/10.1016/j.ijintrel.2011.12.010>.
- Park, Jiyoung, Yukiko Uchida & Shinobu Kitayama. 2016. Cultural variation in implicit independence: An extension of Kitayama et al. (2009). *International Journal of Psychology* 51(4). 269–278. <https://doi.org/10.1002/ijop.12157>.
- Paroz, R. A. 1946. *Elements of southern Sotho*. Morija, Basutoland: Morija Sesuto Book Depot.
- Pastor, Gloria Corpas & Miriam Seghiri. 2010. Size matters: A quantitative approach to corpus representativeness. In Rosa Rabadán, Trinidad Guzmán & Marisa Fernández López (eds.), *Lengua, traducción, recepción: en honor de Julio César Santoyo* “*Language, translation, reception: To honor Julio César Santoyo*,” 111–145. León: Universidad de León, Área de Publicaciones.
- Paul, Waltraud. 2010. Adjectives in Mandarin Chinese. *Adjectives: Formal analyses in syntax and semantics*. 115–151.
- Paul, Waltraud. 2014. Why particles are not particular: Sentence-final particles in Chinese as heads of a split CP. *Studia Linguistica* 68(1). 77–115. <https://doi.org/10.1111/stul.12020>.
- Paul, Waltraud & Shanshan Yan. 2022. Chapter 7. Sentence-final particles in Mandarin Chinese: Syntax, semantics and acquisition. In Xabier Artiagoitia, Arantzazu Elordieta & Sergio Monforte (eds.), *Linguistik Aktuell/Linguistics Today*, vol. 276, 179–206. Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/la.276.07pau>.
- Payne, Thomas E. 1997. *Describing morphosyntax: A guide for field linguists*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511805066>.
- Pesetsky, David. 1987. *Wh-in-situ: Movement and unselective binding*. In Eric J. Reuland & Alice G. B. Ter Meulen (eds.), *The representation of (in) definiteness*, 98–129. Cambridge, Massachusetts: MIT Press.
- Pino, Angel & Isabelle Rabut. 2024a. La diffusion de la littérature taïwanaise en France à travers anthologies et collections. In Angel Pino & Isabelle Rabut (eds.), *La littérature taïwanaise à la rencontre du monde*, 120–144. Paris: You Feng.
- Pino, Angel & Isabelle Rabut. 2024b. Introduction. In Angel Pino & Isabelle Rabut (eds.), *La littérature taïwanaise à la rencontre du monde*, 11–19. Paris: You Feng.
- Poiret, Rafaël. 2020. Construction of a multi-layer parallel treebank for comparative study of Chinese and French syntax. Zhejiang: Zhejiang University PhD dissertation.

- Poletto, Cecilia & Jean-Yves Pollock. 2004. On the left periphery of some Romance *wh*-questions. In Luigi Rizzi (ed.), *The structure of CP and IP*, 251–296. New York: Oxford University Press. <https://doi.org/10.1093/oso/9780195159486.003.0009>.
- Poletto, Cecilia & Jean-Yves Pollock. 2009. Another look at *wh*-questions in Romance: The case of Mendrisiotto and its consequences for the analysis of French *wh-in situ* and embedded interrogatives. In Danièle Torck & W. Leo Wetzels (eds.), *Current issues in linguistic theory*, vol. 303, 199–258. Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/cilt.303.12pol>.
- Poschmann, Claudia. 2008. All declarative questions are attributive? *Belgian Journal of Linguistics* 22. 247–269. <https://doi.org/10.1075/bjl.22.12pos>.
- Puurtinen, Tiina. 2004. Explication of clausal relations: A corpus-based analysis of clause connectives in translated and non-translated Finnish children's literature. In Anna Mauranten & Pekka Kujamäki (eds.), *Translation Universals: Do they exist?*, 165–176. Amsterdam: John Benjamins Publishing Company. <https://doi.org/doi:10.1075/btl.48.13puu>.
- Pym, Anthony. 2008. On Toury's laws of how translators translate. In Anthony Pym, Miriam Shlesinger & Daniel Simeoni (eds.), *Beyond Descriptive Translation Studies Investigations in homage to Gideon Toury*, vol. 75, 311–328. Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/btl.75.24pym>.
- Quillard, Virginie. 2001. La diversité des formes interrogatives : Comment l'interpréter ? *Langage et Société*. Paris: Éditions de la Maison des sciences de l'homme 95(1). 57–72. <https://doi.org/10.3917/ls.095.0057>.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Neil Leech & David Crystal (eds.). 1985. *A comprehensive grammar of the English language*. 2nd edn. London: Longman.
- R Core Team. 2024. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Ramón García, Noelia. 2021. Contrastive linguistics and translation studies interconnected: The corpus-based approach. *Linguistica Antverpiensia, New Series – Themes in Translation Studies* 1. <https://doi.org/10.52034/lanstts.v1i.27>.
- Reinhardt, Janina. 2019. Regularity and variation in French direct interrogatives: The morphosyntax and intonation of question forms in reality tv shows, audio books and teaching materials. Konstanz: Universität Konstanz PhD dissertation.
- Ren, Fangning. 2023. Mandarin overt *wh*-fronting as focus movement. *Proceedings of the Linguistic Society of America* 8(1). 5491. <https://doi.org/10.3765/plsa.v8i1.5491>.
- Reppen, Randi. 2022. Building a corpus: What are key considerations? In *The Routledge handbook of corpus linguistics*, 13–20. 2nd edn. London: Routledge. <https://doi.org/10.4324/9780367076399-2>.

- Richard Kayne & Jean-Yves Pollock. 1978. Stylistic inversion, successive cyclicity, and move NP in French. *Linguistic Inquiry* 9(4). 595–621.
- Richards, J. C., J. Platt & H. Platt. 1992. *Longman dictionary of language teaching and applied linguistics*. 2nd edn. London: Longman.
- Riegel, Martin, Jean-Christophe Pellat & René Rioul. 2009. *Grammaire méthodique du français*. 4th edn. Paris: Presses Universitaires de France (Puf).
- Rissanen, Matti. 2007. Corpus linguistics and historical linguistics. In Anke Lüdeling & Merja Kytö (eds.), *Corpus linguistics: an International Handbook*, 53–68. Berlin: De Gruyter.
- Rivet, Pierrick. 2024. Soixante années de traductions françaises de la littérature taïwanaise (1962-2022). In Angel Pino & Isabelle Rabut (eds.), *La Littérature taïwanaise à la rencontre du monde*, 167–184. Paris: You Feng.
- Rizzi, Luigi. 1990. *Relativized minimality* (Linguistic Inquiry Monographs 16). Cambridge, Massachusetts: MIT Press.
- Rizzi, Luigi. 2006. On the form of chains: Criterial positions and ECP effects. In Lisa Lai-Shen Cheng & Norbert Corver (eds.), *Wh-movement: Moving on*, 97–134. Cambridge, Massachusetts: The MIT Press.
<https://doi.org/10.7551/mitpress/7197.003.0010>.
- Rizzi, Luigi & Ian Roberts. 1989. Complex inversion in French. *Probus* 1(1). 1–30.
<https://doi.org/10.1515/prbs.1989.1.1.1>.
- Robinson, W. P. & Susan J. Rackstraw. 1977. Children's judgements of the sense of questions. *Research in Education* 17(1). 23–39.
<https://doi.org/10.1177/003452377701700103>.
- Roelofsen, Floris. 2019. Semantic theories of questions. In *Oxford research encyclopedia of linguistics*. Oxford: Oxford University Press.
<https://doi.org/10.1093/acrefore/9780199384655.013.504>.
- Römer, Ute. 2011. Corpus research applications in second language teaching. *Annual Review of Applied Linguistics* 31. 205–225.
<https://doi.org/10.1017/S0267190511000055>.
- Rooth, Mats. 1985. Association with focus. Amherst: University of Massachusetts PhD dissertation.
- Rosemeyer, Malte. 2023. French and Spanish *wh*-interrogatives with and without *wh*. *Journal of French Language Studies* 33(2). 227–255.
<https://doi.org/10.1017/S0959269523000042>.
- Ross, John Robert. 1967. Constraints on variables in syntax. Massachusetts: Massachusetts Institute of Technology PhD dissertation.

- Rossi-Gensane, Nathalie, Luisa Fernanda Acosta Córdoba, Biagio Ursi & Margot Lambert. 2021. Les structures interrogatives directes partielles fondées sur *où* dans les dialogues de romans français du XXe siècle. *Journal of French Language Studies* 31(2). 169–191. <https://doi.org/10.1017/S0959269520000253>.
- Rothman, Jason & Roumyana Slabakova. 2018. The generative approach to SLA and its place in modern second language studies. *Studies in Second Language Acquisition*. Cambridge University Press 40(2). 417–442. <https://doi.org/10.1017/S0272263117000134>.
- Rouquier, Magali. 2002. Les interrogatives en ‘qui/qu’est-ce qui/que’ en ancien français et en moyen français. *Cahiers de grammaire* 27(1). 97–120.
- Rouquier, Magali. 2003. La séquence *est-ce* dans les interrogatives en qui/que en ancien et en moyen français. *Journal of French Language Studies* 13(3). 339–362. <https://doi.org/10.1017/S0959269503001182>.
- Rowlett, Paul. 2007. *The syntax of French*. 1st edn. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511618642>.
- Sadock, Jerrold & Arnold Zwicky. 1985. Speech act distinctions in syntax. In Timothy Shopen (ed.), *Language typology and syntactic description*, vol. 1, 155–196. Cambridge: Cambridge University Press.
- Sadock, Jerry. 2012. Formal features of questions. In Jan P. De Ruiter (ed.), *Questions*, 103–122. 1st edn. Cambridge University Press. <https://doi.org/10.1017/CBO9781139045414.008>.
- Salkie, Raphael, Karin Aijmer & Michael Barlow. 1998. Editorial. *Languages in Contrast* 1(1). v–xii. <https://doi.org/10.1075/lic.1.1.01edi>.
- Sandel, Todd, Hsin-I Yueh & Peih-Ying Lu. 2017. Some distinctive Taiwanese communication practices and their cultural meanings. In Donal Carbaugh (ed.), *The handbook of communication in cross-cultural perspective*, 118–128. New York: Routledge.
- Schulze, Wolfgang. 2007. Communication or memory mismatch?: Towards a cognitive typology of questions. In Günter Radden, Klaus-Michael Köpcke, Thomas Berg & Peter Siemund (eds.), *Aspects of meaning construction*, 247–264. Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/z.136.16sch>.
- Searle, John R. 1976. A classification of illocutionary acts. *Language in Society* 5(1). 1–23.
- Semino, Elena & Mick Short. 2004. *Corpus stylistics*. London: Routledge. <https://doi.org/10.4324/9780203494073>.
- Seyfarth, Robert M., Dorothy L. Cheney & Thore J. Bergman. 2005. Primate social cognition and the origins of language. *Trends in Cognitive Sciences* 9(6). 264–266. <https://doi.org/10.1016/j.tics.2005.04.001>.

- Seyferth, Anne, Aarushi Ratna & Kevin C. Chung. 2022. The art of questioning. *Plastic and Reconstructive Surgery* 149(5). 1031–1035. <https://doi.org/10.1097/PRS.00000000000009064>.
- Shei, Chris & Zhao-Ming Gao. 2017. *The Routledge handbook of Chinese translation*. (Ed.) Chris Shei & Zhao-Ming Gao. 1st edn. London; New York: Routledge. <https://doi.org/10.4324/9781315675725>.
- Shlonsky, Ur. 2012. Notes on *Wh* in situ in French. In Laura Brugé, Anna Cardinaletti, Giuliana Giusti, Nicola Munaro & Cecilia Poletto (eds.), *Functional heads*, 242–252. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199746736.003.0019>.
- Shlonsky, Ur. 2017. *Wh* in situ and criterial freezing. In Enoch Aboh, Eric Haeberli, Genoveva Puskás & Manuela Schönenberger (eds.), *Elements of comparative syntax*, 349–370. Berlin: De Gruyter Mouton. <https://doi.org/10.1515/9781501504037-013>.
- Shlonsky, Ur, Sandra Villata & Julie Franck. 2020. Covert Movement in multiple-*wh* questions: Experimental and theoretical Investigations. *Syntax* 23(2). 185–202. <https://doi.org/10.1111/synt.12192>.
- Siemund, Peter. 2001. Interrogative constructions. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher and Wolfgang Raible (ed.), *Language typology and language universals*, 1010–1028. Berlin: Walter de Gruyter.
- Siemund, Peter (ed.). 2013. Interrogative constructions. In *varieties of English: A typological approach*, 237–257. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139028240.013>.
- Siemund, Peter. 2017. Interrogative clauses in English and the social economics of questions. *Journal of Pragmatics* 119. 15–32. <https://doi.org/10.1016/j.pragma.2017.07.010>.
- Simpson, Andrew. 2014. Sentence-final particles. In C.-T. James Huang, Y.-H. Audrey Li & Andrew Simpson (eds.), *The handbook of Chinese linguistics*, 156–179. Hoboken, USA: John Wiley & Sons, Inc. <https://doi.org/10.1002/9781118584552.ch7>.
- Simpson, Andrew & Tanmoy Bhattacharya. 2003. Obligatory overt *wh*-movement in a *wh*-in-Situ language. *Linguistic Inquiry* 34(1). 127–142. <https://doi.org/10.1162/002438903763255940>.
- Sinclair, John. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press. <http://lib.ugent.be/catalog/rug01:000226040>.
- Sinclair, John. 2004a. Corpus and text — Basic principles. *Developing linguistic corpora: A guide to good practice*. <https://users.ox.ac.uk/~martinw/dlc/chapter1.htm>. (22 September, 2023).

- Smirnova, Anastasiia & Anne Abeillé. 2021. Question particles *ça* and *donc* in French: A corpus study. *Linguistic Research* 38(2). 239–269. <https://doi.org/10.17250/khisli.38.2.202106.003>.
- Solomon, Julie. 1990. Fictional questions: Illocutionary force in literary communication. *Journal of the Australasian Universities Language and Literature Association* 74(1). 85–103. <https://doi.org/10.1179/aulla.1990.007>.
- Spencer-Oatey, Helen & Dániel Kádár. 2016. The bases of (im)politeness evaluations. *East Asian Pragmatics* 1(1). 73–106. <https://doi.org/10.1558/eap.v1i1.29084>.
- Spooren, Wilbert & Liesbeth Degand. 2010. Coding coherence relations: Reliability and validity. *Corpus Linguistics and Linguistic Theory* 6(2). <https://doi.org/10.1515/cllt.2010.009>.
- Stark, Elisabeth & Larissa Binder. 2021. L'inversion du sujet clitique en français oral : ultime apanage des interrogatives ? *Langue française* 212(4). 25–40. <https://doi.org/10.3917/lf.212.0025>.
- Starke, Michal. 2001. Move dissolves into merge: A theory of locality. Geneva: University of Geneva PhD dissertation.
- Stenström, Anna-Brita. 1984. *Questions and responses in English conversation* (Lund Studies in English 68). Malmö, Sweden: CWK Gleerup.
- Stivers, Tanya. 2010. An overview of the question–response system in American English conversation. *Journal of Pragmatics* (Question-Response Sequences in Conversation across Ten Languages) 42(10). 2772–2781. <https://doi.org/10.1016/j.pragma.2010.04.011>.
- Stivers, Tanya & N. J. Enfield. 2010. A coding scheme for question–response sequences in conversation. *Journal of Pragmatics* (Question-Response Sequences in Conversation across Ten Languages) 42(10). 2620–2626. <https://doi.org/10.1016/j.pragma.2010.04.002>.
- Stivers, Tanya, N. J. Enfield, Penelope Brown, Christina Englert, Makoto Hayashi, Trine Heinemann, Gertie Hoymann, et al. 2009. Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*. 106(26). 10587–10592. <https://doi.org/10.1073/pnas.0903616106>.
- Stivers, Tanya & Federico Rossano. 2010. Mobilizing response. *Research on Language and Social Interaction* 43(1). 3–31. <https://doi.org/10.1080/08351810903471258>.
- Tancredi, Christopher. 1990. Syntactic association with focus. In *Proceedings from the first meeting of the Formal Linguistic Society of Mid-America*, 289–303. University of Wisconsin Press Madison.
- Tang, Sze-Wing. 2015. A generalized syntactic schema for utterance particles in Chinese. *Lingua Sinica* 1(1). 3. <https://doi.org/10.1186/s40655-015-0005-5>.

- Tang, Xiaobing. 1999. On the concept of Taiwan literature. *Modern China* 25(4). 379–422.
- Tawalbeh, Ayman & Emran Al-Oqaily. 2012. In-directness and politeness in American English and Saudi Arabic requests: A cross-cultural comparison. *Asian Social Science* 8(10). p85. <https://doi.org/10.5539/ass.v8n10p85>.
- Teich, Elke. 2003. *Cross-linguistic variation in system and text: A methodology for the investigation of translations and comparable texts*. Berlin, Boston: De Gruyter Mouton. <https://doi.org/10.1515/9783110896541>.
- Terry, Robert M. 1970. *Faut-il or Est-ce qu'il faut*: Inversion vs. *est-ce que*. *The French Review* 43(3). 480–482.
- Thompson, Sandra A. & Ryoko Suzuki. 2012. The Grammaticalization of final particles. In Bernd Heine & Heiko Narrog (eds.), *The Oxford handbook of grammaticalization*, 668–680. 1st edn. Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199586783.013.0055>.
- Tiee, Hongye. 1986. *A reference grammar of Chinese sentences: with exercises* (Ed.) Donald M. Lance. 2nd edn. Tucson: University of Arizona Press.
- Togeb, K. 1974. *Précis historique de grammaire française* (Etudes Romanes). Copenhagen: Akademisk Forlag.
- Tognini-Bonelli, Elena. 2001. *Corpus Linguistics at work* (Studies in Corpus Linguistics). Vol. 6. Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/scl.6>.
- Tomioka, Satoshi. 2007. Pragmatics of LF intervention effects: Japanese and Korean *Wh*-interrogatives. *Journal of Pragmatics* 39(9). 1570–1590. <https://doi.org/10.1016/j.pragma.2007.03.002>.
- Torreira, Francisco, Seán G. Roberts & Harald Hammarström. 2014. Functional trade-off between lexical tone and intonation: Typological evidence from polar-question marking. In *Proceeding of 4th International Symposium on Tonal Aspects of Languages (TAL 2014)*, 100–103.
- Toury, Gideon. 1979. Interlanguage and its manifestations in translation. *Meta* 24(2). 223–231. <https://doi.org/10.7202/004502ar>.
- Toury, Gideon. 2012. *Descriptive translation studies – and beyond: Revised edition* (Benjamins Translation Library). 2nd edn. Vol. 100. Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/btl.100>.
- Tsai, Wei-Tien Dylan. 1994a. On nominal islands and LF extraction in Chinese. *Natural Language & Linguistic Theory* 12(1). 121–175.
- Tsai, Wei-Tien Dylan. 1994b. On economizing the theory of A-bar dependencies. Massachusetts: Massachusetts Institute of Technology PhD dissertation.

- Tsai, Wei-Tien Dylan. 2008. Left periphery and how-why alternations. *Journal of East Asian Linguistics* 17(2). 83–115. <https://doi.org/10.1007/s10831-008-9021-0>.
- Tseng, Chin-Chin & Dayoung Jeong. 2020. The use of utterance-final particles in Mandarin teaching. *Chinese as a Second Language Research* 9(2). 259–275. <https://doi.org/10.1515/caslar-2020-0010>.
- Ullan, Russell. 1978. Some general characteristics of interrogative systems. In Joseph H. Greenberg (ed.), *Universal of human language*. Stanford: CA: Stanford University Press.
- Van Rooy, Bertus & Haidee Kotze. 2024. Conclusion: Cumulative insights into constrained communication. In Bertus Van Rooy & Haidee Kotze (eds.), *Contact Language Library*, vol. 60, 255–286. Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/coll.60.09van>.
- Velupillai, Viveka. 2012. *An introduction to linguistic typology*. Amsterdam; Philadelphia: John Benjamins Publishing Company.
- Vinay, Jean-Paul & Jean Darbelnet. 1963. *Stylistique comparée du français et de l'anglais*. Paris: Didier.
- Vondříčka, Pavel. 2014. Aligning parallel texts with InterText. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, 1875–1879. Reykjavik, Iceland: European Language Resources Association (ELRA).
- Wagner, René Léon & Jacqueline Pinchon. 1962. *Grammaire du Français: classique et moderne* (Hachette université Langue française). Paris: Hachette.
- Wang, David Der-wei. 2007. Preface. In David Der-wei Wang & Carlos Rojas (eds.), *Writing Taiwan: A new literary history* (Asia-Pacific), vii–x. Durham (N.C.): Duke university press.
- Wang, Keifei & Hongwu Qin. 2010. A parallel corpus-based study of translational Chinese. In *Using Corpora in Contrastive and Translation Studies*, 164–181. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Wang, Wei. 2021. The question-response system in Mandarin conversation. *Pragmatics* 31(4). 589–616. <https://doi.org/10.1075/prag.20019.wan>.
- Wang, William S-Y. & Chaofen Sun (eds.). 2015. *The Oxford handbook of Chinese linguistics*. Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199856336.001.0001>.
- Watanabe, Akira. 1992. Subjacency and S-structure movement of *wh*-in-situ. *Journal of East Asian Linguistics* 1(3). 255–291. <https://doi.org/10.1007/BF00130554>.

- Watanabe, Akira. 2001. *Wh-in-situ languages*. In Mark Baltin & Chris Collins (eds.), *The handbook of contemporary syntactic theory*, 203–225. 1st edn. Oxford: Blackwell. <https://doi.org/10.1002/9780470756416.ch7>.
- Wei, Ting-Chi. 2013. Fragment question and ellipsis in Chinese. *Studies in Chinese Linguistics* 34(3). 151–198.
- Wei, Ting-Chi. 2022. Fragment questions in Mandarin Chinese: Topic movement and pied-piping. *Language and Linguistics*. 266–305. <https://doi.org/10.1075/lali.00010.wei>.
- Wikberg, Kay. 1996. Questions in English and Norwegian: Evidence from the English-Norwegian parallel Corpus. In Carol E. Percy, Charles F. Meyer & Ian Lancashire (eds.), *Synchronic corpus linguistics: Papers from the Sixteenth International Conference on English Language Research on Computerized Corpora (ICAME 16)*, 17–28. Amsterdam: Rodopi.
- Winter, Bodo. 2020. *Statistics for linguists: An introduction using R*. New York London: Routledge.
- Wright, Sue. 2016. French: The rise and fall of a prestige lingua franca. In Sue Wright (ed.), *Language policy and language planning: From nationalism to globalisation*, 134–154. London: Palgrave Macmillan UK. https://doi.org/10.1007/978-1-137-57647-7_6.
- Wu, Ruey-Jiuan Regina. 2004. *Stance in talk: A conversation analysis of Mandarin final particles* (Pragmatics & Beyond New Series). Vol. 117. Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/pbns.117>.
- Xiao, Richard, Lianzhen He & Ming Yue. 2010. In pursuit of the “third code”: Using the ZJU Corpus of Translational Chinese in Translation Studies. In Richard Xiao (ed.), *Using Corpora in Contrastive and Translation Studies*, 182–214. Newcastle-upon-Tyne: Cambridge Scholars Publishing.
- Xiao, Richard & Xian Yao Hu. 2015. *Corpus-based studies of translational Chinese in English-Chinese translation* (New Frontiers in Translation Studies). Berlin, Heidelberg: Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-41363-6>.
- Xu, Beibei. 2012. *Nandao*-questions as a special kind of rhetorical questions. *Semantics and Linguistic Theory* 22. 508. <https://doi.org/10.3765/salt.v22i0.2643>.
- Yang, Yang, Stella Gryllia & Lisa Lai-Shen Cheng. 2020. *Wh*-question or *wh*-declarative? Prosody makes the difference. *Speech Communication* 118. 21–32. <https://doi.org/10.1016/j.specom.2020.02.002>.
- Yeung, Victoria Wai Lan & Yoshihisa Kashima. 2012. Culture and stereotype communication: Are people from Eastern cultures more stereotypical in communication? *Journal of Cross-Cultural Psychology* 43(3). 446–463. <https://doi.org/10.1177/0022022110395138>.

- Yip, Po-ching & Don Rimmington. 2021. *Chinese: An essential grammar* (Routledge Essential Grammars). 3rd edn. London, New York: Routledge.
- Yuan, Jiahong. 2006. Mechanisms of question intonation in Mandarin. In Qiang Huo, Bin Ma, Eng-Siong Chng & Haizhou Li (eds.), *Chinese spoken language processing* (Lecture Notes in Computer Science), vol. 4274, 19–30. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/11939993_7.
- Yuan, Jiahong. 2011. Perception of intonation in Mandarin Chinese. *The Journal of the Acoustical Society of America* 130(6). 4063–4069. <https://doi.org/10.1121/1.3651818>.
- Zhang, Ling. 2014. Segmentless sentence-final particles in Cantonese: An Experimental Study. *Studies in Chinese Linguistics* 35. 47–60.
- Zimmermann, Michael & Katharina Kaiser. 2019. Refining current insights into the *wh*-in-situ interrogative construction in French: The case of Contemporary Hexagonal French. *Romanistisches Jahrbuch* 70(1). 123–157. <https://doi.org/10.1515/roja-2019-0005>.
- Zribi-Hertz, Anne. 2011. Pour un modèle diglossique de description du français: Quelques implications théoriques, didactiques et méthodologiques. *Journal of French Language Studies* 21(2). 231–256. <https://doi.org/10.1017/S0959269510000323>.

References in Mandarin

- 安必諾，戎容（譯），2006，〈台灣文學在德、美、法三國：歷史及現狀一瞥〉，
《中外文學》34(10)，155–165。[Pino, Angel. 2006. (Trans.) Rong Rong. *Chung-Wai Literary Quarterly* 34(10). 155–165.]
- 朱德熙，1999，〈朱德熙文集〉，商務印書館：北京。[Zhu, Dexi. 1999. *Zhu Dexi wen ji*. 1st edn. Beijing: The Commercial Press.]
- 利氏，2001，〈利氏漢法辭典〉巴黎：Desclée De Brouwer。[Institut Ricci (ed.). 2001. *Grand dictionnaire Ricci de la langue chinoise*. Paris: Desclée De Brouwer.]
- 吳桐，2019，〈法漢疑問詞原位結構對比研究〉，〈解放軍外國語學院學報〉
42(2)，116–124。[Wu, Tong. 2019. A comparative study of *wh*-in-situ in French and Chinese. *Journal of PLA University of Foreign Languages*. 42(2). 116–124.]
- 呂叔湘，1999，〈現代漢語八百詞〉北京：商務印書館。[Lu, Shuxiang. 1999. *Xiandai Hanyu babai ci*. Beijing: The Commercial Press.]
- 卓立，2011，〈台灣文學作品法譯與接受〉，〈台灣文學館通訊〉32，18–23。
[Lin-Rosolato, Esther. 2011. The translation and reception of Taiwanese literary works into French. *Newsletter of the National Museum of Taiwan Literature* 32. 18–23]

- 屈承熹，2017，〈形各異而實相通－現代漢語句末情態虛詞與英語語調〉，《華語文教學研究》14(2)，41–64。[Chu, Chauncey C. 2017. Different in form but similar in function: Modern Mandarin utterance-final modality particles and english intonation. *Journal of Chinese Language Teaching* 14(2). 41–64.]
- 林于弘，林佳均。2008。〈九年一貫國語教科書標點符號教材內容分析比較〉，《教科書研究》1(1)。105–121。 <https://doi.org/10.6481/JTR.200806.0105>。[Lin, Yu-Hung & Chia-Chun Lin. 2008. The analysis and comparison of punctuation teaching content in grades 1-9 curriculum mandarin textbook. *Journal of Textbook Research* 1(1), 105-121.]
- 林家儀，邱銘心，2010，〈網路讀者書評文類分析研究：以科學普及類書評為例〉，《教育資料與圖書館學》，48(2)，269–298。
[https://doi.org/10.6120/JoEMLS.201012_48\(2\).0372.RS.AM](https://doi.org/10.6120/JoEMLS.201012_48(2).0372.RS.AM)。[Lin, Chia-Yu & Ming-Hsin Phoebe Chiu. 2010. Analysing online book reviews for scientific popularization genre: A genre analysis study. *Journal of Educational Media & Library Sciences* 48(2), 269-298.]
- 邱佳葳，2023，《誰幫你選書？論書評在不同年代所扮演之角色》中正大學碩士論文。台灣：嘉義。[Chiu, Chia-Wei, *Who chooses your books? On the role of book reviews across different eras*. Chiayi: National Chung Cheng University Master thesis.]
- 徐宏欣，2020，〈要，不要「呢」？疑問句構與篇章功能的互動〉，《華語文教學研究》，17(4)，71-94。[Hsu, Hung-Hsin. 2020. Ne, need it or not: Interface of interrogation and discourse function. *Journal of Chinese Language Teaching* 17(4). 71–94.]
- 袁野，2020，〈漢語疑問副詞原位與前置的非移位製圖範本闡釋〉，《語言暨語言學》21(3)，443–466。
<https://doi.org/10.1075/lali.00067.yua>。[Yuan, Ye. 2020. A non-movement cartographic template approach to Chinese in-situ and fronting wh-adjuncts. *Language and Linguistics*. 21(3). 443-466.]
- 袁暉，管錫華，岳方遂，2002，《漢語標點符號流變史》，武漢：湖北教育出版社。[Yuan, Hui, Xihua Guan & Fangsui Yue. 2002. *Hanyu biaodian fuhao liubian shi*. 1st edn. Wuhan Shi: Hubei Education Press.]
- 國家圖書館。2022，《111 年臺灣圖書出版現況及趨勢報告》，台北：國家圖書館。[National Central Library. 2022. *Taiwan book publishing current situation and trend report in 2022*. Taipei: National Central Library.]
- 國家圖書館。2023，《112 年臺灣圖書出版現況及趨勢報告》，國家圖書館：台北。[National Central Library. 2023. *Taiwan book publishing current situation and trend report in 2023*. Taipei: National Central Library.]

- 教育部國語推行委員會（編），2008，《重訂標點符號手冊》修訂版，中華民國（台灣）教育部：台北市。[National Languages Committee (ed.). 2008. *Revised manual of punctuation*. Taipei: Ministry of Education Republic of China (Taiwan)]
- 陳芳惠，2016，〈翻譯作為方法：台灣小說與世界文學〉，《台灣文學研究》(11)，173–199。[Chen, Fang-Hwey. 2016. Translation as method: The Taiwanese novel with respect to world literature. *Taiwan Literature Studies* (11). 173–199.]
- 湯延池，1981，〈國語疑問句的研究〉，《師大學報》26，219–277。[Tang, Ting-Chi. 1981. A study on Mandarin questions. *Shida xuebao (Journal of National Taiwan Normal University)* 26. 219–277.]
- 湯延池，1984，〈國語疑問句續論〉，《師大學報》29，381–437。[Tang, Ting-Chi. 1984. Further study on Mandarin questions. *Shida xuebao (Journal of National Taiwan Normal University)* 29. 381–437.]
- 詹衛東，郭銳，常寶寶，譚貽榮，陳龍，2019，〈北京大學CCL語料庫的研製〉，《語料庫語言學》6(1)，71–86。[Zhan, Weidong, Rui Guo, Baobao Chang, Yirong Chen & Long Chen. 2019. The building of the CCL corpus: Its design and implementation. *Corpus Linguistics* 6(1). 71–86.]
- 廖宜萱，2019，《法國小說在臺翻譯出版現況分析及展望》國立中央大學碩士論文。台灣：桃園。[Liao, Yi-Hsuan. 2019. *Analysis and outlook on the translation and publication of French fiction in Taiwan*. Taoyuan: National Central University Master thesis.]
- 趙春利，孫麗，2015，〈句末助詞“吧”的分布驗證與語義提取〉，《中國語文》365(2)，121–132。[Zhao, Chunli & Li Sun. 2015. On the distributional verification and semantic extraction of the sentence-final particle ba(吧). *Studies of the Chinese Language* 365(2). 121–132.]
- 齊滄揚，2002，〈“呢”的意義分析和歷史演變〉，《上海師範大學學報(哲學社會科學版)》31(1)，34–45。[Qi, Huyang. 2002. Analyse of 呢 in terms of meaning and historical evolution. *Journal of Shanghai Teachers University (Social Science)* 31(1). 34–45.]
- 編輯小組，2016，〈臺灣法文翻譯之現況與挑戰論壇〉，《翻譯論叢》，9(1)，141–158。[Editor group. 2016. The Current state and future prospects of French translation. *Compilation and Translation Review* 9(1). 141–158.]
- 蔡明月，董蕙茹，2009，〈臺灣地區的世界文學翻譯作品：書目計量分析〉，《圖書館學與資訊科學》35(2)，34–53。[Tsay, Ming-Yueh & Huei-Ru Dong. 2009. A study of world translated literature in taiwan: A bibliometric approach. *Journal of Library and Information Science* 35(2), 34–53.]

關首奇，2011，〈台灣文學在法國的現狀〉，《文史台灣學報》，(3)，131–163。
<https://doi.org/10.29511/TSLH.201112.0004>. [Gaffric, Gwennaël. 2011. The status of Taiwan literature in France. *Taiwan Studies in Literature and History* (3). 131–163. <https://doi.org/10.29511/TSLH.201112.0004>.]

References of novels

Novels in original Taiwan Mandarin and French translation

吳明益。2011。《複眼人》。台北：夏日出版。

紀蔚然。2011。《私家偵探》。台北：INK 印刻文學。

吳明益。2012。《天橋上的魔術師》。台北：夏日出版。

高翊峰。2014。《泡沫戰爭》。台北：寶瓶文化。

張國立。2019。《炒飯狙擊手》。台北：馬可孛羅文化出版。

陳思宏。2019。《鬼地方》。台北：鏡文學出版。

Wu, Ming-yi. 2014. *L'homme aux yeux à facettes*. Translated by Gwennaël Gaffric. Paris: Édition Stock.

Chi, Wei-jan. 2019. *Rue du dragon couché*. Traduit par Emmanuelle Péchenart. Paris : Calmann-Levy: Noir.

Wu, Ming-yi. 2017. *Le Magicien sur la passerelle*. Traduit par Gwennaël Gaffric. Paris: L'Asiathèque.

Kao, Yi-Feng. 2017. *La guerre des bulles*. Traduit par Gwennaël Gaffric. Bordeaux: Mirobole Édition.

Chang, Kuo-Li. 2021. *Le sniper, son wok et son fusil*. Traduit par Alexis Brossollet. Paris: Gallimard: Série Noir.

Chen, Sih-han (Kevin Chen). 2023. *Ghost Town*. Translated by Emmanuelle Péchenart. Paris: Éditions du Seuil.

Novels in original French and Taiwan Mandarin translation

Lévy, Marc. 2010. *Le voleur d'ombres*. Paris: Robert Laffont.

Thilliez, Franck. 2011. *Gataca*. Paris: Fleuve Noir.

Bussi, Michel. 2013. *Un avion sans elle*. Paris: Presses de la Cité.

Barbery, Muriel. 2015. *La vie des elfes*. Paris: Gallimard.

Lemaitre, Pierre. 2016. *Trois jours et une vie*. Paris: Albin Michel.

Le Tellier, Hervé. 2020. *L'Anomalie*. Paris: Gallimard.

馬克·李維。《偷影子的人》。段韻靈譯。台北：商周出版，2011。

法蘭克·西萊斯。《左撇子的殺意》。張喬玫譯。台北：寂寞出版，2012。

米榭·卜熙。《沒有她的飛機》。梁若瑜譯。台北：皇冠出版，2014。

妙莉葉·芭貝里。《精靈少女》。吳馨竹、段韻靈、粘耿嘉合譯。台北：商周出版，2015。

皮耶·勒梅特。《三天一生》。繆詠華譯。台北：時報文化，2018。

埃爾韋·勒·泰利耶。《異常》。陳詠薇譯。台北：木馬文化，2023。